

Névtér, 2018

| | | |
|------------|---|----|
| 1 | Az individuumok tér- és időfüggése..... | 6 |
| 1.1 | A nemzeti névterek fogalma | 6 |
| 1.1.1 | Miért van szükség nemzeti (és globális) névterekre? | 9 |
| 1.1.2 | A névterek szintjei és azok egymásra épülése | 9 |
| 1.2 | Általános modellezési kérdések | 12 |
| 1.2.1 | Többet a KOS-okról | 12 |
| 1.2.2 | Névterek modellezése | 15 |
| 1.3 | Földrajzi névtér | 16 |
| 1.3.1 | A földrajzi entitások modelljének kérdései | 16 |
| 1.3.2 | A földrajzi névtér modellje | 17 |
| 1.3.2.1 | Geoindividuumok azonosítása | 17 |
| 1.3.2.2 | Idő- és eseménykezelés | 18 |
| 1.3.2.3 | Geometriakezelés | 19 |
| 1.3.2.4 | Névkezelés | 19 |
| 1.3.2.5 | Relációk | 20 |
| 1.3.2.6 | Földrajzi névtér logikai adatmodellje | 21 |
| 1.4 | A nyelvi elemek kezelése | 22 |
| 1.5 | Személynévtér | 25 |
| 1.5.1.1 | Személynévtér logikai adatmodellje | 26 |
| 1.6 | Testületi névtér | 27 |
| 1.6.1.1 | Testületi névtér logikai adatmodellje | 27 |
| 1.7 | Eseménynévtér | 28 |
| 1.7.1 | Mi az az idővonal | 29 |
| 1.7.1.1 | Alapfogalmak | 30 |
| 1.7.1.1.1 | idővonal(szolgáltatás) | 30 |
| 1.7.1.1.2 | naptár(szolgáltatás) | 30 |
| 1.7.1.1.3 | idővonalszerver (timeline server) | 30 |
| 1.7.1.1.4 | kalendárszerver (calendar server) | 30 |
| 1.7.1.1.5 | kronológia-adatbázis (chronology database) | 30 |
| 1.7.1.1.6 | idő (time) | 30 |
| 1.7.1.1.7 | esemény (event) | 30 |
| 1.7.1.1.8 | időpontosság (time precision) | 32 |
| 1.7.1.1.9 | időformátum (time format) | 32 |
| 1.7.1.1.10 | relatív idő (relative time) | 32 |
| 1.7.1.1.11 | elemi esemény (elementary event) | 32 |
| 1.7.1.1.12 | naptáresemény (calendar event) | 32 |
| 1.7.1.1.13 | idővászon (time canvas) | 33 |
| 1.7.1.1.14 | sáv (band) | 33 |
| 1.7.1.1.15 | időszakdoboz (period box) | 33 |
| 1.7.1.1.16 | időtengely (time axis) | 33 |
| 1.7.1.1.17 | időfelbontás (time granularity) | 33 |
| 1.7.1.1.18 | időfelbontási szint (granularity level) | 34 |
| 1.7.1.1.19 | időközelítés (time zoom-in) | 34 |
| 1.7.1.1.20 | időtávolítás (time zoom-out) | 34 |
| 1.7.1.1.21 | időablakcsúsztatás (time pan) | 34 |
| 1.7.1.1.22 | időablak (time window) | 34 |
| 1.7.1.1.23 | narratíva (narrative) | 34 |
| 1.7.1.1.24 | interpretáció (interpretation) | 35 |
| 1.7.1.1.25 | narratívaváltás (changing narratives) | 36 |
| 1.7.1.1.26 | eseménykapcsolatok (event relationships) | 36 |
| 1.7.1.1.27 | kulcsesemény (key event) | 36 |

| | | |
|------------|--|----|
| 1.7.1.1.28 | gestalt alakzat (gestalt pattern) | 37 |
| 1.7.2 | Adatmodel | 37 |
| 1.8 | Írásműcímter | 40 |
| 1.9 | Köznévtér | 43 |
| 1.9.1 | Tudásszervezési rendszerek | 43 |
| 1.9.1.1 | Terminuslista | 45 |
| 1.9.1.2 | Taxonómia | 45 |
| 1.9.1.3 | Tezaurusz | 45 |
| 1.9.1.4 | Formális ontológia | 45 |
| 1.9.1.5 | Folkszonómia | 45 |
| 1.9.2 | A köznévterek és a tudásszervezési rendszerek kapcsolata | 46 |
| 1.10 | Névterek közti kapcsolatok | 46 |
| 1.10.1 | Archontológia | 46 |
| 1.10.2 | Iskolázottság | 47 |
| 1.10.3 | Társadalomföldrajz | 47 |
| 1.10.4 | Választás | 47 |
| 1.10.5 | További névtérközi adatkörök | 48 |
| 2 | A névterek fogalma | 49 |
| 2.1 | A névterek fontosságáról | 49 |
| 2.2 | Létező névtérhasználati gyakorlatok | 51 |
| 2.3 | Fájlnevek | 51 |
| 2.4 | Domainnevek | 52 |
| 2.5 | Emailcímek | 52 |
| 2.6 | Telefonszámok | 53 |
| 2.7 | Földrajzi helyek | 53 |
| 2.7.1 | Postai címek | 53 |
| 2.7.2 | Földrajzi térképek, földrajzi nevek | 53 |
| 2.7.2.1 | GeoNames | 54 |
| 2.7.2.2 | Getty TGN | 55 |
| 2.7.2.3 | Google Maps | 57 |
| 2.7.2.4 | OpenStreetMap | 58 |
| 2.8 | Webes névazonosítási szabványok | 59 |
| 2.9 | Személynevek | 60 |
| 2.10 | Műcímek | 61 |
| 2.10.1 | Könyvek, folyóiratok | 62 |
| 2.10.1.1 | ISBN | 62 |
| 2.10.1.2 | ISSN | 65 |
| 2.10.1.3 | ISTC | 67 |
| 2.10.1.4 | DOI | 68 |
| 2.10.1.5 | NBN | 68 |
| 2.10.2 | Zeneszámok | 68 |
| 2.10.2.1 | ISMN | 68 |
| 2.10.2.2 | ISRC | 70 |
| 2.10.2.3 | ISWC | 70 |
| 2.10.3 | Filmek, audiovizuális tartalmak | 73 |
| 2.10.3.1 | ISAN | 73 |
| 2.10.3.2 | IMDb | 82 |
| 2.11 | Személyek azonosító rendszerei | 84 |
| 2.11.1 | Tudományos kutatók és cikkek azonosító rendszerei | 84 |
| 2.11.1.1 | MTMT | 85 |
| 2.11.1.2 | ODT | 87 |
| 2.11.1.3 | ORCID | 88 |
| 2.11.1.4 | ResearcherID | 89 |
| 2.11.1.5 | Web of Science | 92 |

| | | |
|----------|---|-----|
| 2.11.1.6 | Google Scholar..... | 93 |
| 2.11.1.7 | Scopus..... | 94 |
| 2.11.1.8 | ResearchGate..... | 95 |
| 2.11.1.9 | Academia.edu..... | 97 |
| 2.11.2 | További azonosító rendszerek személyekre, helyekre..... | 97 |
| 2.11.2.1 | VIAF..... | 98 |
| 2.11.2.2 | ISNI..... | 99 |
| 2.11.2.3 | PIM személynévtér..... | 104 |
| 2.11.2.4 | ULAN..... | 105 |
| 2.11.2.5 | GDN..... | 107 |
| 2.11.2.6 | BNF..... | 107 |
| 2.11.2.7 | A könyvtárak besorolási állománya..... | 108 |
| 2.11.2.8 | ISADN..... | 109 |
| 2.11.2.9 | Wikidata..... | 109 |
| 2.12 | Névstruktúra..... | 110 |
| 2.13 | Relációelmélet..... | 111 |
| 2.13.1 | Relációelméleti alapfogalmak..... | 111 |
| 2.13.1.1 | A reláció fogalma..... | 111 |
| 2.13.1.2 | Relációtulajdonságok..... | 112 |
| 2.13.1.3 | Függvény..... | 113 |
| 2.13.1.4 | Művelet..... | 113 |
| 2.13.1.5 | Nevezetes matematikai relációk..... | 114 |
| 2.13.1.6 | Relációműveletek..... | 115 |
| 2.13.1.7 | Reifikáció..... | 116 |
| 2.13.2 | Ontológiai alaprelációk..... | 116 |
| 2.13.3 | KOS-relációk..... | 120 |
| 2.13.3.1 | Generikus reláció..... | 121 |
| 2.13.3.2 | Partitív reláció..... | 122 |
| 2.13.3.3 | Szinonima..... | 124 |
| 2.13.3.4 | Homonima..... | 124 |
| 2.13.3.5 | Meghatározója (előzménye, következménye) reláció..... | 125 |
| 2.13.3.6 | Hierarchikus alá-fölérendeltje (tartalmazás) reláció..... | 125 |
| 2.13.3.7 | Lexikografikus rendezési reláció..... | 126 |
| 2.13.3.8 | Rokona (egyéb) reláció..... | 126 |
| 2.14 | A használatba vehető relációk listája..... | 126 |
| 2.15 | Tudásszervezési rendszerek..... | 127 |
| 2.15.1 | Formális keretrendszer..... | 127 |
| 2.15.2 | Terminuslista..... | 129 |
| 2.15.2.1 | Formális jellemzés..... | 130 |
| 2.15.3 | Taxonómia..... | 131 |
| 2.15.3.1 | Formális jellemzés..... | 135 |
| 2.15.4 | Tezaurusz..... | 136 |
| 2.15.4.1 | Formális jellemzés..... | 142 |
| 2.15.5 | Folkszonómia..... | 143 |
| 2.15.5.1 | Formális jellemzés..... | 159 |
| 2.15.6 | Formális ontológia..... | 160 |
| 2.15.6.1 | Formális jellemzés..... | 160 |
| 2.16 | Dokumentumleíró rendszerek..... | 161 |
| 2.17 | Tulajdonnév és megnevezés..... | 161 |
| 2.18 | A névtér összetevői..... | 161 |
| 2.19 | A tulajdonnévtér meghatározása..... | 162 |
| 2.20 | A nevek egyértelműsítése..... | 163 |
| 2.21 | Lokális, nemzeti és nemzetközi névtér..... | 166 |
| 2.22 | A név fogalma, funkciója..... | 166 |

| | | |
|-----------|--|-----|
| 2.23 | A névtér fogalma, funkciója..... | 166 |
| 2.24 | Lokális névterek..... | 168 |
| 2.25 | Nemzeti névterek | 168 |
| 2.25.1 | A nemzeti névtér szűkebb értelmezése | 169 |
| 2.25.2 | A nemzeti névtér tágabb értelmezése (névtár) | 170 |
| 2.26 | Globális névterek | 171 |
| 2.27 | A névtér fogalmának ontológiai elkötelezettségei | 172 |
| 2.28 | A természeti és társadalmi tények különbsége..... | 172 |
| 2.29 | Névterek és tudásszervezési rendszerek..... | 172 |
| 2.29.1 | A tulajdonnévterek struktúrája | 173 |
| 2.29.1.1 | A személynévtér struktúrája | 173 |
| 2.29.1.2 | A területi névtér struktúrája | 174 |
| 2.29.1.3 | A földrajzi névtér struktúrája | 176 |
| 2.29.1.4 | Az írásműcímter struktúrája..... | 181 |
| 2.29.1.5 | Az audiovizuálismű-címterek struktúrája | 184 |
| 2.29.2 | A köznévterek struktúrája | 184 |
| 2.30 | A névtérdefiníció fejlesztésre gyakorolt hatásai | 185 |
| 3 | KOS-alapú keresés..... | 187 |
| 3.1 | Bevezetés | 187 |
| 3.2 | Alapfogalmak..... | 188 |
| 3.2.1.1 | Információ..... | 190 |
| 3.2.1.1.1 | Szöveg | 190 |
| 3.2.1.1.2 | Hipertext | 195 |
| 3.2.1.1.3 | Grafonyelvi technika | 196 |
| 3.2.1.1.4 | Adatbázis | 199 |
| 3.2.1.1.5 | Tudásszervezési rendszer..... | 203 |
| 3.2.1.2 | Dokumentum..... | 211 |
| 3.2.1.3 | Metainformáció..... | 213 |
| 3.2.1.4 | Gyűjtemény, információs tartomány..... | 216 |
| 3.2.1.5 | Információelérés és metaadat..... | 219 |
| 3.3 | Információelérés | 220 |
| 3.3.1.1 | Információs igény | 221 |
| 3.3.1.2 | Relevanciakezelés | 222 |
| 3.3.1.2.1 | Relevancia mint mértékkategória | 222 |
| 3.3.1.2.2 | Relevancia mint érték kategória | 223 |
| 3.3.1.2.3 | Relevanciafogalmak értékelése | 225 |
| 3.3.1.3 | Információs igények kielégítése és a felhasználó aktivitás..... | 225 |
| 3.3.1.3.1 | Keresés..... | 226 |
| 3.3.1.3.2 | Böngészés | 226 |
| 3.3.1.3.3 | Műsorkövetés..... | 227 |
| 3.3.1.4 | Információs igények kielégítése és az információs térstruktúra | 228 |
| 3.3.1.4.1 | Térbeli navigáció | 229 |
| 3.3.1.4.2 | Szemantikus navigáció | 238 |
| 3.3.1.4.3 | Társas navigáció | 258 |
| 3.3.1.5 | Az információelérés típusai | 268 |
| 3.4 | Gyakorlati következmények | 268 |
| 3.4.1.1 | A névterekben használt kiemelt relációk pontosítása | 268 |
| 3.4.1.2 | Relációtípus és relációpéldány..... | 273 |
| 3.4.1.2.1 | A mintát adó filozófiai megoldás: az eseménypéldány fogalma | 274 |
| 3.4.1.2.2 | Relációpéldány a triplet-logika szolgálatában | 276 |
| 3.4.1.3 | A névterekben és KOS-ban tárolt tudás viszonya..... | 279 |
| 3.4.1.3.1 | Névtérmodell, nyelvi modell | 279 |
| 3.4.1.3.2 | A nevek kezelése a nyelvi modellben..... | 285 |
| 3.4.1.3.3 | A tudásszervezési rendszerek kezelése a nyelvi modellben | 288 |

| | | |
|-----------|--|-----|
| 3.4.1.3.4 | Névterek és a nyelvi modell viszonya | 289 |
| 3.4.1.4 | Relevanciakezelés a névtérben..... | 289 |
| 3.4.1.4.1 | Értékkezelés a névtérben | 289 |
| 3.4.1.4.2 | Mértékkezelés a névtérben | 292 |
| 4 | Tér- és időérzékeny keresés..... | 297 |
| 4.1 | Térérzékeny keresés..... | 297 |
| 4.1.1.1 | Partitív reláció..... | 297 |
| 4.1.1.1.1 | A partitív relációt leíró elmélet: mereológia..... | 297 |
| 4.1.1.1.2 | Ontológiai alaprelációk..... | 299 |
| 4.1.1.2 | Térinformatikai áttekintés..... | 302 |
| 4.1.1.3 | Relációs adatbázisok geometria kezelésének lehetőségei..... | 303 |
| 4.1.1.3.2 | Tér adatok kezelése..... | 305 |
| 4.1.1.3.3 | Geometriák egymáshoz való viszonyának meghatározása | 306 |
| 4.1.1.3.4 | Térbeli adatok műveletei | 310 |
| 4.1.1.3.5 | Geometriák feldolgozása | 314 |
| 4.1.1.3.6 | Topológia téradatbázisban | 319 |
| 4.1.1.3.7 | Térbeli referencia rendszerek alkalmazhatóság szempontjából..... | 331 |
| 4.2 | Gyakorlati következmények | 333 |
| 4.2.1.1 | A térérzékeny keresés gyakorlati lehetőségei a névterekben..... | 333 |
| 4.2.1.2 | Időérzékeny keresés | 335 |
| 5 | Hivatkozások | 337 |
| 5.1.1 | Szabványok, szolgáltatások..... | 347 |

1 Az individuumok tér- és időfüggése

1.1 A nemzeti névterek fogalma

A körülöttünk levő dolgokat gyakran valamilyen névvel illetjük. A világban való eligazodásunk fontos része a nevek használata. Nevén szólítjuk az ismerőseinket, nevet adunk a háziállatainknak, a szervezeteinknek, elnevezzük (megcímezzük) a földrajzi tér számunkra fontos elemeit, részeit, de nevet (címet) adunk a nekünk fontos dokumentumoknak (filmeknek, könyveknek, zeneműveknek stb.), sőt, olykor a használati tárgyaink közül is meg-megnevezünk párat (az autónknak kötelező módon adunk speciális nevet, rendszámot, de az is előfordul, hogy valaki "emberszerű" névvel illeti saját kocsiját). A nevek használatának értelme az, hogy a nevekkel dolgokra tudunk hivatkozni, "rámutatni". Ez a nagyon általános körbeírás megengedi, hogy a névhasználat két – egymástól jelentős mértékben különböző – módját is megragadhasuk vele. Amennyiben a nevek használatával valamely dologra (vagy dolgok egy halmazára) mutathatunk rá, akkor fogalmilag érdemes elkülöníteni egymástól ennek a relációnak a két oldalát (relátumát). A **név** önmagában nem létezik, csak akkor tudjuk értelmesen megragadni, ha figyelembe vesszük az(oka)t a dolgo(ka)t, ami(k)re a név hivatkozik. Ez utóbbiakat hívhatjuk **névhordozóknak**.

A nevek, a névhasználat jelenségét a névhordozókkal együtt lehet csak jól megérteni, pontosan leírni.

Amikor nevekről beszélünk, akkor mindig valamilyen entitás(típus) mint névhordozó(típus) megnevezési és névhasználati gyakorlatáról beszélünk.

A **névtér** (namespace) nevek valamilyen strukturába rendezett halmaza vagyis nevek adatbázisa, valamint a nevek beazonosítását, a névadatbázis javítását, bővítését lehetővé tevő alkalmazás. A logikai minőségét tekintve a névtérnek két típusát,

a köznévtérrel és tulajdonnévtérrel különöníthetjük el egymástól.

A **köznévtér** elemei nyelvtani státusukat tekintve köznevek, amelyekkel valamilyen általános (típus vagy osztály) fogalomra hivatkozunk, ami által ezek terjedelmébe mindig egyedi előfordulások valamilyen halmaza tartozik. Az általános fogalmak fontos tulajdonsága, hogy – elvileg – bármelyikük alá mindig rendelhető, képezhető valamilyen – szűkebb terjedelmű – fogalom a **generikus alárendeltje** reláció segítségével. Az általános fogalmaink (közneveink) a világról való tudásunk kifejezői, ezért valamilyen elv szerinti rendszerbe szervezésüket szokás **tudásszervezési rendszer** (KOS – Knowledge Organization System) névvel is illetni. A tudásszervezési rendszerek legfontosabb (és egyben legismertebb) típusai az **osztályozási rendszerek** (pl. ETO), **tezauruszok** (pl. Köztaurusz), **ontológiák** (pl. DOLCE).¹

A névterek másik nagy típusát a **tulajdonnévtér** jelenti. Ebben az esetben a névtér elemei nyelvtani státusukat tekintve tulajdonnevek, amelyekkel mindig valamilyen egyedtípus konkrét egyedeinek előfordulásaira hivatkozunk. A tulajdonnevek használatának lényegi célja egyedi módon hivatkozni valamilyen konkrét egyedre valamilyen kontextuson belül. Jól ismert az a tény, hogy ezt a célt a gyakorlatban sokszor nem érik el (legtöbbször a kontextusok nem megfelelő kezelése miatt), de ez a

¹ A tudásszervezési rendszerekről egy későbbi fejezetben bővebben szót ejtünk.

sikertelenség nem változtatja meg sem a névhasználati szándék eredeti tartalmát, sem a tulajdonnevek logikai minőségét. A tulajdonnevek lényegük szerint nem bonthatók fel, nem rendelhető alájuk szűkebb terjedelmű fogalom, a legfontosabb reláció számukra az a típusmeghatározás, amely alapján tudjuk, hogy milyen típusú egyed (tulajdon)neveként (merek jelölőjeként) használjuk őket.²

Nagyon gondosan el kell választani a kétfajta névtípust egymástól, mert a kezelésük, rendszerbe szervezésük sok szempontból eltérő logikát, más megoldásokat, technológiákat, szabályzatokat kíván meg. A tulajdonnévterek további felosztását aszerint végezhetjük el, hogy milyen névhordozó entitástípusba tartozó egyedek megnevezéséről van szó. A közgyűjtemények világában többféle általános egyedtípust határolhatunk el, amelyek közül van olyan, amit további altípusokra bonthatunk, van olyan, amit nem, illetve van, ami jobban, és van, ami kevésbé van átfedésben a különböző közgyűjtemények között. A legfontosabb tulajdonnévterek azok, amelyek minden kulturális területen használatba vehetők, ezek építésével kell az együttműködést kezdeni.

- földrajzi helyek
- ágensek
- dokumentumok
- események

A földrajzi helyek, földrajzi entitások alá sorolhatunk poligon, vagy vonal- vagy pontszerű dolgokat (a földrajzi térben megfelelő határokkal, vonalakkal, illetve pontokkal jelölve őket) úgy mint: országok, megyék, települések vagy utcák, folyók, vezetékek vagy boltok, bankautomaták, templomok, házak. Ezek mindegyikéhez rendelhetünk nevet és ilyenkor beszélünk **földrajzi nevekről**.

A közgyűjteményi gyakorlatban az ágensek azok a szereplők, akik a dokumentumok megalkotásában, létrehozásában, sokszorosításában, gyártásában, terjesztésében, finanszírozásában részt vesznek. Az ágensek két nagy típusa a **személy** és a **testület**. A közgyűjteményi – és azon belül elsősorban a könyvtári – gyakorlatban az ágensek egyértelmű nyilvántartására, kezelésére jöttek először létre azok az authority rendszerek, amelyek a névterek első – még nem teljes funkcionalitású – használatának tekinthetők. A fontosabb szerep, az ebből következő nagyobb érdek, valamint a könnyebb megvalósíthatóság miatt a **személynévterek** építésében érték el eddig a legnagyobb eredményeket. A **testületi névterek** emberek valamilyen csoportját, szervezeteket, cégeket, csoportosulásokat egyként meghivatkozó testületi nevek rendezett és kontrollált rendszere, amelyek építésében kevesebb előrehaladás történt eddig.

A közgyűjtemények számára legfontosabb entitás a **dokumentum**, igaz, ebből a különféle altípusok más és más közgyűjtemény számára lehetnek fontosak. Többféle módon is csoportosíthatjuk a dokumentum fogalmát, és ez a kérdés most nem pusztán elméleti, de gyakorlati szempontok miatt is különösen fontos. A hagyományos felosztás a hosszú évtizedek, évszázadok alatt kialakult intézményi elkülönültség mentén csoportosította a dokumentumokat. Ez az intézményi felosztás részben az információ, részben az információt hordozó médium típusa szerint végezte el a rendszerezést, amikor tévémsorszámokról, filmről, VHS-kazettáról, dvd-ről, online videóról, könyvről, periodikáról, cikkről, weboldalról, e-bookról, bakelitlemezeiről, cd-lemezeiről, cd-rom lemezeiről, plakátról, kottáról stb. mint önálló dokumentumtípusról beszélt. Ebben a kevert elvű felosztásban külön dokumentumtípusnak számított, tehát külön kellett leírni, ha ugyanaz a mozgókép mozifilmként, DVD-n, tévében, VHS-kazettán, online videótékán keresztül volt megnézhető, de ebben a logikában a papír alapú könyv és az elektronikus könyv is kétféle dokumentumtípusnak számított. A különböző médiumok, illetve az ezekhez kapcsolódó különböző intézmények szerinti elkülönítés logikája természetesnek látszott a legtöbb érintett szereplő számára, és az effajta magátólértéketeket mindig messzemenően tiszteletben kell tartani, mert az intézményesült szervezeti gyakorlatok, rutinok hihetetlen erővel képesek fenntartani önmagukat. A nagy veszély azonban az, hogy a digitális univerzum kiteljesedésével, a teljes digitalizálódás folyamatával érvényüket veszthetik a korábban megingathatatlanak vélt hitek, gyakorlatok, kiürülnek a korábbi logikák, praxisok, amikor a korábbi – jó – intézményi gyakorlatok diszfunkcionálisakká vál(hat)nak. információ típusok szerint szöveg, kép, mozgókép, hang típusairól kell beszélnünk. A leglátványosabban, legszemléletesebben a mozgókép példáján lehet érzékeltetni ezt a jelenséget a korábban már említett felsorolásra utalva, hogy ti. ugyanazt a filmet nézhetjük meg a moziban, a tévében, a számítógép képernyőjén dvd-lemezeiről vagy egy okostelefonon a hálózaton

² A névterek fogalmának pontos meghatározását egy másik tanulmányban végezzük el.

keresztül letöltve. A nagy kérdés persze az, hogy kell-e, szabad-e ezzel a váltással foglalkoznunk, figyelembe kell-e venni a váltásból eredő következményeket. Az archívumi, könyvtári szakma már régóta kidolgozta azt a fogalmi modellt (az FRBR-t), ami elvileg alkalmas lenne arra, hogy ezt a jelenséget megfelelő módon meg lehessen ragadni, de mindeddig csak nagyon kevés sikeres és nagy adattartalommal rendelkező projektről lehetett hallani, ami a megvalósítás nyilvánvaló nehézségeire (és veszélyeire) utal. Ezt a kérdéskört nagyon megfontoltan végig kell vitatni, mert bármilyen döntést is hozunk, mindegyiknek nagyon komoly következményei lesznek.

Kell-e, lehet-e, szabad-e használni az FRBR-modellt?

Van még egy olyan entitástípus, amely esetében beszélhetnénk névterekről, de egy ilyen névtér felépítése útjában komoly akadályok lennének. A dolgok időbeliségét eseményszerű fogalmak segítségével fejezzük ki, és a konkrét eseményeket elég gyakran meg is nevezzük, ezért felvethető lenne, hogy a partikuláris események számára is fel lehetne építeni egy **eseménynévteret**. Ez az elvi lehetőség azonban nagyon nehezen lenne megvalósítható a gyakorlatban egyszerűen azért, mert nem vagy csak nagyon nehezen tehetnék teljessé, konzisztenssé az ilyen névteret. Mivel fontos eseményeket valóban gyakran megnevezünk (első világháború kirobbanása, 1948-as magyar szabadságharc kezdete, a pákozdi csata stb.), könnyen az a csalóka kép alakulhat ki bennünk, hogy az individuális eseményeket mindig meg lehet nevezni. Ez elvileg igaz is, gyakorlatilag nem. Egy teljesség igényére tartó eseménynévtér könnyen és villámgyorsan beláthatatlan méretűvé dagadhatna, hiszen gyakorlatilag minden más entitást és az entitások egymáshoz való viszonyát is mindig az időben kellene/lehetne kifejezni, ami megfoghatatlan méretű, kontrollálhatatlan feladattá terebélyesedhetne pillanatok alatt. Az természetesen megfontolható, hogy előre rögzített elvek mentén valamilyen nagyon korlátozott "hatókörű" eseménynévteret el lehet kezdeni felépíteni, de csak a szükségszerű korlátoltság, behatároltság tudatában szabad egy ilyen munkába belekezdeni.

Mielőtt a névterek építésébe belevágnánk, fontos figyelmeztetni magunkat, hogy a hálózat világában van egy – már elég régóta működő – névtérkezelő rendszer, amelytől sokat lehet és kell tanulni. Az internetes domainnevek regisztrálását, nyilvántartását, felismerését végző **DNS-rendszer** (domain name system) egy olyan szolgáltatás, amely egyedi domainneveket regisztrálását (kiosztását) és az egyedi IP-címekhez való rendelését biztosítja. Mind az IP-címek, mind a domainnevek egyediek, individuálisak, tehát technikai értelemben tulajdonneveknek tekinthetőek. A domainnevek mint tulajdonnevek kiosztása, felismerése, feloldása a web kezdetétől fogva működik mégpedig úgy, hogy ezt a feladatot nem egyetlen, hanem sok szervezet végzi egyszerre, osztott módon. A névterek közt meglevő nyilvánvaló különbségek ellenére

a lehető legtöbb jó megoldást át kell venni a DNS-praxisból,

hiszen ez már évtizedek óta sikeresen bizonyítja a saját létét megalapozó elvek hatékonyságát, működőképességét. A domainnevek kezelésének praxisából már előzetesen is ki lehet ragadni egy olyan elvet, amely az osztott adatkezelés működőképességét biztosítja. Ez a nyitottság, a névtér működéséhez szükséges bármiféle adat szabad, mindenféle korlátozástól mentes megismerhetőségének és használatának az elve. Ezt az elvet legalább négyféle területre, adat- és tudástípusra érdemes kiterjeszteni úgymint:

- **nyílt modell** (open model): minden névteret kiszolgáló rendszer strukturáltságát biztosító adatmodellt nyilvánossá kell tenni, hogy bárki pontosan megismerhesse az adatrendszerek belső logikáját
- **szabad adat** (open data): a névterek minden adatát bárki használhatja (valamilyen CC-licenz alapján)
- **nyílt forráskódú program** (open source application): a névterek működéséhez szükséges bármilyen alkalmazási igényt nyílt forráskódú programokkal kell kiszolgálni, amiből következően ezeket a programokat bárki továbbhasználhatja saját céljaira
- **(részben) nyílt együttműködés** (open or mixed collaboration): bizonyos névterek esetében a teljes kollaborációt, más esetekben a közgyűteményi kollaborációt kell vagy lehet megengedni,

de még a korlátozott módú együttműködési rendszerekben is lehetővé kell tenni az intézményen kívüli kollaborációs tevékenységek eredményeinek valamilyen szintű – csúsztatott, kontrollált – becsatlakoztatását.

1.1.1 Miért van szükség nemzeti (és globális) névterekre?

A különböző tartalmú gyűjtemények saját gyűjteményi körüknek megfelelő dokumentumokat gyűjtenek és tesznek elérhetővé. A hálózati elérés megszabadította a felhasználókat, látogatókat a térbeli és időbeli kényszerektől, vagyis a hálózati gyűjteményeket bárhol, bármikor meglátogathatjuk. Ez nyilvánvalóan hatalmas előrelépés a hagyományos gyűjteményi szolgáltatások kínálatához képest. A hálózat ebben az értelemben valóban eltörli a térbeli és időbeli korlátokat, vagyis nyitottabbá teszi világunkat. Ebben a konstrukcióban azonban egyfajta zárttság továbbra is megmarad, noha a hálózati technológiák ugyanúgy lehetővé tennék meghaladni, eltüntetni ezt a fajta zártsgot is. Amikor már bármely gyűjteményt, bármely dokumentumot, bármikor elérhetünk a hálózaton keresztül, akkor még mindig szembesülnünk kell az intézményi elkülönültség tényével – és bizonyos esetekben – hátrányaival. Arra vagyunk kényszerítve ugyanis, hogy egymás után bejárjuk mindazokat a hálózati helyeket, ahol a minket érdeklő adatokat reméljük megtalálni. Ez az hálózati elkülönültség tükrözi a hagyományos intézményi elkülönültséget, ennyiben természetesnek mondható, de a felhasználói élmény, a látogatók szempontjából tekintve ez fölösleges és értelmelen elkülönültség. A látogatókank az lenne kívánatos, hogy egyetlen keresés alapján megtalálják az összes olyan dokumentumot, adatot, ami őket érdekli, és amely dokumentumok és adatok esetleg szét vannak szórva a különböző gyűjtemények között. A felhasználói igény a gyűjtemények közti teljes átjárhatóság, interoperabilitás, őket nem érdekli az intézményi elkülönültség, hiszen az esetek döntő többségében nem intézményeket, hanem az intézmények által kezelt dokumentumokat, adatokat keresik. A feladat tehát az, hogy biztosítani tudjuk valahogyan ezt a fajta – szemantikai alapú – átjárhatóságot vagyis az adatszintű interoperabilitást. A kérdés már csak az, hogy mindezt hogy lehet elérni. A válasz egyszerű: az intézmények közti névterek felépítésével és közös használatával.

A gyűjtemények közötti átjárhatóságot a közös névterek segítségével lehet biztosítani.

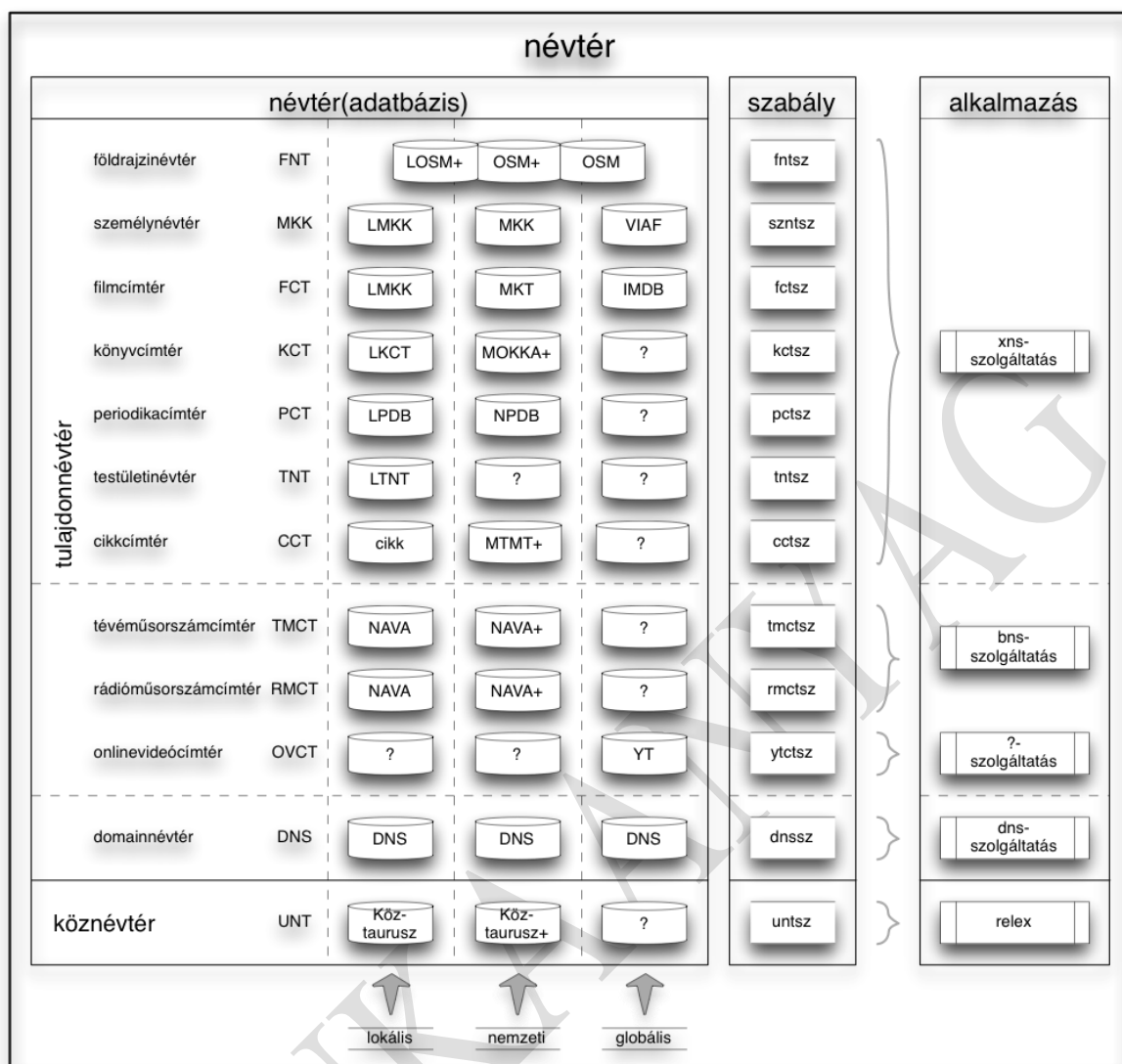
Azért az intézményközi adatátjárhatóság a kulcs, mert az (online) gyűjtemények más gyűjtőkörrel rendelkeznek ugyan, tehát az általuk szolgáltatott dokumentumok között nincs átfedés (vagy kevés van), viszont a dokumentumokkal kapcsolatba hozható névhordozó entitások, személyek, testületek, földrajzi helyek, illetve a dokumentumok leírásához használt fogalomkészletek - legalább részben – közösek. És ha ezeket közös névterek segítségével írjuk le, akkor ezek a közös pontok természetes ugrópontokat, bejárési útvonaljelzőket jelentenek a tartalmukat tekintve nagyon különböző gyűjtemények között.

1.1.2 A névterek szintjei és azok egymásra épülése

A névterek használatának technikai értelme abban van, hogy egyértelműsítjük, illetve normalizáljuk mind magukat a névhordozó entitások, mind az adott névtípusba tartozó nevek halmazát. Az egyértelműsítés azt jelenti, hogy egyfelől elkülönítjük egymástól az azonos névvel rendelkező entitásokat (névhomonimitás feloldása), másfelől összekapcsoljuk egymással az ugyanazon névhordozókhoz tartozó különböző neveket (névszinonimitás beazonosítása). A normalizálás lényege az, hogy egy információt (itt: minden névvel kapcsolatos információt) egyetlen helyen tárolunk, és a nevek minden más használatát azonosító számokkal helyettesítjük. Ez megszünteti a redundáns és inkonzisztens adatkezelés lehetőségét. Másodlagos előny származik ebből amiatt, hogy az így tárolt adatokat folyamatosan és könnyen újrahasznosíthatjuk. A névterek építésének azonban van egy komoly bekerülési költsége, és a névterek használatának előnyei csak a kezdeti munkák, a kezdeti költségek befektetése után (annak eredményeként) jelentkeznek. Amíg a névterek építése az egyik oldalon kezdetköltséges, addig a másik oldalon erőteljes networkhatás érvényesül, vagyis minél többen használják az adott névteret (akár intézményi, akár intézményközi kooperációban, annál jobban tapasztalhatóvá, érzékelhetővé válnak a közös használat előnyei. A normalizáláshoz szükséges

felderítési, egyértelműsítési munkákat minél többen végzik, annál előbb és annál kiterjedtebb módon végzi el a közösség egésze a szükséges munkákat, és annál többen és annál gyorsabban élvezik a közös névtérhasználat előnyeit. A névterek különböző szintű adatrendszerben használhatók, és a névtérhasználó intézményi kört tekintve lehet lokális, nemzeti, illetve globális névterekről beszélni. A névterek használatának elemi színtereit az egyes adatrendszereket, a **lokális névtereket** építő intézmények jelentik. Ekkor az intézményen belüli kommunikációt lehet támogatni, javítani (akár az adatokat kezelő emberek között, akár az azonos entitástípust használó adatrendszerek közötti). Ha minden egyes intézmény kiépítené a saját lokális névtérét, a közös névtérhasználathoz még normalizálni kellene az egyes – addig különálló – névterek adatbázisait, és a későbbiekben is minden egyes névtérgazda belépése során el kéne végezni ezt az egységesítési munkát. A névterek bekerülési költségei tehát mind a nemzeti névterek, mind a globális névterekhez való csatlakozás esetében jelentkeznek. Ez egy további akadálya a névtérépítési munkáknak, de ezt azonnal ellentételezni lehet azzal, ha utalunk arra, hogy az előnyök minősége és nagyságrendje is együtt változik a névtérhasználó közösség körének bővülésével, és miatt érdemes ebbe a munkába belevágni. A **nemzeti névtér** egy adott ország, esetleg egy adott nyelvi közösség lokális névtéreit integráló közös névtér, míg a **globális névtér** valamely névtér típus esetén egy nemzetközi kooperációban működő, több nemzeti és/vagy lokális névtérgazda által fenntartott névtér jelent. Jelen írás a magyar nemzeti névterek lehetőségeit, a megoldandó feladatokat próbálja meg egy stratégiai anyagban összefoglalni. Mind a nemzeti, mind a globális névterek használatát úgy képzelhetjük csak el, hogy a névtérhasználat feltétele a lokális névtér létezése. Elvileg el lehetne képzelni olyan konstrukciót is, amelyben egy adatrendszer közvetlen és folyamatos kapcsolatban állva a nemzeti vagy globális névtérszolgáltatóval és a névadatokat folyamatosan a távoli rendszerből szedi ki, de legalább produkciós (és még más egyéb) okok miatt ez az elvi lehetőség nem látszik gyakorlatban sem megvalósíthatónak, sem megvalósítandónak.

Ha a közgyűjtemények számára fontos névtérépítkezés stratégiáját keressük, akkor érdemes minden fontos névtérrel számba venni még akkor is, ha tudjuk, hogy nem számíthatunk arra, hogy minden területen elinduljanak a munkálatok. Az összkép felfestése érdekében az alábbi ábrában igyekeztünk minden fontos szegmenst, potenciális fontos szereplőt, összetevőt, feladatot feltüntetni.



Az ábra értelmezése előtt újra hangsúlyozni kell, hogy a fenti képen minden fontos feladat fel van tüntetve, de ezek közül sokakat biztosan hátrébb kell majd sorolni, és csak később lehet belekezdeni a velük kapcsolatos munkákba. Vannak olyan névtér típusok, amelyekre vonatkozóan könnyebben megmondható, hogy kik lehetnek érintettek a fenntartásukban, kik lehetnek a főfelelősök az induló munkáknak, hol lehet esetleg már a jelenben nemzetközi partnert találni, és vannak olyan területek, ahol inkább csak a kérdéseket lehet feltenni, és a válaszokon még sokat kell gondolkodni. Az ábrán szerepelnek mindazon potenciális névterek, amelyek a közgyűjteményi gyakorlatban használhatók lennének. Egy nemzeti névtér felépítésére irányuló stratégiának figyelembe kell venni minden olyan dokumentumtípust, amely a nemzeti örökségünk fenntartásához hozzájárulhat. Ezen dokumentumtípusok nagyobbik részét nyilván hozzá lehet rendelni a hagyományos intézményi struktúrában működő, már létező szervezetekhez, de vannak olyan dokumentumok is, amelyekre nézve még nincsenek se gyűjteményezési elvárások, se ilyen gyakorlatok, miközben egyre nyilvánvalóbban látszik, hogy ezeket a magyar kulturális örökség körébe tartozónak kell minősítenünk. Az nyilvánvaló, hogy mind a hagyományos, mind az elektronikus könyvek gyűjtése a könyvtárak feladata, mint ahogy a filmek a Filmintézet (MNFA), a broadcast (tévé- és rádió) műsorszámok a NAVA gyűjtési körébe (felelősségébe) tartoznak. De mit kell kezdenünk az online videókakl, amelyek gyűjtésének láthatóan vannak csomópontjai a neten, de a nemzeti kultúra megőrzésével kapcsolatban szempontokról még semmit sem tudunk mondani.

A Filmintézetben elindult a MozgóKépTár (MKT) portál fejlesztése, amelynek egy olyan filmes adatbázist kell hozzáférhetővé tenni a neten, amelyen belül már létezik lokális névtér a filmekre, személyekre, testületekre, földrajzi helyekre vonatkozóan. Ezek közül a filmes lokális névtér "igényt

tarthatna" arra, hogy a filmes nemzeti névtér legyen egyben, a többi lokális névtér esetében pedig a vonatkozó nemzeti névtérfejlesztések indulásakor az MKT induló partner lehetne. A földrajzi névtérre vonatkozóan külön fejezetben fejtjük ki az elképzeléseinket. A személynévtér esetében nincs olyan közgyűjteményi szereplő, amelyekre azt mondhatnánk, hogy az intézmény személyadatbázisa (névtére) természetes módon adhatná a nemzeti névtér alapját. A legnagyobb névállománnyal nyilván az Országos Levéltár rendelkezik, de ez a névtérépítés szempontjából inkább hátránynak tekinthető. Van viszont több olyan globális névtér ezen a területen, amelyik "pályázhat" arra, hogy a nemzeti személynévtér számára "partnerül" válasszuk, a VIAF, illetve az ISNI. A projekt elején meg kell vizsgálni, hogy ez valóban így van-e, és ha igen, a két globális névtér közül melyiket érdemes választani, és a döntés után milyen módon lehet a kapcsolatrendszer kiépíteni a nemzeti és globális szint között. A területi névtér területén az első szakaszban csak annyit érdemes vállalni, hogy lokális névtér építését megkezdjük, a korlátozott erőforrások védelme érdekében ezen a területen nem érdemes várni a gyorsabb előrehaladásra. A könyvekre és periodikákra biztosan érdemes lenne egy-egy névtérépítési alprojektet elindítani, hiszen ezen dokumentumtípusok körébe tartozó dokumentumok számossága még elég kicsi ahhoz, hogy egy névtér-projekt végét látni lehessen, aminek sikerességéhez az is hozzájárulhat, hogy ezen dokumentumok feldolgozottsági szintje is magasnak mondható. Azt a kérdést meg kell vizsgálni, hogy egy könyv és periodikanévtér alapját milyen adatrendszerekből lenne érdemes kibányászni. A cikkek esetében első pillanatra nehezebbnek tűnik a névtér felállítása, de lehet, hogy alaposabb "helyismeret" alapján ez a tézis hamisnak kell majd tartanunk. Elképzelhető, hogy a cikknévteret valamilyen szempont alapján korlátozottan lenne érdemes elkezdni (például a tudományos cikkek gyűjtő MTMT mögé állítani egy ilyen névteret).

1.2 Általános modellezési kérdések

A világról való tudásunk kifejezésére fogalmakat teremtünk, amelyeket nyelvi konstrukciók segítségével kommunikálunk. A nyelvi megnyilatkozásaink során állításokat teszünk a világról. Természetesen tehetünk állításokat magukról a fogalmakról (nyelvi konstrukciókról) is. Ezek a sajátos metanyelvi állítások a fogalmak (konstrukciók) közti viszonyokra vonatkozhatnak. A fogalmaink, illetve nyelvi konstrukcióink közt sokféle kapcsolat létezhet, amelyek reprezentálására **tudásszervezési rendszereket** (Knowledge Organization Systems, KOS) képezhetünk. A tudásszervezési rendszerek fogalmakat rendeznek el valamilyen struktúrába. Típusai abban különböznek egymástól, hogy milyen relációkat (és ezzel milyen struktúrákat) engednek meg a rendszerbe szervezett fogalmak között. Minél több relációt vehetünk fel az elemek között, annál "bonyolultabb" struktúrához jutunk. A bonyolultabb struktúrának nagyobb a leíróereje (ez az előnye), viszont nehezebb mind a KOS építése, mind a KOS használata (ez a hátránya).

1.2.1 Többet a KOS-okról

Az archívumi munka során háromféle tudásszervezési rendszert használnak leginkább: **listákat**, mutatókat, indexeket (például egy archívum, egy könyvesbolt, egy videotéka által őrzött dokumentumok, könyvek, filmek listáját), **osztályozási, klasszifikációs rendszereket** (például az ETO-t) és **tezauruszokat** (például a Köztauruszt).³

A listák felsorolják valamely általános fogalom alá sorolható egyedi példányok (előfordulások) azonosítóit, neveit, címeit (személyneveket, könyvcímeiket, filmcímeiket, szervezetneveket stb.). Ebben az esetben a listába tartozó elemek között csak annyi a kapcsolat, hogy mind összetartoznak az egyetlen fogalmi dimenzióban. Mondhatjuk úgy is, hogy mindegyik elem ugyanabba a általános fogalomba,

³ Itt csak vázlatos áttekintést adunk a tudásszervezési rendszerekről, a tanulmány későbbi fejezetében a pontosabb leírásukat is megadjuk.

típusba, osztályba tartozik. Ekkor az egyedeket és az őket egybefogó általános fogalmat a példány/előfordulása/eleme relációval kötjük össze.

Az osztályozási rendszerek annyiban lépnek túl a listákon, hogy a kezelni kívánt partikuláris fogalmakat több fogalmi dimenzióban rendezik el. Ennek a struktúrának már két alaprelációja van. Az elrendezésre váró egyedeket mindig a példány relációval fogjuk egybe valamely általános fogalomhoz kötődően, de itt már több általános fogalmat használunk, amelyek között egy új relációtípussal, a generikus relációval teremtünk kapcsolatot. Minden általános fogalomra megadhatjuk, hogy mely egyedfogalmak tartoznak a terjedelmébe, és minden általános fogalomhoz hozzárendelhetünk egy generikusan fölérendelt másik általános fogalmat. A két – egymástól nagyon különböző – alapreláció léte megadja annak lehetőségét is, hogy a generikus relációval összekapcsolt általános fogalmakat önmagukban, az alájuk tartozó partikuláris példányok nélkül kezeljük. Egy klasszifikációs rendszer lényegét a generikus hierarchiába szervezett fogalmak jelentik, de sosem szabad elfelejteni, hogy milyen célokra kell/lehet használunk az ilyen tudásszervezési rendszereket: mindig valamilyen partikuslárís jelenségegyüttes, dokumentumok bizonyos konkrét halmazának tartalmi leírására, amikor pedig minden esetben a példány relációt kell alkalmaznunk.

A teauruszok összetettebbek a klasszifikációkhoz, hiszen a "sima" osztályba sorolási (valamint a példány) reláción túl további – pontosan definiált – relációkat is tartalmaznak. A **generikus alá- és fölérendeltségi relációpár** mellett (ami tehát az osztályozási tevékenység alapját jelenti) a teaurusz relációja még a **partitív alá- és fölérendeltje relációkettős**, az **előzménye, oka/"utózmánya"**, **következménye** pár, valamint a mindezeket túli kapcsolatokat jelezni képes **rokonsága reláció**.

A teaurusz tehát tartalmaz legalább egy osztályozási rendszert is (amilyen mértékben a generikus relációk szerepet kapnak a teauruszon belül), de a teauruszon belül további részstruktúrákat írhatunk le a partitív relációkettős, illetve a temporális/kauzatív szemantikai relációpár segítségével. A teaurusz nagyobb leíró ereje abból a tényből fakad, hogy a feldolgozandó ismeretterület dokumentumpéldányait többféle – jól-rendeztett – fogalmi struktúra segítségével jellemezhetjük. Van még egy további előnye a teauruszoknak: magasabb szinten kezelhetjük velük a nyelvhasználatunkban tetten érhető terminológiai többértelműség jelenségét.

Mondhatjuk, hogy fenti szemantikai relációk a fogalmak között léteznek, hiszen bármely nyelven, bármilyen nyelvi konstrukcióval fejezzük is ki őket, szemantikailag mindig ugyanahhoz az eredményhez jutunk. A "kutya-állat", "eb-állat", "dog-animal", a nyelvi szinten létező, különböző nyelvi konstrukcióban kifejezett, három állítás egyetlen generikus relációpár segítségével reprezentálható a fogalmi szinten. A nyelv rugalmassága miatt ugyanazt a fogalmat többféle nyelvi konstrukcióval is megragadhatjuk, és hogy ez a szinonimitás ne legyen zavaró a fogalmak közti kapcsolatok leképezésekor, a teauruszok (illetve általában a tudásszervezési rendszerek) elkülönítik egymástól a nyelvi konstrukciók két típusát. Az ugyanazon fogalmat jelölő kifejezések közül kiemelnek egyet, amit **preferált terminusnak** neveznek el, míg a fogalomhoz tartozó többi nyelvi konstrukciót **utalószóként** minősítve összekapcsolják a preferált terminussal egy **lásd** relációval (aminek a **helyette** reláció lesz az inverze). A nyelvi konstrukciók ilyen felosztása arra jó, hogy a preferált terminusok segítségével a fogalmak és a nyelvi kifejezések között egyértelmű (1:1-es) kapcsolatot lehessen teremteni egyfelől, illetve az utalószavak révén növelni lehessen a tudásszervezési rendszer kifejezőerejét másfelől. Ezek a relációk más minőségűek a korábbi szemantikai kapcsolatokhoz képest, hiszen amíg a **lásd** és a **helyette** relációk a nyelvi konstrukciók, addig a generikus, a partitív, a temporális/kauzatív reláció a fogalmak között létezik (ezen nem változtat semmit az a tény, hogy ezeket a relációkat a preferált terminusok segítségével fejezzük ki).

A teaurusz jóval nagyobb kifejezőereje – a többi tudásszervezési rendszerhez képest – a gazdagabb relációstruktúrában rejlik. A nehezkesebb kezelésmód miatt további relációkat már nem vettek fel a teauruszokba (illetve az ezeket definiáló szabványokba). Ennek viszont az lett a következménye, hogy bizonyos kapcsolatokat szemantikailag nem megalapozott módon lehetett csak a létező teauruszrelációkkal leírni. Mindez persze kisebb jelentőségű, megszokott, így vállalható hiba volt. Az

osztályozási rendszerek (mint például az ETO) ennél még nagyobb pongyolást követtek el folyamatosan azzal, hogy egybemoszták a hierarchikus jellegű (generikus, partitív, időbeli) relációkat egymással az egyetlen klasszifikációs (igazából csak valamilyen általános alá-fölrendelést kifejező) reláció alkalmazásával.

A klasszifikációs rendszerekhez hasonlóan a teauruszokra is igaz, hogy az általános fogalmakból álló részük önmagában is értelmes, önálló rendszert alkot, sőt, teaurusz fogalma alatt inkább az általános fogalmi részt értik – a teaurusz elemeivel jellemzett partikuláris fogalmak nélkül. A 'példánya' reláció ugyan benne van a teauruszt leíró szabványokban, de a gyakorlatban ritkábban használják, sőt, az is előfordul, hogy a generikus relációt és a példánya relációt egynek tekintik. Márpedig a névterek "kiszolgálására" olyan fogalmi keretre van szükség, amely a névterek elemeit, tehát a példánya relációt is jól kezeli. Olyan névterekről beszélünk ugyanis, amelyek tulajdonnévterek. Ha szemantikailag megalapozott tulajdonnévtereket akarunk építeni, akkor érdemes felkészíteni a rendszerünket arra, hogy a partikularitás minden problémáját kezelni tudjuk. A kérdés az, hogy milyen elvárásokat kell megfogalmaznunk (és kielégítenünk) az egyedi fogalmak kezelésével kapcsolatban.

A nyelvi konstrukciók világában megkülönböztetünk közneveket és tulajdonneveket egymástól azon az alapon, hogy milyen a kifejezésekkel jelölt fogalmak extenziója. Ha a fogalommal/kifejezéssel – legalább potenciálisan – a dolgok egy halmazát jelöljük, akkor beszélünk köznévről, ha a fogalommal/kifejezéssel egyetlen dologra referálunk, akkor van szó tulajdonnévről. A lényeg: a tulajdonnévvel (vagy az egyedi fogalommal) mindig valaminek a példányát, valamilyen instanciát ragadhatunk meg, a köznévvvel (vagy a generikus fogalommal) pedig a létező dolgok, instanciák valamely halmazát jelöljük valamilyen közös tulajdonságuk révén összekapcsolva őket. A kétféle fogalom közti alapvető különbség azon a ponton látszik legszemléletesebben, amikor megpróbálunk új fogalmat alkotni a már létező fogalmainkra támaszkodva. Egy generikus fogalomból – elvileg - mindig létrehozhatunk egy új generikus fogalmat úgy, hogy egy újabb tulajdonságot rendelünk hozzá, amivel potenciálisan tovább szűkítjük ugyan az új fogalom/kifejezés terjedelmébe tartozó előfordulások listáját, de még mindig a dolgok egy halmazára referálunk. Ezzel szemben az egyedi fogalom/kifejezés alá már nem képezhetünk egy újabb fogalmat/kifejezést. Ha van egy 'város' fogalmunk, akkor képezhetünk olyan új fogalmakat, mint 'nagyváros' vagy 'főváros', amikor is az új fogalmak terjedelme kisebb lesz, mint az eredeti fogalomé, de az új fogalmak alá is több példányt berendelhetünk. Ha viszont azt mondjuk házak, ingatlanok és más műtárgyak egy adott halmazára, hogy az 'Budapest', akkor az így értelmezett egyedi fogalom/kifejezés alá már nem rendelhetünk semmit. Az már egy másik kérdés, hogy ebben az esetben is létrehozhatunk új fogalmakat/kifejezéseket ('Újpest', 'XIII. kerület', 'Nagykörút'), és ezeket egymás alá vagy fölé rendelhetjük, de ilyenkor nem a generikus, hanem a partitív reláció mentén tesszük (tehetjük) meg mindezt.

A fenti problémára a jó megoldást, a megfelelő magyarázatot akkor találhatjuk meg, ha figyelembe vesszük, hogy itt a generikus alá-, fölrendeltje és a példánya/instanciája reláció különbségéről van szó. Amikor egy generikus fogalmat képezünk, akkor a generikus relációt használjuk erre, ami másodrendű reláció, de csak abban segíthet nekünk, hogy jobban megértsük majd a generikus és partitív reláció közti különbséget. Itt fontosabb az a tény, hogy a generikus reláció tranzitív, hiszen ez a minőség biztosítja minden osztályozási művelet talán legmeghatározóbb vonását, a tulajdonságok öröklődését az egymásba ágyazott osztályok/fogalmak mentén. A partitív reláció is tranzitív, viszont ez a fogalom elsőrendű kategória, és a pontos meghatározására, a típusainak elkülönítésére egzakt mereológiai elméletek, pontos logikai definíciók vonatkoznak (Szakadát 2005). Amit a filozófia partikulárisnak nevez, azt más ismeretterületeken a **példány**, az **előfordulás**, az **eset**, az **instancia** vagy az **elem** terminusaival ragadják meg. Ezek alapvető különbsége az általános fogalmakhoz (osztályokhoz, típusokhoz) képest az egyediségüket jelentő, jelző, biztosító tér-idő vonatkozásuk. Az általános, generikus fogalmakhoz nem tudunk, nem lehet térbeli és időbeli minőségeket kapcsolni, az individuumokra vonatkozó egyedi fogalmakhoz viszont nem csak tudunk és lehet, de kell is ilyen kapcsolatokat rendelni, ha egyértelműen meg akarjuk ragadni őket. Egy emberre, egy szervezetre, egy településre, egy épületre akkor tudunk egyértelműen referálni, ha megadjuk a téridő-koordinátáit (az a kérdés, hogy az egyediséget minden esetben vagy csak az esetek többségében lehet ezáltal biztosítani,

az itt tárgyaltak szempontjából nem tűnik lényegesnek). A kérdés, vajon miként tehetjük ezt meg. Sajnos az ismert tudásszervezési rendszerek (a klasszifikációk, a tezauszok) eszköztára ezt nem igazán teszi lehetővé. Ha olyan KOS-infrastruktúrát akarunk felépíteni és működtetni, amely képes lehet kiszolgálni ezt az igényt, akkor ennek a rendszernek a mélyén ott lehetnek a tezausz komponensei is, de szükség van további relációk felvételére. Mielőtt kifejtjénénk, milyen módon lehet ezt a bővítést elvégezni, szükség van arra, hogy a tulajdonnévterek építéséhez elengedhetetlen példánya reláció sajátos minőségére rávilágítsunk.

Amikor létre akarunk hozni egy új partikuláris fogalmat, azt ugyanúgy hozzá kell kötnünk egy generikus fogalomhoz, mint ahogy az új generikus fogalom definiálásakor is ezt kell tennünk (mindig meg kell adnunk egy partikuláris fogalom típusát). Van azonban egy nagyon fontos különbség a kétféle alárendelés között: az egyedi fogalmak/tulajdonevek létrehozásakor nem a generikus alárendeltje, hanem az instanciája/példánya reláció segítségével tehetjük meg mindezt. Ez azért lényeges, mert ez a reláció nem tranzitív (szemben a generikus vagy a partitív relációval), sőt, az instanciája reláció intranzitív. Ez az intranzitivitás részben (de lehet, hogy teljes egészében) annak köszönhető, hogy az instanciája reláció heterogén, hiszen egy osztályfogalmat köt össze egy individuumfogalommal. Ezzel szemben a generikus (és partitív) reláció homogén, hiszen azonos minőségű relátumokat köt össze egymással. Ez a heterogén minőség magyarázza meg azt, miért nem tudunk egy individuumot jelző tulajdonnév alá egy másik individuumra utaló tulajdonnevet rendelni.

1.2.2 Névterek modellezése

Ahhoz, hogy a névterek modellezésének legfontosabb kiinduló tézisének rögzíthessük, előbb meg kell válaszolnunk azt a kérdést, hogy mit is kezelünk a névterek segítségével: **neveket** vagy **névfordozókat** (vagy mindkettőt). A válasz az, hogy elvileg és értelme szerint névfordozókat (személyeket, testületeket, földrajzi helyeket stb.), viszont a gyakorlatban mindkettőt vagyis névfordozókat és neveket egyszerre. A tulajdonnevek használatának igazi célja mindig az, hogy velük egyértelműen rá tudjunk mutatni valamely egyedi dologra (névfordozóra). Általános tapasztalat azonban az, hogy a nevek használata nem biztosít feltétlen egyértelmű azonosíthatóságot. Ez lehet a magyarázata annak is, hogy bármennyire egyszerű, nyilvánvalónak tűnő a fenti válasz, a tényleges névterek felépítésekor többféleképpen emelték be ezt az elvet saját modelljükbe. A Getty vagy a GeoNames a földrajzi helyeknek oszt ki azonosítót, és az eltérő névalakokat mind ezekhez rendeli (tehát a névváltozatok nem kapnak egyedi azonosítót). A KSH Helységnévtárában – ezzel szemben – a nevekhez rendelnek azonosítókat (noha a rendszer lényege szerint a névfordozókra, konkrétan a településekre koncentrálnak). Van tehát két entitástípusunk, a nevek és a névfordozók. Ha mind a névfordozókat, mind a neveket kezelni szeretnénk a névterünkben, márpedig erre szükség van, akkor:

mindkét entitástípus (a névfordozók és a nevek) előfordulásait egyedileg azonosítani kell.

Az egyedi azonosítást számítógépek segítségével könnyen megtehetjük, ha a gép által kiosztott egyedi azonosítókat rendelünk valamely entitástípus minden konkrét előfordulásához. A nevek és névfordozók azonosításához azonban nem elég gépi azonosítókat kiosztani, mert az ember számára ez nem ad elég információt a dolgok tényleges azonosítására. Szükség van ezen egyedi dolgok **lokalizálására** is - a térben és az időben. Az individuumok mindig a térben és időben léteznek, így a parikularitás, az individuális minőségét megragadhatjuk ezzel a lokalizálással, az individuális nevek és névfordozók időhöz és térhez kötésével. Az, hogy az individuumok térben és időben kötötten léteznek, azt jelenti, hogy azzal írhatjuk le őket, ha megadjuk az időhöz való viszonyukat (létezésük kezdő- és záróidejét), valamint el kell helyeznünk a térben is őket (vagyis meg kell adnunk a geometriájukat a földrajzi térhez rendeltén). A következő fontos modellezési kérdés az, hogy mire kell vonatkoztatni és hogyan kell megvalósítani ezt a kettős lokalizációs igényt. A következőkben a geoindividuumokra vagyis a földrajzi helyekre vonatkozóan válaszoljuk meg ezt a kérdést.

1.3 Földrajzi névtér

A földrajzi neveknek vannak olyan jellegzetességei, amelyek más névtípusokhoz képest egyedinek mondhatóak, ráadásul a névtérépítkezés szempontjából is meghatározó jelentőséggel bírnak. Ezeket a sajátosságokat figyelembe kell venni a földrajzi névtér modelljének kialakításakor. A földrajzi névtérnek a modellezésbeli különbségeken túl van még egy további, számossággal kapcsolatos fontos specialitása is, ami érvként használható amellet, hogy a földrajzi névtérrel kezjük el az építkezést. A földrajzi helyek, nevek felülről erősen korlátosak, számosságuk nem nő lineárisan együtt az idővel, míg a többi tulajdonnév esetében inkább az a jellemző, hogy idő haladtával folyamatosan nő az adott névtér tartománya. Mindig megjelennek új ágensek, mindig keletkeznek új dokumentumok, mindig lesznek új események, de földrajzi helyből, földrajzi névből ez nem így van. A földrajzi nevek még csak-csak változnak, de a földrajzi helyek már sokkal kevésbé.

1.3.1 A földrajzi entitások modelljének kérdései

A földrajzi entitások különlegessége, hogy a földrajzi térben léteznek, így egyfelől valamilyen geometriai tulajdonsághalmazzal (röviden: geometriával) jellemezhetők, másfelől mindig elhelyezhetők, lokalizálhatók a földrajzi tér egészén belül (hogy mennyire pontosan és mennyire egyértelműen, az már egy másik kérdés). Mondhatjuk, hogy a földrajzi helyeknek mindig létezik valamilyen geometriája, de ezt az állítást még pontosítanunk kell. Nagy kérdés ugyanis, hogy mit és mi alapján tekinthetünk földrajzi entitásnak. A válaszhoz egy nagyon fontos filozófiai fogalomkettőt kell beemelnünk az értelmezési keretünkbe.

A filozófusok elkülönítik egymástól a dolgok (azon belül a földrajzi entitások) kétféle típusát, amikor azt mondják, hogy beszélhetünk **fiat** és **bona fide** dolgokról (Smith, Varzi 2000). A fiat entitások az emberi akarat termékei, valamilyen konstituálási aktus eredményeként léteznek, míg a bona fide entitások létezése, létrejötte az emberi akarattól független. Amikor azt mondjuk a tér egy darabjára, hogy ez az utca, a mellette levő másik rész pedig a járda, akkor mindkét dolog csak azért létezik, mert így döntöttünk, és fizikailag-földrajzilag korábban is létező földdarabkára mutatva létrehoztuk az adott konkrét utcát és adott konkrét járdát. Amikor egy városhoz csatolják a szomszédfalut, amivel egyrészt megnő a város területe, másrészt a szomszédfalu megszűnik önálló településként és a város egy városrészévé válik, akkor is arról az emberi konstitutív aktusról van szó, amely újradefiniálja az addig létező települések közigazgatási státusát és területeit. Ezek mind fiat objektumok. Ezzel szemben egy folyó, egy hegy, egy tó akkor is ott van a földrajzi térben, és akkor is folyó, hegy vagy tó marad, ha mi, emberek, ezt nem deklaráljuk, nem konstituáljuk, nem csinálunk semmit.

Más szakterületen ez a kettősség a **természeti tények**, illetve **társadalmi tények** szembeállításában jelenik meg (Searle 1995)

A kétféle entitás megkülönböztetés után (és alapján) már rögzíthetjük, hogy mit és hogyan kell figyelembe vennünk a földrajzi entitások leírásakor. Mindkét objektumtípus előfordulhat a földrajzi objektumok között, tehát beszélhetünk fiat és bona fide földrajzi entitásokról (a fiat objektumokra példa az ország, város, tér, utca, a bona fide objektumokra a folyó, hegy, medence, tó, óceán fogalma). Modellezési szempontból a fiat objektumok leképezése a nehezebb, hiszen ezek leírásához, értelmezéséhez több ismeretre van szükségünk, így ezekre érdemes fókuszálni a földrajzi objektumok modellezése során. A földrajzi entitások legfontosabb jellemzőit – konstitutív jellegük miatt – mindig az időben elhelyezve, időbeli paraméterekkel ellátva kell leírnunk.

A földrajzi entitások tulajdonságainak mindig van időbeli kezdő- és végpontja.

Az az elvárás, hogy minden fontos tulajdonság leírásába be kell venni a kezdő- és záróidőjelzést, nem jelenti azt, hogy mindig meg tudnánk adni pontosan ezeket az időbeli értékeket. Egyrészt nem is lehet

minden esetben egyetlen időpontot kijelölni az időben, ami valamilyen tulajdonság pontos kezdetét vagy végét jelölné, vagy azért nem, mert nem tudjuk a pontos értéket, vagy azért nem, mert nem köthető egyetlen időponthoz a szóbanforgó esemény. Másrészt ha tudunk is valamit az adott eseményről, akkor is kérdés még, hogy milyen pontosságú ez az ismeretünk. A modellünkbe be kell vezetni az **időpontosság** fogalmát, amelynek segítségével mindig jelezni tudjuk, hogy valamely időbeli adat milyen pontossággal rendelkezik (perc-, óra-, nap-, hónap-, évponos stb.).

A földrajzi hely legfontosabb paramétere a **geometriája**, aminek mindig van típusa (pont-, vonal- vagy poligonszerű), és ez a geometria változhat az időben (például egy település határvonala változik). A földrajzi entitásoknak mindig van egy társadalmi, intézményi, közigazgatási **státusminősége**, ami megintcsak változhat az időben. Gondoljunk arra a példára, amikor egy kisebb települést hozzácsatolnak egy nagyobbhoz. Ilyen esetben az az intézményi besorolás változik meg a csatolt település esetében, ami a falu közigazgatási önállóságára vonatkozik: megszüntetik önálló települési jogállását, és a másik település városrésze lesz. Eközben a "befogadó" város státusa, besorolása nem változik. Az ilyen besorolási, osztályozási konstitutív aktusok körébe tartozik még valamely település más járásba, más megyébe, más országba való besorolása. A földrajzi objektumok harmadik jellemzője az azonosítására használt megnevezés, a **földrajzi név**. A nevek is konstitutív aktusok eredményeként jönnek létre, tehát ezek leírásakor is időbeli paramétereket kell felvennünk. Abból a tényből, hogy mindhárom alaptulajdonság esetében meg kell engednünk azt, hogy azok időben változzanak, következik az a modellezési elvárás, hogy mindhárom tulajdonságtípus esetében a konkrét tulajdonságértékeket úgy kell a földrajzi entitások előfordulásaihoz rendelni, hogy meg kell engednünk az 1:N-es kapcsolatot köztük.

A földrajzi entitások három alaptulajdonságát, a geometriát, a státust és a nevet időben változó módon, 1:N-es kapcsolattípus mentén kell felvennünk.

Egy földrajzi entitás leírásához mindig meg kell tudnunk adni a három alaptulajdonságának aktuális értékét. A földrajzi helyek nevét tíválisan meg tudjuk adni, hiszen minden diskurzusban így hivatkozunk rájuk.

1.3.2 A földrajzi névtér modellje

1.3.2.1 Geoindividuumok azonosítása

A földrajzi helyek, a geoindividuumok modellezéséhez figyelembe kell venni azt a fogalmi kettősséget, ami a **társadalmi és természeti tények** elkülönítése révén lehet megragadni. Ezt a szembeállítást a földrajzi objektumok esetében a **fiat** vs. **bona fide** fogalmak segítségével fejezik ki. A bona fide objektumok fogalmával olyan geoentitásokat ragadhatunk meg, amelyeknek határai nem emberi konstituálás eredményeként léteznek, így tehát a velük kapcsolatos adatokat természeti tényként kezelhetjük. A folyók, hegyek, dombok vagy az ember által épített objektumok (házak, templomok, vasútállomások) a példák erre. Ezzel szemben a fiat objektumok alapvonása az, hogy az objektumok határai ember által konstituáltak, ahogy így van ez a települések, a járások, az országok esetében. A geoentitások másik két fontos tulajdonsága, az entitások nevei, illetve az entitások típusbesorolásai esetében nem létezik ez a kettősség, ezek a minősítések emberi konstituálás eredményeként jönnek létre. Ebből a tényből fakad az a modellezési elvárás, hogy a konstitutív adatokhoz időbeli minőséget kell rendelnünk, illetve meg kell engednünk azt is, hogy ugyanahhoz a geoindividuumhoz időben változó módon, többféle minősítést lehessen hozzárendelni.

A geoindividuum típusát a nevekhez kapcsolódóan lehet származtatni, ezek az adatok is ugyanazokból a forrásokból szerezhetők be, mint a nevek. A geometriai adatok beemelése sokáig nehézkes volt, térképeken keresztül volt megvalósítható. A térinformatikai eszközök elterjedésével azonban ma már

ezen adatrétteg becsatlakoztatása is megoldható, noha ezen a ponton még meg kell oldani a geometriai adatok időbeli változásának kezelhetőségét.

1.3.2.2 Idő- és eseménykezelés

Az időkezelés modellezési kérdéseinek áttekintését kezdhetjük azzal a tétellel, amit már kimondtunk korábban is, de nem lehet elégszer hangsúlyozni.

A földrajzi entitások tulajdonságainak, kapcsolatainak mindig van időbeli kezdő- és végpontja.

Az az elvárás, hogy minden fontos tulajdonság leírásába be kell venni a kezdő- és záróidőjelzést, nem jelenti azt, hogy mindig meg tudnánk adni pontosan ezeket az időbeli értékeket. Van, hogy napra pontosan tudunk valamilyen esemény kezdetéről, végétől, de előfordulhat az is, hogy csak az esemény évét, esetleg csak az évtizedét (évszázadát) ismerjük. Ennek kezelésére a modellünkbe

be kell vezetni az időpontosság fogalmát, amelynek segítségével mindig jelezni tudjuk, hogy valamilyen időbeli adat milyen pontossággal rendelkezik.

Az időpontosság fogalmával azt fejezhetjük ki, hogy az adott esemény dátumát milyen pontosan ismerjük. Ez lehet perc-, óra-, nap-, hónap-, év-, évtized-, évszázad- vagy akár csak évezredpontos. Ezt az ismeretet úgy kezelhetjük, hogy minden eseményhez felvesszünk egy dátumot (amit a modellben dátum típusú értéként definiálhatunk). Amennyiben napra-pontos információval rendelkezünk az eseményre vonatkozóan, akkor beírhatjuk azt a (dátum)értéket. Ha viszont pontatlanabb ismeretünk van csak, akkor beírhatunk egy "szabályos" dátumértéket, és egy dátumpontosság mezőben jelezhetjük, hogy mennyire pontos az időmegjelölés (és a dátum-értékből csak az így definált "értékes" részt, az évet, az évtizedet, az évszázadot vesszük figyelembe). Példa: ha csak annyit tudunk, hogy valamikor a XVI. században történt valami, akkor dátumként felvesszük az '1500-01-01' értéket, és a dátumpontosság mezőben jelezzük, hogy az adatunk évszázadpontos, ha napra/óra/percre pontosan tudjuk az esemény idejét, akkor beírjuk azt a dátum mezőbe és a dátumpontosságként a nap/óra/percpontos minősítést tesszük el.

A névterekben (különösen a földrajzi névtérben) előforduló, dátum-jellegű információ kezelésével kapcsolatban érvényesíthetünk egy további elvet:

az időkezelést nem a kapcsolatokhoz rendelt dátumadatok, hanem események felvételével érdemes megvalósítani.

Az időkezelés "hagyományos" megoldása azt jelentené, hogy a különböző események (névváltozás, státusváltás, geometriai változás) időadatait beírjuk az adott entitás rekordjának megfelelő – dátum típusú – mezőjébe (és egy másik mezőben felvesszük a dátum pontosságát jelző időpontosság értékét is). Az ezzel szemben javasolt megoldás szerint az események adatait külön táblában tároljuk, az eseménypéldányokat azonosítókkal látjuk el, és az eseményekre mindig ezeken az eseményazonosítókon keresztül hivatkozunk a többi entitás táblájában. Így az eseményekhez kapcsolható időadatot egy helyen tároljuk, de nem ez az igazi előnye, értelme ennek a megoldásnak. Fontosabb előny az, hogy ezzel a technikával kezelni tudjuk az "összetett" események problémáját. Egy esemény ugyanis sok esetben több – egymással kapcsolatban levő – részeseményből "áll össze", és ezt az összetartozást valahogyan jelölni érdemes. A legnyilvánvalóbb példa erre az, amikor megváltozik

egy település neve. Ebben az esetben két individuumról (két névről) van szó, amelyekre vonatkozóan külön rekordban kell az adatokat tárolnunk. Ugyanazt a dátumot tehát két helyre is be kell írni. A normalizálás követelménye itt azt kívánja meg, hogy felvegyünk egy eseménypéldányt az események közé (amelynek adhatunk időértéket, pontosságot), és a két névrekord adatai közé minden esetben ugyanazon esemény egyedi azonosítójával hivatkozzunk. Ennek a megoldásnak további előnye még, hogy nem kell felvenni egy újabb táblában azt az információt, amely összekapcsolja a két eseményt (az egyik név "megszűntét" és a másik név "keletkezését") egymással. Még egyértelműbben látszik a megoldás előnye, ha egy olyan esetre gondolunk, amikor egy városban egy időben több falut is beolvasztanak. Ekkor még több – ugyanazzal a dátummal rendelkező – részeseményt kellene felvennünk a nevek, geometriák, státuszok változásait leíró rekordokba, ami helyett előnyösebb és egyértelműbb egyetlen eseményt felvenni az esemény táblába, és erre hivatkozni a különböző típusú változásokat jelző mezőkben.

A fentiek alapján a változások kezelésére azt az egységes megoldást alkalmazhatjuk, hogy a geoindividuumok partikuláris (időhöz kötött) jellegét úgy fejezzük ki, hogy a kezdő- és záróesemény azonosítóit hivatkozzuk meg a geoindividuumok tábláiban, és az eseményeket egységesen egy esemény táblában kezeljük.

1.3.2.3 Geometriakezelés

A földrajzi névtér geoindividuumait meg lehet jeleníteni egy digitális térképen, amennyiben az adott geoindividuumnak van digitális reprezentációja. Ezt a reprezentációt az OpenStreetMap (OSM) közösséggel együttműködve érdemes felépíteni, de a geometriai adatok birtokában ki lehet (és ki is érdemes) építeni a kapcsolatot a Google Maps felé is. Utóbbi magasabb színvonalú szolgáltatást kínál, viszont zárt (pontosabban nem teljesen nyitott), ami nem teszi lehetővé a teljesen nyílt, kétirányú kooperációt. Az OSM-kapcsolaton belül két kulcskérdés van. Az egyik fontos feladat: meg kell oldani mindkét irányban a **szinkronizáció** lehetőségét. Mindkét irányban tudni és jelezni kell, hogy változtak az adatok és ezeket valahogyan át kell vezetni a másik rendszer irányába. A másik feladat az OSM jelenidejűségéből adódik: meg kell oldani a múltbeli (történeti) adatok kezelését. Itt nyilván elsősorban a geometriai adatok történetiségéről van szó. Ennek az lehet a módja, hogy az OSM létező rendszere mellett fel kell építenünk egy történeti OSM-t (HOSM – Historical OSM) a földrajzi névtér részeként. Ez a rendszer az eredeti OSM egyfajta részleges - klónjaként is értelmezhető, hiszen infrastrukturálisan mindenben támaszkodhat (támaszkodnia kell) a futó - kortárs - OSM-rendszer adataira.

1.3.2.4 Névkezelés

Az individuumokat nevekké azonosítjuk, de ügyelni kell arra is, hogy a neveknek mindkét típusát kezelni tudjunk. A hagyományos nevek mellett ugyanis az individuumok azonosítására számokkal kifejezett egyedi azonosítókat is alkalmazunk, és ezeket ugyanúgy a nevek egyik típusaként foghatjuk fel, mint bármely más, természetes nyelven kifejezett, "karakter-alapú" nevet. Az ilyen "mesterséges azonosítók" funkciója ugyanaz, mint a természetes nevéké: egyértelműen azonosítani és ezáltal hivatkozhatóvá tenni egy névhordozó individuumot. A különbség csak annyi köztük, hogy amíg az ember a természetes nyelvi kifejezéseket kezeli könnyebben, addig a számítógép számára a számok feldolgozása egyszerűbb. Modelllezési szempontból tehát lehetne névként értelmezni az azonosítókat, praktikus érvek miatt azonban érdemesebbnek tűnik külön tartani őket (részben a szokáskényszer miatt, részben a karakterek és a számok eltérő adattípusait így saját logikájuk szerint lehet tárolni). Kell viszont olyan keresési opciót is biztosítani, amely a "kétféle névadatban" egyszerre képes keresni (ehhez csak annyit kell tenni, hogy egy join művelettel egyesíteni kell a neveket és azonosítókat tartalmazó táblák tartalmát). Konkrét példaként hivatkozhatunk itt az irányítószámok (irányítószám-körzetek), a települések KSH-kódjai, a GeoNames- vagy a Getty-kódok szerinti kereshetőségre.

A nevek esetében is érvényesíteni kell azt az elvet, hogy a nevek partikuláris jellegét úgy ragadhatjuk meg az adatbázisban, hogy az geoindividuumokhoz 1:N-es relációban kapcsolt neveket (azonosítókat)

kezdő- és záródátummal (eseménnyel) látjuk el. Természetesen itt is szükség van az időpontosság jelzésére.

A névkezelés további kérdése az, hogy kell-e **kitüntetett névalak**, és ha igen, hogyan határozzuk meg azt. Az archívumi gyakorlatban sok esetben használnak kitüntetett névalakokat, és névtér működtetése során igazodni kell ehhez a szokáshoz, igényhez. Mivel azonban a preferált névalak konkrét értéke erősen kontextusfüggő, ezt a funkcionalitást is kontextusfüggő módon kell biztosítani. Ez annyit jelent, hogy biztosítani kell, hogy a preferált névalak a használati esettől, a kontextustól függő módon lehessen más és más. Ha ezt így tesszük, akkor a kollaborációban résztvevő intézmények saját preferált névalakot jelölhetnek ki maguknak. Természetesen ebben az esetben is szükség lesz arra, hogy a nemzeti névtéren belül is legyen egy alapértelmezett, preferált névalak, hiszen itt is szükséges – egyértelmű és egy-elemű – listákat kezelni, de ezt nem kell az együttműködő partnereknek átvenniük, ha nem akarják.

1.3.2.5 Relációk

A geoindividuumok esetében a neveket, a társadalmi státusok és a geometriák konkrét előfordulásait 1:N-es kapcsolatban rendeljük a geoindividuumokhoz. Mindhárom reláció esetében fel kell venni a sémánkba a három kapcsolat időbeli lokalizálását biztosító mezőket. A relációk megengedik, hogy egy geoindividuum mindhárom relációban több "értéket" is felvegyen – akár ugyanabban az időben, hiszen egyszerre lehet több neve vagy több típusbesorolása egy konkrét geoindividuumnak. Elvileg nem lehetne megengedni, hogy egy időben több geometriája is legyen egy geoindividuumnak, de gyakorlati szempontból érdemes ezt mégis megengedni azért, hogy a többféle forrásból származó, olykor eltérő adatokat is meg tudjuk őrizni.

Az individuumok és a nevek között valamilyen referencia jellegű kapcsolatot, a típusba sorolások esetében generikus alárendeltje relációt (is a subclass of, is a type of, is a sort, is a category, is a kind of), az entitások és előfordulásai között a példánya (is an instance of, is an occurrence of) relációkat kell értelmeznünk. A geoindividuumok közti partitív alá-, fölérendelési relációkat külön táblában érdemes kezelni, és itt biztosítani kell, hogy egy időben többféle partitív besorolást is el lehessen végezni az individuumok között. A geoindividuumok világában bizonyos relációk mentén lehetnek **partíciók**, de nem minden partitív felosztás jelent egyben tiszta és szabályos partícióra bontást. A megyék, a járások, a települések szabályos partíciót alkotnak (bár a fővárost egyfajta kivételkezeléssel lehet csak beilleszteni megyei, járási partíciók logikájába), de a településrészek, tájegységek, természetföldrajzi geoindividuumok már nem minősíthetők partícióknak. A névterek szolgáltatásakor, a névtérben való keresések biztosításakor nagyon fontos szempont kell legyen, hogy a felhasználók egyértelműen láthassák, kereshessék a különböző logikájú partitív kapcsolatrendszereket, partíciókat. Ha megengedhető, hogy a partitív (vagy a generikus) reláció mentén többféle besorolás is lehetséges legyen, akkor azt is biztosítani kell, hogy a különböző szempontú, különböző típusú alárendeléseket el tudjuk különíteni egymástól. Ezt megtehetnénk úgy, hogy az adott reláció alá rendelve altípusokat definiálunk, de ez talán fölösleges (bár praktikus könnyebben kezelhető lenne sok minden). Az is megoldást jelenthet, ha pontos megkülönböztetést teszünk a kezelt entitások között. Ekkor a relációkat nem kell annak megfelelően megtöbbszörözni, hogy milyen típusú entitások között érvényesülnek. A példa kedvéért: ha tudjuk, hogy két geoindividuum között partitív alá-, fölérendeltség létezik (például Szentendre része Pest megyének, illetve a Duna egy része határos Szentendrével), akkor az összekapcsolt geoindividuumok típusából fakadóan tudhatjuk, hogy más partitív felosztási logikába tartozik a két eset). Vannak helyzetek azonban, amikor igenis érdemes, sőt kötelező lehet, hogy a partitív relációból többet definiáljunk. Akkor van erre szükség, amikor partitív relációt akarunk felvenni konkrét földrajzi individuumok között egyfelől, illetve a partikuláris geoentitásokat tipizáló általános fogalmak (osztályok, típusok) között másfelől. Mindkét szándék értelmes, de más relációt kell érvényesítenünk az általános fogalmak, illetve individuumok között.

A tulajdonnévterek legfontosabb relációja a példánya reláció, hiszen a névtérben valamely entitáshoz kapcsolható elemeket, konkrét névhordozókat akarunk egyértelműen azonosítani, amihez konkrét

előfordulásokat, példányokat kell kezelünk. A példány reláció központi jelentősége más formában is megfogalmazható:

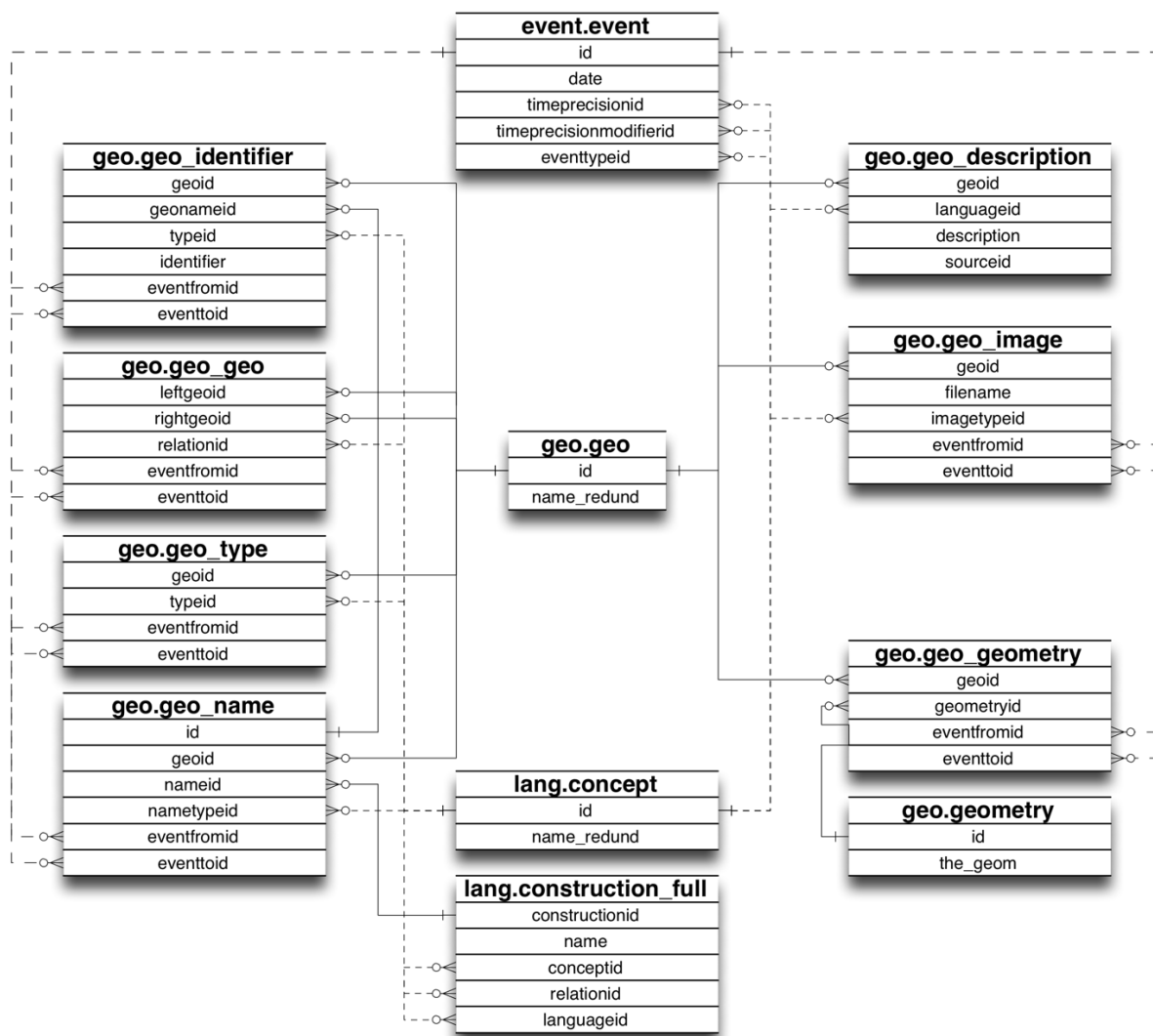
az individuumokat mindig valamilyen általános fogalom alá kell rendelni (vagyis mindig meg kell adni a geoontológiai típusukat).

A tezauszokban használt, háromféle hierarchikus reláció közül mindenképpen használnunk kell a generikus alárendeltje és a generikus fölérendeltje relációpárt (egymás inverzeként), a partitív alárendeltje (pars) és a partitív fölérendeltje (totum) relációpárt (egymás inverzeként), viszont nincs szükségünk az előzménye/oksága és utózmánya/következménye relációkettősre. Fontos megjegyezni, hogy a korábban és más rendszerekben gyakran használt ISA relációt nem szabad megengedünk, mert pontatlan meghatározásával több kárt okozhatunk, mint amennyi előny származik a – látszólagos – egyszerűségéből.

A tezuszban használt deskriptor, nem-deszkriptor kategóriákra itt nincs szükség. Ez a két kategória a szinonimitás kezelésére alkalmas. A tulajdonnévterekben a szinonimitást nem kell az átkános fogalmak szintjén kezelni, mert itt alapvetően individuális fogalmakkal foglalkozunk. A példányok szintjén is előfordulhat ugyan a szinonimitás jelensége, hiszen egy konkrét névhordozó egyednek lehet több neve, ami szemantikailag ide tartozik, de ezt a problémát a – kontextusfüggő – kitüntetett névalak kategóriájának bevezetésével lehet megoldani.

1.3.2.6 Földrajzi névtér logikai adatmodellje

A fent röviden kifejtett megfontolások értelmében a földrajzi névtér az alábbi logikai adatmodell alapján lehet felépíteni.



1.4 A nyelvi elemek kezelése

A névtérket képessé kell tenni, hogy igazítani lehessen őket nemzetközi névterekhez. Ezzel kapcsolatban kétféle elvárás lehet megfogalmazni. Egyfelől biztosítani kell, hogy tetszőleges nyelvű neveket fel lehessen venni a rendszerbe, és a nevekhez hozzá lehessen rendelni a nyelvi eredetet elégségesen reprezentáló adatokat. Másfelől biztosítani kell azt is, hogy névterekben használt fogalomkészlet többnyelvű legyen. Induláskor kétnyelvű (angol-magyar) rendszert várhatunk el realisan, de a rendszert úgy kell megtervezni és felépíteni, hogy a későbbiekben más nyelvek fele is könnyedén lehessen nyitni.

Minden nyelvi elem egy teljesen önálló nyelvi rétegben van tárolva, és a névtérben csak idegen kulcsú hivatkozásokat tárolunk, amelyek a nyelvi réteg meghatározott szegmensébe mutatnak. A névtér lényegét adó tulajdonnevek a nyelvi adatréteg konstrukciós állományának elemeiből kerülnek ki, míg a geointitások (a névhordozók és a nevek) tipizáláshoz szükséges köznév-állományok fogalmként vannak a névtér megfelelő pontjaihoz kötve. A nyelvi rétegben a fogalmak nyelvfüggetlen módon vannak tárolva, és a fogalmakhoz tetszőleges nyelven lehet nyelvi konstrukciókat rendelni. Ez adja meg annak lehetőségét, hogy a névtér felhasználói felületét bármilyen nyelvű címkeállománnyal fel lehessen tölteni (ha megvannak a fogalmakhoz köthető konstrukciók az adott nyelven). A nyelvi réteg elkülönült

kezelése kis mértékben megnehezíti a névtér adatainak kezelését, de a nyelvészeti elemzések számára rengeteg többletfunkciót képes biztosítani.

A nyelvi réteg elkülönítése azt is lehetővé teszi, hogy az entitások attribútumait, illetve kapcsolatait úgy írjuk le, hogy az adott attribútumot vagy kapcsolatot jellemző köznevet (általános fogalmat) a nyelvi rétegen keresztül kötjük be (vagyis csak a nyelvi-fogalmi rétegbe mutató idegenkulcsokat írunk be a névtér-adatbázisba). Ez a megoldás abból a szempontból nagyon rugalmassá teszi a névtéradatbázis építését, hogy ha a nyelvi réteg fogalmi szegmensébe felveszünk új elemeket, akkor azokra azonnal lehet hivatkozni a névtéradatbázisban anélkül, hogy a névtéradatbázis sémáján bármilyen változtatásra lenne szükség. Ha valaki egy teljesen új fogalmi sémával szeretné leírni a névtér geoindividuumaikat, azok neveit, akkor ezt bármikor megteheti úgy, hogy a nyelvi-fogalmi rétegben felépíti a saját fogalomkészletét, és annak megfelelő elemeit a geoindividuumokhoz, nevekhez rendeli a névtéradatbázisban. Arra is mód van, hogy a fogalmi sémák közti megfeleltetéssel tetszőleges adatcsere-, részleges vagy teljes export/import-folyamatokat lehet menedzselni.

Egy nyelvi konstrukció mindig valamilyen nyelven kifejezett megnyilatkozás, amely az adott nyelv morféma-készletéből a nyelv – tág értelemben vett – szintaktikai szabályait követve épül fel és a jelentését valamilyen fogalmi struktúra valamely eleméhez kapcsolódóan "nyeri el". A nyelvi konstrukciókat tehát formailag morféma-kból állítjuk elő, tartalmilag fogalmak segítségével kötjük meg. A különböző nyelveken kifejezett nyelvi konstrukciók kapcsolódhatnak ugyanahhoz a fogalomhoz. Elvileg bármely fogalmat kifejezhetünk bármely nyelven, ezért a fogalmat tekinthetjük nyelvfüggetlen entitásnak. A konstrukcióhoz mindig tartoznia kell egy nyelvnek, és ha egy konstrukciónak több nyelven is teljesen megegyezik mind a formai alakja, mind a tartalmi jelentése, akkor annyi konstrukcióról beszélhetünk, ahány nyelven használják.

| concept | construction | language |
|------------|--------------|----------|
| "kutyaság" | dog | angol |
| "kutyaság" | kutya | magyar |
| "kutyaság" | eb | magyar |

A morféma-k – vizuális-írási környezetben – valamilyen string segítségével reprezentálhatóak. Ugyanazt a stringet több nyelv is használhatja, sőt, egy nyelv esetében is gyakran előfordul, hogy a nyelv egy stringből több morféma-t állít elő. Az ugyanazon stringet használó morféma-kat úgy tudjuk egyértelműen elkülöníteni egymástól, hogy a stringhez való kapcsolatukban megadjuk meg azt, hogy milyen nyelvű morféma-ról van szó, illetve milyen alapvető nyelvi típusbesorolása (legtöbb esetben: milyen szófaja) van a morféma-nak. A 'fog' string mind a magyar, mind az angol nyelv része, de más konstrukció lesz belőle a két nyelvben, míg a magyarban képezhető kétféle konstrukciót azon az alapon különböztetjük meg, hogy másféle szófajba soroljuk a kétféle használati módját.

| string | morph | language | morphotype |
|--------|-------|----------|------------|
| fog | fog | angol | főnév |
| fog | fog | magyar | főnév |
| fog | fog | magyar | ige |

Morféma-k a szavak, melyek önállóan is használhatóak, de morféma-k a toldalékok is, amelyek önmagukban nem lehetnek megnyilatkozások, csak más morféma-kkal együtt kerülhetnek nyelvi használatba. A morféma-kat sokféle jeggyel jellemezzük, hogy a nyelvi működés módjuk teljes funkcionalitását meg tudjuk ragadni. Ezek a jegyek a morféma-k – többszörös – típusbesorolásainak tekinthetők. Egy nyelv konstrukció állhat több morféma-ból is. Ilyenkor – a legtöbb nyelven – úgy képezzük a konstrukciót, hogy annak morféma-komponenseit egymás után illesztjük (konkatenáljuk).

megtalált = meg | talál | t

A konstrukciókat sokféle tipizálási logika mentén lehet jellemezni. A modellben a konstrukciókat a morfémák– rangsorolt – felsorolásával lehet megadni.

| construction | component | rank |
|----------------------|-----------|------|
| Decsi-Nagy-Holt-Duna | decsi | 1 |
| Decsi-Nagy-Holt-Duna | - | 2 |
| Decsi-Nagy-Holt-Duna | nagy | 3 |
| Decsi-Nagy-Holt-Duna | - | 4 |
| Decsi-Nagy-Holt-Duna | holt | 5 |
| Decsi-Nagy-Holt-Duna | - | 6 |
| Decsi-Nagy-Holt-Duna | duna | 7 |

A tulajdonnevekben alkalmazott nagybetűs írásmód "kötelmét" abból a – morfémák szintjén számon tartott – információból lehet származtatni, hogy a komponensmorfémák milyen alaptípusba vannak sorolva. A konstrukciót egy függvény állítja elő a komponensmorfémák konkatenálásával úgy, hogy eközben figyelembe veszi a morfémák alaptípusát és morfológiai jegyeit. A morfológiai jegyek között van az a szabály is, hogy a konkatenálás során kell-e és ha igen, melyik oldalon kell szünetet hagyni a morfémák között.

A nyelvi konstrukciókkal segítségével közölt jelentéseket úgy modellezhetjük, hogy a jelentést a 'fogalom' entitással reprezentáljuk. A fogalmak jelentését primitív alapfogalmak megadásával, illetve a fogalmak egymás közti kapcsolatain keresztül érthetjük/adhatjuk meg. A fogalmak közti relációk közt legfontosabb a 'generikus alá-fölérendeltje' reláció, de a fogalmi struktúrába – elvileg – bármilyen relációt fel lehet venni. Az adott nyelvi adott konstrukció jelentését úgy adhatjuk meg, hogy a konstrukciót hozzákapcsoljuk a kiválasztott – nyelvfüggetlen – jelentéshordozó fogalomhoz. Értelemszerűen több konstrukciót is lehet ugyanahhoz a jelentéshez kapcsolni – akár több, akár egyetlen nyelven (előbbi esetben fordításról, utóbbiban szinonimitásról van szó). Ha egy névtérben az adatelemek leírása, tipizálása során fogalmakat használunk, akkor a tényleges nyelvi konstrukciókat úgy tudjuk megjeleníteni, ha egy nyelvet kiválasztva leválogatjuk a megfelelő konstrukciókat a fogalmakat és konstrukciókat összekapcsoló tábla segítségével. A kommunikációs felület többnyelvűsítése tehát csak annyit jelent, hogy a felületeken használt fogalmakhoz leválogatjuk a megfelelő nyelvi kifejezéseket. A fentiek alapján a nyelvi réteg fogalmi modellje a következő lehet.

őket. Ilyen nevezetes esemény a személy születése és halála. Ha az idő- és térpontosság kategóriáit kellően finom felbontás mentén használjuk, akkor már az egyik nevezetes esemény, a születés is elég lehetne a személyek elkülönítésére, hiszen az egyetűjű ikrekre még mondhatnánk azt, hogy a tér centiméterre azonos pontján születtek (ha ezt mérni tudnánk), de az idő dimenziójában biztos találhatnánk – legalább percekben mérhető – különbséget az ikrek születési ideje között. A személyek eseményeken keresztüli megkülönböztetésének elvi lehetősége a gyakorlatban nem teljesen működik, mert nem tudjuk kellő pontossággal összeszedni a szükséges adatokat minden személy esetében.

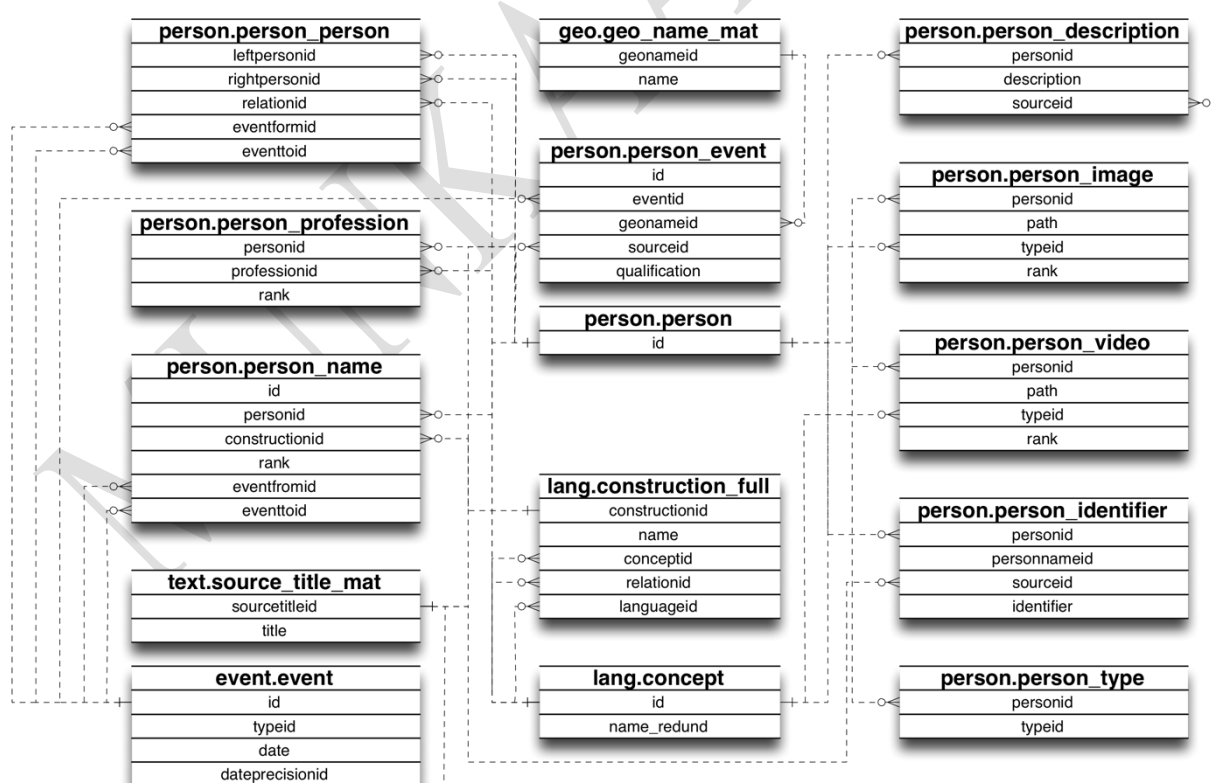
A személyek tipizálását többféle szempont szerint el lehet végezni. A közgyűteményi adatbázisok közti átjárhatóság megteremtésére szolgáló személynévtérben a személyek nemének, nemzetiségének, valamint az élete során művelt foglalkozások listájának megadása elégséges alapot nyújthat ahhoz, hogy a személyt (a neveivel és nevezetes eseményadataival együtt) egyértelműen beazonosíthassuk.

A fentiekén túl érdemes még néhány olyan adattípust is felvenni a személynévtér modelljébe, amelyek az egyértelmű azonosításhoz már nem szükségesek, de a személyek jellemzéséhez hozzájárulhatnak. Ilyen adatok például a személyek közti rokonsági kapcsolatok vagy a személyekhez köthető kulturális dokumentumok listája, képek, videók, személyes bibliográfiák, filmográfiák, diszkográfiák stb.

Az intézményközi kapcsolatok kialakítása és fenntartása végett természetesen szükség van még arra is, hogy a névtérben tárolni lehessen más névterek azonosítóit is, hiszen a különböző névterek közti átjárhatóságot pont ez ilyen összerendelések teremthetik meg. A digitális környezetben az egyértelmű azonosítás feladatát mesterséges egyedi azonosítók kiosztásával lehet megoldani, és a névterek, kulturális adatrendszerek közti interoperabilitás a különböző kontextusokban kiosztott azonosítók összekapcsolásával biztosítható.

1.5.1.1 Személynévtér logikai adatmodellje

Az előző fejezetben kifejtett megfontolások alapján a személynévtér modelljét az alábbi ábrán foglalhatjuk össze.



1.6 Testületi névtér

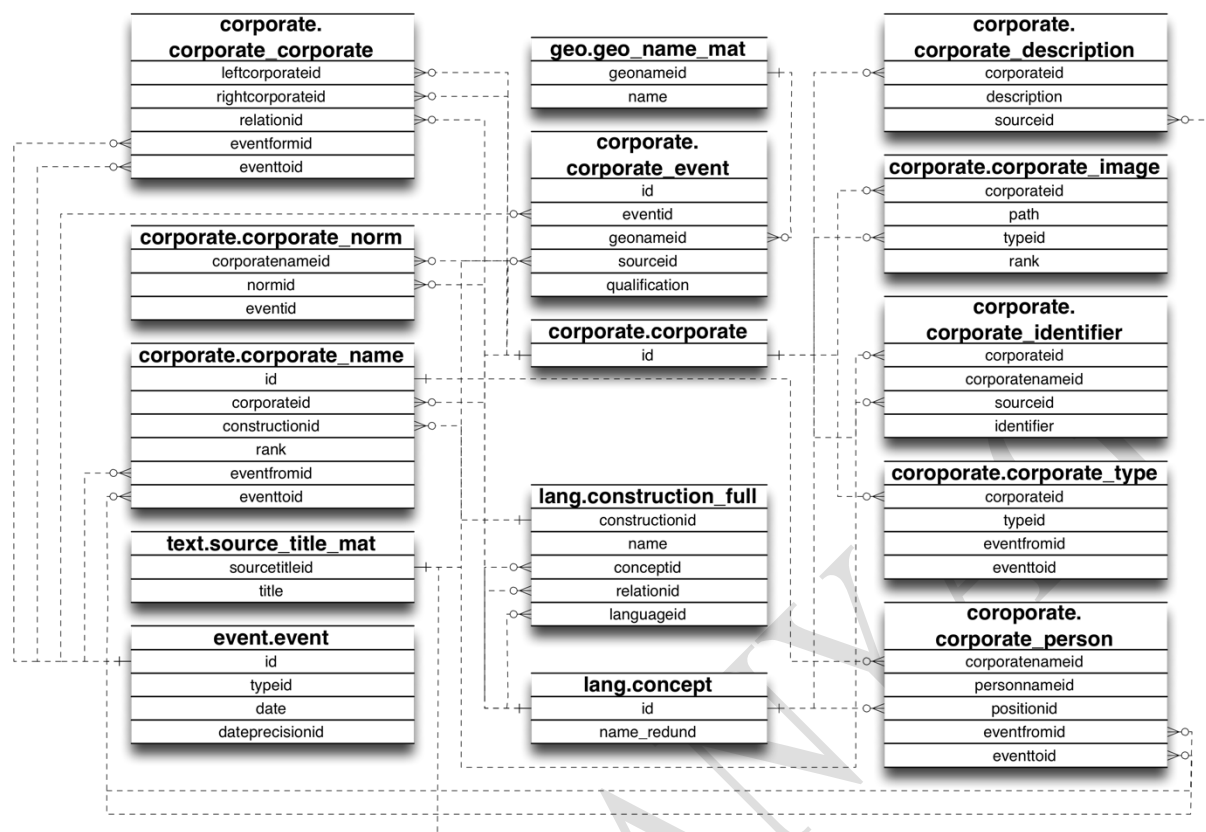
Ahogy azt már korábban jeleztük, a testület fogalma az ágens alá sorolható be. Testületről beszélünk akkor, ha emberek valamilyen csoportját összekapcsoljuk a testületi létüket meghatározó szabályok, értékek, szerepek együttesével. Testületekről van szó akkor, amikor a hétköznapi életben szervezetek, intézmények, vállalatok, egyesületek, társadalmi szervezetek, cégek, csapatok, pártok stb. kérdéseiről beszélünk.

Szemben a személyekkel, amikor a leírásukhoz domináns módon természeti tényeket használhatunk, úgy a testületek jellemzésekor inkább a társadalmi tények a meghatározók.

A testületek modellje értelemszerűen nagyon sokban hasonlít (hasonlítania kell) a személynévtér modelljéhez. A testületekhez ugyanúgy rendelhetünk neveket, ugyanúgy tipizálni lehet és kell őket (akár több fogalmi dimenzióban is). A döntő különbség a személyekhez képest az, hogy a testületeket mint társadalmi konstrukciókat nem tudunk természeti tényekkel jellemezni. A testületek nem a valóságos fizikai téridőben, hanem a konstruált, szimbolikus, társadalmi téridőben léteznek, és ezért rájuk vonatkoznak azok a kényszerek, amelyek a fizikai létezők esetében mindig fennállnak. A testületeket is hozzárendelhetjük a tér és idő megfelelő pontjaihoz, mondhatjuk, hogy egy testület valahol, valamikor létrejött ("megszületett") vagy megszűnt ("meghalt"), de ezeket az állításokat csak metaforikus értelemben használhatjuk. Az azonosításhoz, megkülönböztetéshez sok esetben elégségesek lehetnek az ilyenfajta adatok is, de a társadalmi konstruáltság alapvető ténye megoldhatatlannak tűnő filozófiai kérdéseket vet fel. Tegyük fel, hogy egy politikai párt megalakul valamikor, majd később beszünteti a tevékenységét. Ilyenkor a névvel, a típusbesorolással, illetve a megalakulás és megszűnés dátumával elég jól behatárolhatjuk ezt a testületet, és megkülönböztethetjük egy később ugyanazon a néven alakult politikai formációtól, tehát az azonosítás (és elkülönítés) sikeres lehet ilyen adatokra támaszkodva. Arra a kérdésre azonban nem tudunk sem ezen adatok, sem más adatok alapján megnyugtató választ adni, mi van akkor, ha a valamikor megszűnt párt újrakezdi a működését a "megszűnése" után (ahogy számos példa volt erre a magyar történelemben az 1989-90-es rendszerváltás időszakában). Mikor és milyen alapon mondhatjuk ilyen esetekben, hogy a történelmi idő két pontján létező testület ugyanazon individuum-e vagy sem.

1.6.1.1 Testületi névtér logikai adatmodellje

A testületi névtér adatmodellje az alábbi ábrában foglalható össze.



1.7 Eseménynévtér

Az eseménynévtér – alapértelmezés szerint – olyan eseményeket tartalmaz, amelyek más névterekből származnak: a személyek születési és halálozási eseményei, a földrajzi helyek geometriájában, nevében, típusbesorolásában történt változások, a testületek megalakulása, megszűnése, testületekben a személyek által betöltött pozíciók változása stb. Az eseménynévtér a más névterekből "származtatott" események gyűjtőhelye. Az effajta igények kielégítéséhez szükség van egy alapos – az egyes névtereken belől keletkező események típusait lefedő – eseménytipológiára, és arra, hogy az események időpontosságát rugalmasan tudjuk kezelni.

Az eseményeket természetesen lehet "saját jogukon" is kezelni, és definiálhatunk önálló logikájú eseménynévteret is, amelyben mindenféle eseményeket megengedünk felvenni, és ami minden olyan igényt képes lehet kielégíteni, amelyek az idővonal vagy kronológia jellegű szolgáltatásokkal kapcsolatban általában felmerülhetnek. Ennek az igénynek akkor felelhetünk meg a valóságban, ha megpróbáljuk önmagukban is modellezni az események jelenségét. A következőkben erre teszünk kísérletet.

A kronológia vagy idővonal – az időhöz köthető – események valamilyen csoportját összekapcsoló, azokat adott felületen egyszerre megjelenítő szolgáltatás. Mivel a világ értelmezésének, leírásának egyik módja a világ dolgait eseményeken keresztül próbálja megragadni, alapkérdés, hogy milyen eseményeket kell/érdemes figyelembe vennünk, kezelnünk, megjelenítenünk a projekt keretein belül. Nyilván nem kell mindenféle eseményt kezelnünk, csak azokkal érdemes foglalkoznunk, amelyek erős módon kapcsolódnak valamelyik névtérszegmenshez. Korábban már említettük (meg is indokoltuk), hogy az egyes névtérszegmensek (személy, testület, földrajzi hely) leírásakor az entitások időfüggő minőségeit, kapcsolatait eseményszerűen írjuk le. Ha viszont az egyes névtérszegmensekben szükségünk van eseményekre, akkor adódik a lehetőség (és elvárás is egyben), hogy ezeket az eseményeket önmagukban is, de az adott névtérszegmensen egészen belül lehessen kezelni. Az egyes nemzeti névtérszegmensekbe mindig beilleszthetünk egy olyan idővonalat, amely a szegmens összes

fontos eseményét képes – a kronológiákra jellemző – önálló logika mentén kezelni. Az egyes névtérsezmensekhez rendelt eseményeket azonban összegezni is lehet és önálló szolgáltatásként definiálható egy olyan idővonal, amely az összes névtér eseményeit egyben tartalmazza.

Az idővonal többféle értelmezésének és hasznosításának lehetősége azonban nem változtat azon a tényen, hogy magát a kronológiát (idővonalat), vagyis események összekapcsolt rendszerét egységesen, az idővonal szolgáltatás logikájának megfelelően kell kezelni. A továbbiakban összeszedjük azokat a szempontokat, definíciókat, amelyekre szükség lehet az idővonal-szolgáltatás megtervezéséhez és fejlesztéséhez.

1.7.1 Mi az az idővonal

- A KRONOLÓGIA (CHRONOLOGY) egyedi események időben elrendezett szolgáltatása, amihez szükség van egyrészt egy, az eseményeket leíró adatok kronológiai adatbázisra, másrészt egy, az adatok időbeliségét kezelni képes alkalmazásra.
- A KRONOLÓGIAI ADATBÁZIS (CHRONOLOGY DATABASE - CDB) egyedi eseményeket és a köztük levő kapcsolatokat leíró, időfüggő adatok strukturált gyűjteménye. Az események leírásához szükség van további entitás- és relációtípusok (helyszín, ágens, tárgyak, nem megfogható dolgok stb.) modellbe emelésére.
- A KRONOLÓGIASZERVER (CHRONOLOGY SERVER - CS) események (kronológiai adatbázisok) adatait megjelenítő alkalmazás, amely képes az események időbeliségével kapcsolatos funkciók, metódusok kezelésére.
- A kronológiának (kronológia-adatbázisnak és -szervernek) két fő típusa létezik a kezelt események típusától függően: IDŐVONAL (TIMELINE - TL) és NAPTÁR (CALENDAR - CAL). Az idővonal szolgáltatásban az események abszolút idejét kezeljük, míg a kalendárszolgáltatásban az események relatív, évfüggetlen időbeli értékét vesszük csak figyelembe. Ebben az értelemben a naptár/kalendárium egy adott, (évente) ismétlődő eseményhalmaznak felel meg.
- Az események idejét mindig valamilyen IDŐPONTOSSÁG (TIME PRECISION) mellett tudjuk megadni, és egy kronológiai adatbázisra mindig megadható (és jellemzi azt) az időkezelés pontosságának alkalmazott mértéke. Azért van szükség időpontosság-kezelésre, mert gyakran előfordul, hogy a különböző eseményeket eltérő pontossággal lehet megadni, viszont valamilyen időtengelyen az összes eseményt rendezni kell tudni, amihez a különböző adattípusként (szöveges, dátum- vagy numerikus adatként) megadott dátumértékeket homogenizálni kell (hiszen csak ekkor lehet részleges rendezést végrehajtani az összes esemény dátumértéke szerint).
- Az idővonal-adatbázisban egy esemény mindig egyedi, partikuláris, ami annyit jelent, hogy mindig kell legyen valamilyen egyedi (konkrét, partikuláris) idő- és térbeli kapcsolódása. Másként szólva: egy – idővonal – esemény nem lehet általános (univerzális).
- A naptáradatbázisban az események kétféle módon értelmezhetőek. Egyfelől lehet őket általánosnak tekinteni, hiszen évről évre (ugyanúgy és ugyanakkor) megjelennek a kalendáriumokban, másfelől viszont lehet őket egyediként is kezelni, hiszen mindig hozzájuk lehet rendelni a "befogadó" évet, ami által egyediként lehet interpretálni őket. Amikor egy nevezetes esemény nevezetes (mondjuk, 100 éves) évfordulóhoz érkezik (a nevezetesség minőségének a legtöbb esetben pusztán számmisztikai alapja van), akkor látszik igazán, hogy a naptárba kerülésnek mindig egyedi oka van, mégha ez a kalendárium periodikus jellege miatt sokszor nem annyira nyilvánvaló.
- Amikor egy új – egyedi – eseményt felvesszünk a CDB-be, akkor mindig meg kell adnunk az (esemény)típusát. Egy egyedi eseménynek sosem lehet másolata, szemben az eseménytípussal, amely tetszés szerint ismételtető.
- Egy idővonal (adatbázis és alkalmazás) számára gyakorlati szempontból az idő-függőség fontosabb, mint a tér-függőség, noha elméletileg mindkét kapcsolat egyaránt lényeges. Mivel a térbeli adatok praktikus szempontból nem nélkülözhetetlenek a kronológiaszolgáltatás számára, ezért előfordulhat, hogy egy ilyen szolgáltatásban nem kezelik az események térhez kötöttségét, de a kezelt események egész halmazára vonatkozóan akkor is tudni, rögzíteni kell (legalább a szolgáltatás címében), hogy a feldolgozott eseményeknek milyen térbeli referenciája van (ez

lehet egy ország, egy kontinens vagy akár az egész Föld, ha az emberiség történelmének valamilyen tematikus szegmensével kapcsolatos adatokról van szó).

- Amennyiben az események térbeli vonatkozásait is felvesszük a modellbe, akkor a térbeli adatok TÉR PONT OSSÁGÁT (LOCATION PRECISION) is kezelni kell (az időpontossághoz hasonlóan).
- Az időben rendezhető adatokat nemcsak kronológiai jellegű szolgáltatásban lehet kezelni, léteznek más műfajok is. Az időbeliségnek fontos szerepe van, mégsem tekinthető elsődleges rendezőelvnek az olyan tematikusan önállóan tekintett műfajokban, mint az életrajz, a helytörténet vagy a gazdasági, társadalmi folyamatok idősoros elemzése.

1.7.1.1 Alapfogalmak

A kronológiaszerver modellezéséhez, felépítéséhez szükséges alapfogalmakat a következőképpen definiálhatjuk.

1.7.1.1.1 idővonal(szolgáltatás)

- Időfüggő, elemi eseményeket tartalmazó, idővonalszerver által működtetett szolgáltatás. Két komponense van:
 - idővonalszerver
 - idővonal-adatbázis
- Amennyiben nem zavaró, akkor az 'idővonal' és a 'kronológia' terminusokat szinonimáknak tekintjük és egymással felcserélhető módon használjuk.

1.7.1.1.2 naptár(szolgáltatás)

- Évfüggetlen kalendáreseeményekre szűrt kronológia-adatbázisra épülő, kalendárszerver által működtetett szolgáltatás. Két komponense van:
 - kalendárszerver
 - szűrt kronológia-adatbázis + speciális naptár-típusú adatkészlet (például keresztnév adatbázis)

1.7.1.1.3 idővonalszerver (timeline server)

- Az elemi események kezelésére alkalmas alkalmazás.

1.7.1.1.4 kalendárszerver (calendar server)

- A naptáreseemények kezelésére alkalmas alkalmazás.

1.7.1.1.5 kronológia-adatbázis (chronology database)

- Események tulajdonságait és kapcsolatait leíró adatok adott séma alapján elrendező adatbázisa.

1.7.1.1.6 idő (time)

- Az idő az idő, szükségtelennek (és nehéznek) tűnik definiálni a fogalmat. Ami fontos és kezelhető itt: az eseményeknek kétféle reprezentációs módja létezik. A kronológia alapértelmezett prezentációs formája a végtelenített időtengelyre vetített lineáris eseménykezelés, de lehetséges másfajta megjelenítési technológia is (például havi, heti, napi nézetben való eseménykezelés), amikor valamilyen ciklikusságot mutató reprezentációs formát alkalmazunk. Vizuálisan kétféle módon kezelhetjük az időt:
 - lineáris koordinátarendszerben, illetve
 - poláris koordinátarendszerben ábrázolva.

1.7.1.1.7 esemény (event)

- A kronológia alapegysége az esemény. Esemény bármi lehet, amihez valamilyen dátumszerű adatot lehet rendelni, és van valamilyen neve (címe) és tartalmi leírása. Az események megragadásához minimálisan szükséges adatkészlet tehát a következő:
 - az esemény időbeli vonatkozása,
 - az esemény neve,
 - az esemény leírása.

- Az esemény három szükséges adatmezője (dátum, név és leírás) mellé további dokumentummezőket is fel lehet venni, pl. mutatni lehet az eseményt reprezentáló kép-, videó-, hangdokumentumok elérhetőségére, és ezeket meg lehet jeleníteni valahogyan az idővásznon is (akár külön időrétegekben, akár popupablakokban, akár máshogy).
- Az eseményeket időbeli értékük alapján idővonalba (kronológiai rendbe, sorba) rendezhetjük.
- Egy esemény lehet:
 - történés (instantesemény, kiterjedés nélküli, pontszerű esemény), illetve
 - folyamat (terjedelemmel rendelkező, tartamszerű esemény).

pontszerű esemény: egy személy születése, házasságkötése, halála
 folyamat: egy személy élete, egy háború, egy előadás

- A történéshez egyetlen időpontot tudunk rendelni, a folyamathoz viszont mindig egy időszakasz tartozik, vagy másként: a folyamatnak mindig van kezdő- és záróidőpontja.
- Az időkezelés kétféle reprezentációs technikájához illeszkedően elkülöníthetünk kétféle eseménytípust:
 - tiszta eseményt (röviden eseményt), illetve
 - naptáreseményt vagy kalendáreseményt.
- A kalendáresemények periodikusan ismétlődő események, amelyek valamilyen időponttól fogva adott periodicitás mentén újra és újra létezhetnek. A kalendáreseményeket csak reprezentáció esetén, poláris kordinátarendszerben érdemes megjeleníteni (pl. valakinek a születésnapját évente kitenni egy végtelenített időtengelyre nem lenne szerencsés, de egy ciklusreprezentációban (kalendárium formátum esetén) ugyanez már értelmes lehet.
- Egy esemény lehet:
 - elemi esemény, illetve
 - konténeresemény (kompozit vagy összetett esemény).
- Amíg egy konténeresemény több, beágyazott eseményt tartalmaz(hat), az elemi esemény mindig "önmagából" áll. Elméletileg nagyon nehéz éles határvonalat húzni a két eseménytípus közé.

elemi esemény: egy személy születése, egy személy halála

konténeresemény: II. világháború (tartalmazza a sztálingrádi csatát, Lengyelország német megszállását, az atombomba ledobását Japánra), az első Orbán-kormány időszaka 1998-2002 között (tartalmazza a következő évi költségvetési törvények betervezését, a korona úsztatását Esztergomba).

- Ha lehetséges más eseményeket tartalmazó esemény vagyis konténeresemény, akkor ez igaz fordítva is, vagyis lehetséges eseményekbe ágyazódó esemény, azaz részesemény. Például a háború mint esemény részeseménye egy csata vagy egy béketárgyalás.
- Az eseményeknek szemantikusan sokféle, gyakran mélyen tagolt szemantikai struktúrája van. A különböző eseménytípusoknak lehet nagyon eltérő szemantikai szerkezete, adatmodellje.

Egy személy születésének mint eseménynek egészen más az adatstruktúrája, mint egy archontológiai eseménynek, mondjuk, egy miniszteri kinevezésnek.

- Minél gazdagabban, mélyebben struktúrált eseménymodell áll rendelkezésünkre, annál nagyobb rugalmasságot érhetünk el a szolgáltatásban, annál gazdagabb felhasználó élmény nyújtására lehetünk képesek, ezért nagyon fontos kérdés, hogy milyen adatstruktúrával írjuk le az esemény fogalmát. Az esemény minél több tulajdonságát és kapcsolatát felvesszük a modellünkbe, annál több linket tudunk felajánlani az események kapcsán, ami javítja a felhasználói élmény minőségét, hiszen az esemény mélyebb megértését és több irányú továbblépési lehetőséget biztosít.

1.7.1.1.8 időpontosság (time precision)

- Az események idejét eltérő pontossággal adhatjuk meg: egy eseményt megadhatunk az időtengelyen percre, órára, napra, hónapra, évre, évtizedre stb. pontosan.

nap-pontosság: 1989.03.02.

másodperc-pontosság: 1978.12.22. 10:24:10

évtized-pontosság: 1940-es évek, 1960-70-es évek

évszázad-pontosság: 18. század

továbbiak: 1956 ősze, 3. évezred, éjjel, este, az év közepén, a hónap elején, szilveszterkor

- Az időpontosság egymásba ágyazódó típusai egységesen kezelhetők egy időpontosság generikus fogalma alá besorolva.

1.7.1.1.9 időformátum (time format)

- Tudni kell kezelni azt, hogy ugyanaz az időpont többféle formátumban is megjeleníthető ugyanazon a nyelven.

1988.05.15., 15 May 1988, 1988-feb-15

- Tudni kell kezelni azt, hogy ugyanazt az időpontot többféle formátumban kell megjeleníteni eltérő nyelveken.

1988. május 15., 15 May 1988

1.7.1.1.10 relatív idő (relative time)

- Előfordulhat, hogy egy esemény időbeliségét egy másik eseményhez képest, annak nevére hivatkozva, relatív módon határozzuk meg. Ilyenkor a viszonyítási alapot jelentő események időadatait kell származtatni a hivatkozó eseményre.

a napóleoni háborúk után, X. házassága előtt

1.7.1.1.11 elemi esemény (elementary event)

- Az elemi esemény megjeleníthető egy egyszerű, lineáris Descartes koordinátarendszerben.
- Az idővonal-szolgáltatásban az eseményeket egy végtelenített, lineáris időtengelyen ábrázolhatjuk (reprezentációs célokból a lineáris időtengely helyett alkalmazni lehet másfajta időkezelési módot, például nagyon hosszú időszakot átfogó kronológiák esetében használni lehet logaritmikus időskálát).
- Az elemi esemény fogalmát tekinthetjük az alapértelmezett eseménynek. Ha nem zavaró, akkor az 'elemi esemény' terminusa helyett használhatjuk a rövideb 'esemény' kifejezést is.

1.7.1.1.12 naptáresemény (calendar event)

- Az időfüggő eseményadatok megjelenítésének másik módja az, amikor kihasználjuk bizonyos események periodikusan ismétlődő jellegét. Vélhetőleg a Föld Nap körüli körforgásából adódó, a természeti és biológiai jelenségek körében megfigyelhető periodicitás jelentette és jelenti azt a mintát, ami a társadalmi jelenségekhez kapcsolható események körében is kialakított és fenntart ismétlődéseket. Ezeket kalendárium vagy naptár jellegű eseményeknek minősítjük, és az időhöz való viszonyukat úgy kezeljük, hogy az éven belül elfoglalt helyüket tartjuk igazán

számon, ezeket a relatív időadatokat rendeljük az ilyen eseményekhez. Ezt a periodikus jelleget a poláris koordinátarendszerben lehet a legkönnyebben, legadekvátabb módon megragadni.

- A naptáreseemények speciális tulajdonságát kezelni képes alkalmazást nevezhetjük kalendárszervernek (vagy naptárszervernek).

1.7.1.1.13 idővászon (time canvas)

- Az idővászon az a sík, amelyen a kronológiai adatok (az események és a hozzájuk kapcsolódó további adatok) megjeleníthetők.
- Az idővászont fel lehet osztani több, diszjunkt területre, ahol a kronológiai adatbázis eltérő szegmenseit lehet megjeleníteni.

1.7.1.1.14 sáv (band)

- Az eseménysáv, az idősáv vagy röviden sáv ('event band' vagy 'band') tematikusan összetartozó események adatait megjelenítő felület az idővászon valamely elkülönített részén.
- Az idősávot (a rajta megjelenő eseményadatokkal együtt) mindkét irányban lehet tologatni az időtengely két irányában (a múlt és a jövő idő felé).
- Amennyiben egy idővászonon több idősáv is megjelenik, akkor az idővonal-alkalmazásnak képesnek kell lennie arra is, hogy ezek a sávok egymással szinkronban mozgathatók legyenek (ha a felhasználó az egyik sávban mozgatta a sávot az időtengelyhez képest, akkor az alkalmazás a másik sávban is ugyanahhoz az időponthoz szinkronizálja az eseményeket), de biztosítani kell azt is, hogy a sávok aszinkron módban legyenek használhatók. Ez az elvárás nem azt jelenti, hogy minden időrétegben feltétlen ugyanolyan felbontásnak kell szerepelnie, akár szinkron, akár aszinkron működésmód az aktuális elvárás, az időrétegekben, idősávokban lehet eltérő időfelbontás.
- Az idővonal-kiszolgáló alkalmazással szemben támasztott alapvető követelmény kell legyen az az elvárás, hogy a különböző sávoknak lehessen különálló, saját időfelbontása.

1.7.1.1.15 időszakdoboz (period box)

- Az időszakdoboz egy események valamilyen szempont szerint megszürt halmazát tartalmazó, négyzetes terület, amit értelmezhetünk a kiválasztott események konténerének.
- Az időszakdoboz önálló egységként hivatkozható, és egy sávon belül helyezkedik el.
- Az időszakdobozok egymásba ágyazhatók, vagyis egy korszakon belül definálhatunk rész- vagy alkorszakokat).
- Az időszakdobozhoz saját címet (nevet) és leírást kell hozzárendelnünk, de az időhivatkozásait nem önállóan kell megadnunk, hanem azokat örököltetni kell az időszak által magába foglalt eseményektől (az időszak kezdődátuma megegyezik a magába foglalt legkorábbi esemény kezdőidejével, míg az időszak záródátumaként az időszakba tartozó legkésőbbi esemény záródátumát kell értelmeznünk).

1.7.1.1.16 időtengely (time axis)

- Az időtengely az idő mértékét mutatja egy időablakon belül adott – vagy a felhasználó által kiválasztott, vagy a rendszer által automatikusan beállított – időfelbontás mellett.
- A – választott vagy automatikusan beállított – különböző felbontási szinteken az időtengelyen megjelenített időhöz eltérő időpontosság társul.
- Az időtengely alapértelmezett formája a lineáris Descartes koordinátarendszeren alapul.
- Az időtengely másik lehetséges használati módja poláris koordinátarendszeren alapul, és adekvát módon a naptáreseemények kezelésére alkalmas, de elképzelhető olyan megoldás is, hogy – valamilyen spirális interfészen keresztül – végtelenített időtengelyként funkcionáljon.

1.7.1.1.17 időfelbontás (time granularity)

- Könnyen előfordulhat, hogy egy idővonal-adatbázis annyi eseményt tartalmaz, hogy azok egyszerre már nem jeleníthetők meg az idővászonon belül. Ennek két oka lehet. Az mindig igaz, hogy a teljes eseményhalmazból a legkorábbi és a legkésőbbi eseményhez kapcsolt időpontok adják a legszélesebb időablak két végpontját. Előállhat olyan helyzet, amikor ez a maximális időablak megfelelő módon elhelyezhető a rendelkezésre álló idővászonon belül, vagyis az

adatbázis teljes időhorizontja elfér a képernyőn. Természetesen ilyenkor is jelentkezhet elhelyezési probléma, mert ha túl sok az esemény, amit egyszerre kell kirakni a képernyőre, akkor "besűrűsödési" probléma keletkezik. Ezt azonban még kezelni lehet úgy, hogy az időtengelyre ortogonális dimenzióban széthúzzuk az idővásznot, és az így megnövelt helyre szétterítve már nagyobb eséllyel férhetnek el az adatok. Ugyanerre a problémára természetesen megoldást jelenthet még az események valamifajta – interaktív, tehát felhasználóoldali – szűrése is. Lehet azonban más oka is annak, hogy az idővásznonra nem férnek el az események, amit már nem lehet orvosolni a fenti technikák egyikével sem. Gyakori esetként számolhatunk ugyanis azzal, hogy a kezelendő teljes időhorizont vagyis a maximális időablak jóval szélesebb, mint a rendelkezésre álló idővásznon mérete. Ilyenkor hiába növeljük az idővásznon méretét a másik dimenzióban, hiába szűrjük meg az eseményeket, a megjelenítési probléma ugyanúgy megmarad. Ekkor kell igénybe venni az időkezelésnek általában és ezen belül az idővásznon látható időtengelyrészhez rendelt időskála változ(tathat)ó felbontásának technikáját. Biztosítani kell, hogy ugyanakkora méretű idővásznon eltérő felbontású időszak legyen megjeleníthető. Ebben az esetben az időfelbontás éppen adott mértéke, szintje, valamint az éppen megjelenített időszakba eső események száma fogja meghatározni az események sűrűségét az aktuális idővásznon belül.

1.7.1.1.18 időfelbontási szint (granularity level)

- Az időfelbontás az a mérték, ami meghatározza (jelzi), hogy az idővásznon belül látható időtengelyrészen milyen két végpont közötti időszakot jelenítünk meg. Az időfelbontás az időablak és az idővásznon maximális szélességének aktuális arányát fejezi ki.

1.7.1.1.19 időközelítés (time zoom-in)

- Finomabb időfelbontás kérése/elérése, amelynek eredményeként az idővásznon kisebb időablak jelenik meg.

1.7.1.1.20 időtávolítás (time zoom-out)

- Durvább időfelbontás kérése/elérése, amelynek eredményeként az idővásznon nagyobb időablak jelenik meg.

1.7.1.1.21 időablakcsúsztatás (time pan)

- Változatlan időablakméret és változatlan időfelbontás mellett mozgatni az időablakot az időtengely mentén előre vagy hátra abból a célból, hogy időben korábbi vagy későbbi események jelenjenek meg az idővásznon.

1.7.1.1.22 időablak (time window)

- Az időablak azt az időszávot jelenti, ami az idővásznon látható az aktuális időfelbontás mellett. Az aktuális időfelbontás szintje és a rendelkezésre álló idővásznon mérete együttesen határozza meg, hogy milyen szélességű időszávban láthatjuk az eseményeket az idővásznon. Praktikusan az időablak két szélső értékét az a két időpont adja, amelyek az időtengelyen láthatóak az idővásznon két szélén.

1.7.1.1.23 narratíva (narrative)

- Egy komoly, adott területen a teljességre törekvő kronológia-adatbázis nagyon hamar elérheti azt az eseményszámot, azt az adatmennyiséget, amikor a szolgáltatott adatok teljesége egyfelől áttekinthetelenné, másfelől érdektelenné válik a látogatók számára. Emiatt van nagyon nagy szükség arra, hogy a különböző kérdésekkel, érdeklődési irányokkal érkező felhasználók mindig találjanak megfelelő belépési pontokat, haladási útvonalakat a szolgáltatáson belül. Az eseményhalmaz egészen belüli fókuszok kialakítását narratívák segítségével lehet biztosítani. Narratívának tekinthetünk bármely tetszőleges (tehát önkényes) válogatást a kronológia-adatbázis teljes eseménykínálatából, amit valamilyen történettel, "mesével" összekapcsol valaki (felhasználó, szerkesztő, "mesemondó").

egy személy életeseeményei, életrajza, egyéni válogatás a leghíresebb, legsikeresebb filmpremierekből, híres tengeri katasztrófák, földrengések története

- A narratívát értelmezhetjük szűkített kronológiaként, mint egy kis kronológiát a nagy kronológián belül.
- A térinformatikában bevett érdeklődési vagy érdekes pontok (point of interest) mintájára itt beszélhetünk érdekes eseményekről (line of interest -LOI). Ezek funkciója, értelme, haszna az lehet, hogy fókusz, figyelmet irányítanak egy eseménnyvonulatra, segítséget nyújtanak az események tengerében (a teljes kronológiai adatbázisban) eligazodni.
- Az idővonal-alkalmazás egyik legfontosabb funkciója lehet annak biztosítása, hogy a felhasználó könnyen tudjon válogatni, kapcsolni a lehetséges, felkínált narratívák között.
- az eseményeket sok szempont alapján tematikus csoportokba lehet szervezni (politikai, kulturális, filmes események, születések, halálozások stb.)
- az események tematikus csoportjait az idővásznon tematikus rétegekben lehet megjeleníteni, amelyekre vonatkozóan a felhasználói felületen rétegkezelő képeséget kell kialakítani a felhasználók számára (rétegválasztás, réteggikapcsolás, rétegbekapcsolás, rétegsorrend-kialakítás)
- Az eseményekre kényszereket lehet alkalmazni a dátumértékek alapján (pl. kezdődátum < záródátum vagy befoglaló esemény kezdődátuma <= részeseemény kezdődátuma)
- finomabb adatmodell esetén saját sémával rendelkező eseménytípusokat lehet definálni (személyes életeseemény, archontológiai esemény, publikálási esemény stb.)
- saját sémával rendelkező eseménytípusok esetén lehetséges típuskényszereket definálni
- saját sémával rendelkező eseménytípusok esetén lehetséges önálló névttereket létrehozni és azok adattartományában szűrési, keresési műveleteket végrehajtani. Például úgy szűrni egy adott időréteg (vagy akár az egész idővásznon) tartalmát, hogy egy, a felhasználó kiválasztott eseménytípus példányait vagy egy kiválasztott konkrét személy, szervezet, dokumentum, földrajzi objektum stb. kapcsolati körébe tartozó eseményeket jelenítjük meg
- eseménylánc = önálló eseményeket valaki összekapcsol (pl. oksági elven) eseményszekvencia
- beágyazott hierarchia: az 56-os felkelés magába foglalja a Rádió ostromát, a Sztálin szobor ledöntését, a szovjet ki-, majd bevonulást

1.7.1.1.24 interpretáció (interpretation)

- Az interpretáció az események (részben) értékelő jellegű leírása erősen értékvonatkozású, érzelmi töltetű terminusok használatával. Ebből a megközelítésből nézve az interpretáció a tények és a hozzájuk kapcsolt – nem feltétlenül tudatos, nem feltételül szándékos – vélemények, értékelések együttesét jelenti. Egy eseményt háromféleképpen lehet értékelni: pozitív, negatív, illetve semleges módon.

Egy eseménynek mindig lehet többfajta értelmezése. Ugyanazt az politikai eseményt lehet pozitívan értékelve (forradalomként) vagy negatív módon (a csőcselék randalírozásaként) interpretálni, de lehet írni értéksemleges(ebb), technikai kategóriák segítségével is (tömegmegmozdulás).

- A kronológiába kiválasztott eseményhalmaz és az események interpretációja természetesen szorosan összekapcsolódik egymással, de ez az összefonódás gyakorlati okokkal magyarázható, elméleti szempontból nem mondható szükségszerűnek. Az eseményeket leíró tények és mindezek interpretációja elválaszthatók egymástól, könnyen elképzelhető, hogy két narratíva ugyanabból az eseményhalmazból áll, mégis különböznek egymástól a tényekkel összekapcsolt eltérő értékvonatkozások, azaz az eltérő interpretációk miatt.
- Az interpretáció részben gyengébb, részben másfajta értelmezését jelenti az az értékalapú, tehát valamilyen szinten elfogult kiválasztás, ami az események elvileg végtelen láncolatából besorolja a kronológiai adatbázisba vagy annak egy részletébe, egy narratívába az arra érdemesnek, fontosnak tartott eseményeket. Ebben az esetben az interpretáció egyfajta válogatásnak (szűrésnek) fogható fel.

1.7.1.1.25 narratívaváltás (changing narratives)

- Miközben a felhasználók barangolnak a felkínált idővonal-események (valamely narratíva alapján előzetesen kiválasztott eseményhalmaz elemei) között, az idővonal-szolgáltatásnak a lehető legtöbb – az aktuálshoz képest eltérő – narratívát is fel kell tudni ajánlani. Ezt úgy lehet megvalósítani, hogy az aktuális narratíván belül olyan linkeket helyezünk el, amelyek mentén a felhasználó más narratívákba ugorhat át.

Tegyük fel, hogy a felhasználó az amerikai politikatörténet eseményeit nézegeti (vagyis az amerikai politikatörténeti narratívában van). Amikor arra az eseményre "lép", aminek neve: "Meggyilkolták J.F. Kennedy elnököt Dallasban", akkor egyfelől maradhat továbbra is az aktuális narratívában és nézheti tovább a gyilkosságot követő politikatörténeti eseményeket, másfelől választhat magának más narratívát is (a felkínált linkekre kattintva):

- választhatja JFK életrajzi narratíváját (JFK nevére kattintva)
 - továbbugorhat a Dallas helytörténetét bemutató helytörténeti narratívára (Dallas nevére kattintva)
 - kérheti az aznapi események teljes listáját – Amerikában vagy akár az egész Földön (egy külön ugrópont segítségével)
 - továbbléphet a politikai gyilkosságok történetét bemutató narratívára (a "meggyilkolták" szóra kattintva)
- Az idővonal minden individuumának (személyeknek, testületeknek, szervezeteknek, földrajzi helyeknek, dokumentumoknak, nevezetes (számozott, megjelölt) tárgyakkal) lehet saját történelme, narratívája, ezért az idővonalon megjelenő minden egyedi dolog elméletileg egy új narratíva belépési pontjaként funkcionálhat. Ennek az elvi lehetőségnek a kihasználását praktikus nehezíti, hogy ehhez az adatbázisba külön, önálló egységként kell felvenni az események minden összetevőjét, kapcsolatát.

1.7.1.1.26 eseménykapcsolatok (event relationships)

- az események között a dátum szerinti rendezés értelemszerűen megteremt egy szinkronicitási kapcsolatot (ez a kronológiák kidolgozásának legfontosabb hozadéka)
- az események között további kapcsolatokat lehet definálni, pl. okozati (előzmény/következmény) reláció vagy tematikus összetartozás, vagy eseményekben érintett „szereplők” (személyek, szervezetek, helyszínek stb.)
- ezeket a szemantikus tartalmú eseménykapcsolatokat (relációs adatokat) vagy csak emberi közreműködéssel lehet feltölteni (amit az automatikus forrásfeldolgozás nem tud biztosítani), vagy olyan adatsémára van szükség, amely lehetővé teszi adott entitások szerinti keresés, szűrés műveletét.

1.7.1.1.27 kulcsesemény (key event)

- Egy narratíván belül lehet egy (vagy több, de nem sok) olyan esemény, amelynek kiemelt szerepe, nagyobb hangsúlya van a narratíva egészén belül. Ez(eke)t nevezhetjük kulcseseményeknek. Az ilyen minősítés "kiosztása" nyilvánvalóan az értékelés egyik formája, így ez az adat elvileg az interpretációs rétegbe tartozik. Kulcseseménynek minősíteni egy eseményt egyben jelenti azt is, hogy:
 - egy vagy több esemény fontosabb a többihez képest,
 - a nem kulcsesemények interpretációja, fontosságátulajdonítása függ a kulcseseményektől (az utóbbiaknak tulajdonított jelentőség módjától, okától, formájától),
 - a fentiekből következően a narratíváknak van (lehet) struktúrájuk.
- A kulcseseménnyé minősítés szubjektív döntés, értékelés, ez a minőség nyilván nem az esemény inherens tulajdonsága. Az idővonalservernek biztosítani kell egy adminisztratív felületet, amelyen keresztül a felhasználók, szerkesztők elvégezhetik az effajta minősítéseket.

1.7.1.1.28 gestalt alakzat (gestalt pattern)

- Az események idővonal szerű elrendezésének egyik episztemológiai potenciálja, ígérete, hogy az időben rendezett adatokat egy reprezentációs térben megjelenítve gestaltélményeket szerezhetünk, vagyis olyan átfogó, egész-jellegű kapcsolatokat fedezhetünk fel az események között, amelyeket az egyedi eseményekre fókuszálva nem ismerhetünk fel.
 - Ha valaki 1960-ban született Magyarországon, akkor 1978-ban lett nagykorú, és 18 évet élt a kommunista rezsimben, de a rendszerváltás után több évtizedet tölthetett el demokratikus viszonyok között. A két időszakban (részben) mások voltak a jellemző eseménytípusok, a kulcseseményekhez kapcsolódó érzelmek, értékek, tapasztalatok, és a mindezekre vonatkozó szubjektív, személyes emlékek. Ezekről függetlenül az öregedéssel is más tapasztalatok, más élmények, más hangsúlyok fényében értékeli (értékeli át) ugyanazokat az eseményeket.
 - Teaching the history of media devices it can be showed only with the help of a timeline, how the different media devices (magic lantern, camera obscura, photograph, motion picture illusion technologies, cinematography and so on) had to be followed each other.
 - Ha a tudománytörténetet szélesebb kultúrtörténeti kontextusba ágyazva vizsgálja valaki (ahogy tette ezt Simonyi Károly a fizika kultúrtörténetéről szóló könyvében), akkor a társadalmi élet különböző színterein zajló események egymásra vetítése érdekes együttjárásokra, egymásra hatásokra deríthet fényt, és az egymástól távol eső szférák időbeli kötődésének "felmutatása" sokszor alkalmas lehet arra, hogy megérezzük, megérezzessük a korszellem, a Zeitgeist összetevőit, tartalmát.
 - Megfelelő fejlesztéssel a kronológia arra is alkalmassá tehető, hogy kezelni lehessen a "relatív időt", vagyis fel lehet kínálni azt, hogy a felhasználó "belebújjon valaki más életébe" és az ő időrendjén keresztül nézhesse az eseményeket.

1.7.2 Adatmodel

- Az események modellezésének fő kérdése az, hogy milyen belső struktúrát adunk az eseményeknek, pontosabban: van-e általános modellje az esemény fogalmának, aminek aztán lehetnek további altípusa, alosztályai. Egy demográfiai eseményt (például születés) nyilván más összetevőkre kell bontani, mint egy archontológiai eseményt (kinevezést egy testületi pozícióba) vagy egy háborús eseményt (mondjuk egy világháborús csatát vagy egy sokoldalú békeszerződés aláírását). Ebben a megközelítésben fordított irányban kell a modellezést indítani: ekkor a kulcsfontosságú kérdés úgy szól, hogy milyen főbb eseménytípusokat kell kezelnünk. Ha ezeket sikerül behatárolnunk, akkor ezek általánosításaként már eljuthatunk az általános eseménymodellig.
- Van azonban egy olyan modellezési lehetőség is, amelyben nem kell előzetesen rögzíteni az általános eseménymodellt, amely megengedi, hogy bármikor (időben később) lehessen új – tetszőlegesen összetett, akárhogy strukturált – eseménytípusokat felvenni a rendszerbe, természetesen anélkül, hogy a fogalmi modell alapsémáján lényegszerűen módosítani kelljen (Davidson 1967).
- A fogalmi modell középpontjában az a tétel áll, hogy az eseményeket reifikálni kell, vagyis minden eseményt példányosítani kell, ami megengedi azt, hogy tetszőleges fogalmi összetevővel bővíteni lehessen – akár menet közben is – az adott eseménytípus modelljét.
 - A problémát egy példával – predikátumlogikai szinten – szemléltethetjük. Vegyük a következő eseményt: 'János megkeni a kenyerét.' Ezt fogalmilag egy kétargumentumú predikátummal írhatjuk le:

megken(János, kenyér)
 - Ha azonban pontosítom az esemény leírását, amit többféleképpen tehetek meg, és azt mondom: 'János megkeni vajjal a kenyerét.' vagy 'János megkeni lassan a kenyerét.' vagy 'János 2013.03.30-án megkeni a kenyerét.' vagy veszem ezek aggregátumát,

akkor mindig új predikátumokat (fogalmi modelleket) kell felvennem. Például a “leghosszabb” predikátum így nézne ki:

megken(János,lassan,kenyér,vaj,2013.03.30)

- A fenti megoldás problémája az, hogy
 - itt előzetesen rögzíteni kell a leggazdagabb eseménytípus-modellt, illetve
 - nem lehet következtetni a specifikusabb predikátumokból az általánosabb predikátumokra vonatkozóan, pedig ezt nyilvánvaló módon meg kéne tudnunk tenni.

Ha tudom, hogy a

megken(János,lassan,kenyér,vaj,2013.03.30)

tartalmilag igaz, akkor abból nem tudok következtetni egy másik predikátum, a

megken(János,kenyér)

igazságára, hiszen ez két – egymástól független, más argumentumszámmal rendelkező – predikátum.

- A megoldás a következő. Bevezetjük az eseménypéldányok (e) fogalmát és ezekre vonatkoztatjuk az eseménypredikátumok (relációk) hatókörét. Ha vannak eseménypéldányok, amelyek különböző eseménypredikátumokkal írhatók le (különböző eseményosztályokba tartoznak), akkor:
 - az individuális eseményekre lehet kvantorokat alkalmazni
 - az individuális eseményeket több eseményosztályba be lehet sorolni
 - egy konkrét individuális eseményt úgy lehet leírni, hogy diszjunkcióval (ÉS-kapcsolattal) összekötjük azokat az eseményosztályokat, amelyek alá besorolható az adott eseménypéldány
- A fentiek azt jelentik, hogy az esemény leírásakor az összekapcsolt eseményosztályok mindegyikében szerepelni kell az individuális eseménynek. Nézzük meg, hogy lehet ilyen módon modellezni a fenti példánkat:
 - 1. megoldás:
$$\exists e(\text{megken}(\text{János},\text{kenyér},e) \wedge \text{kenőanyag}(\text{vaj},e) \wedge \text{kenési_idő}(\text{2013.03.30},e) \wedge \text{kenés_módja}(\text{lassan},e))$$
 - 2. megoldás:
$$\exists e(\text{kenő_ágens}(\text{János},e) \wedge \text{megkent}(\text{kenyér},e) \wedge \text{kenőanyag}(\text{vaj},e) \wedge \text{kenési_idő}(\text{2013.03.30},e) \wedge \text{kenés_módja}(\text{lassan},e))$$
- Az előző pontban látottak alapján látszik, hogy még így is szükség van modellezési döntésekre, lehet kétargumentumú összetevőkre bontani a sémát, lehet többargumentumú predikátumokat használni, de ez részletkérdés. Az a fontos, hogy nem kell elköteleződnünk az elején valamilyen általános eseménymodell mellett, hiszen ez a modellezési megoldás megengedi azt, hogy egymás mellett “éljenek” a különböző típusú események. Egy-egy eseménytípust természetesen le kell előzetesen modellezni, de:
 - az eseménytípusok egymástól független (különböző) sémákkal rendelkezhetnek
 - az eseménytípus-modelleket menet közben is lehet bővíteni

- csak arra van szükség, hogy minden eseménytípusra mindig rendelkezésre álljon egy olyan – up-to-date – programmodul, amely képes összerakni az eseményösszetevőkből az esemény teljes leírását. Ez a feltétel azt követeli meg a rendszertől, hogy ha módosítunk egy adott eseménytípus modelljén, akkor ennek megfelelően módosítsuk az aggregáló modul tartalmát is.
- A koncepcionális modell alapján könnyen kezelni lehet az időpontosság problémakörét. Egy esemény időértékét ugyanis nem egyetlen adatkapcsolattal (egyetlen adatmezővel) írhatjuk le, hanem a dátummeghatározás elemi összetevőit külön-külön adhatjuk meg (és annyival, amennyi adat az adott időpontosság mentén rendelkezésre áll). A különböző pontosságú adatokat összetevőkre bontva mindig meg lehet adni, ami alapján egy időaggregáló modulnak kell összeállítania a megjelenítés során a dátumot (és ugyanez a modul végezheti el az események időbeli rendezését is). Példák:
 - $\text{date}(e)=1989.10.13. 8:23 \rightarrow e_year(1989,e) \wedge e_month(10,e) \wedge e_day(13,e) \wedge e_hour(e,8) \wedge e_minute(23,e)$
 - $\text{date}(e)=1920\text{-as évek közepén} \rightarrow e_decade(1920,e) \wedge e_timemodifier(\text{közepén},e)$
 - $\text{date}(e)=\text{tizenkilencedik század elején} \rightarrow e_century(19,e) \wedge e_timemodifier(\text{elején},e)$
- Ez a modellezési koncepció nagyon rugalmas abban a tekintetben, hogy egy eseményről mindig csak annyi adatot kell beírni az adatbázisba, amennyi rendelkezésre áll. Nincsenek üres mezők. Viszont bármikor lehet kiegészíteni az esemény leírását, ha új adat kerül elő úgy, hogy egy új eseményosztály felvételével hozzákapcsoljuk az új adatot az adott eseménypéldányhoz.
- Ez a modellezési koncepció nagyon hasonlít a gráfrepresentációs megoldásokra.
- Ez a modellezési koncepció részlegesen hasonlít a NoSQL-alapú adatkezelési filozófiára.
- Ez a modellezési koncepció nagyon hasonlít az RDF-mondatok hármasságára, bár többféle RDF-mondattípust kell kezelni benne.
- Egy individuális eseményt tehát úgy írhatunk le, hogy az adott eseménypéldányt összekötjük vagy egy másik individuummal (emberrel, testülettel, helynévvel, időponttal, konkrét dokumentummal, filmmel, zeneszámmal, könyvvel stb.) vagy egy eseményontológiából származó fogalommal (lassan, vaj, kedvetlenül stb.).
- Amit a modellezés egyszerűségén és rugalmasságán nyerünk, azt picit meg kell fizetni azon az ágon, hogy fel kell építeni egy eseményontológiát, ami alapján az események részbeni minősítését el lehet végezni.
- A modell működhet individuum-névterek nélkül is, de a rendszer funkcionalitását, hasznosságát nagyságrenddel megnöveli, ha névterekből lehet kivenni az események individuumkapcsolatait. Ez több munkát kíván meg, de amiatt megéri, hogy az individuum-névterek használatakor az individuumok (személyek, testületek, helyszínek stb.) azonosítva vannak, és ennek köszönhetően ezekre vonatkozóan szűréseket, rendezéseket lehet elvégezni.
- A modellezési koncepció további nagy előnye még az, hogy könnyen lehet többnyelvűvé tenni az adatbázist (és ezzel együtt az alkalmazás egészét). Ha ugyanis az eseményösszetevőket is absztraháljuk, vagyis nyelvfüggetlen fogalomként kezeljük, akkor az eseményösszetevők listáját (vagy szerkezetét) tetszőleges nyelvre rá lehet vetíteni, és a lekérdezések során a kiválasztott nyelven lehet visszaadni az eseményösszetevő megnevezését. Ez három helyen követeli meg a fogalmi összetevők fordítását:
 - az eseményontológia elemeit kell fordítani (lassan, kedvetlenül stb.), tehát a predikátumok argumentumaiba kerülő fogalmakat,
 - az eseményösszetevők (predikátumok) neveit is le kell fordítani (megken, kenő_ágens stb.)
 - az eseményaggregáló modulnak is tudnia kell kezelni az összetevők nyelvi fordításait
- Az a koncepció rugalmas. Megengedi azt is, hogy:
 - egy adott eseménytípus modelljét lehesen menet közben változtatni (a sémaváltozásokat mindig követni kell az aggregáló modul változtatásával), illetve

- egy új eseménytípust is fel lehet venni bármikor a modellbe (ehhez esetleg bővíteni kell az eseményösszetevők, illetve az eseményontológia elemeit, illetve meg kell írni az új típushoz tartozó aggregáló modult.
- Ez a modell természetesen megengedi azt is, hogy bizonyos eseménytípusok esetében lehessen alosztályokat képezni és lehessen tulajdonságokat örököltetni a származási kapcsolatok mentén.
- Mindez praktikus szempontból azt jelenti, hogy el lehet úgy indítani az adatbázis (és alkalmazás) építését, hogy a leggyakoribb eseménytípusokkal kezdjük a modellezést, és ezekre vonatkozóan sok adattal fel lehet tölteni a rendszert, majd a következő körökben lehet az egyre bonyolultabb eseménytípusokat felvenni a rendszerbe.
- Az események tipizálásakor figyelembe vehetjük azt a klasszikus felosztást, amely szerint elkülöníthetünk **tevékenységeket**, **teljesítményeket**, **eredményeket**, valamint **állapotokat** (activities, accomplishments, achievements, states) egymástól (Ryle 1949; Vendler 1957, Kuti et al. 2006).

1.8 Írásműcímter

A nemzeti névterek körébe olyan név- vagy címtereket is be lehet sorolni, amelyek adott típusú kulturális dokumentumok azonosítására alkalmasak a dokumentumok címei és további adattípusok alapján. A filmes világ számára ilyen névtérkategória lehet a filmcímter, a könyvtári világ számára pedig az írásos dokumentumok címtere. Utóbbira nehéz jó címet adni, és nehéz a határait is megvonni, vagyis azt a kérdést megválaszolni, hogy milyen írásos dokumentumok kezelését kell megoldani egy ilyen írásműcímter keretein belül és milyeneket nem. Nem minden írásos dokumentum tartozik a könyvtári világ fennhatósága alá, hiszen a levéltárakban is alapvetően írásos dokumentumokat őriznek. Ha megtesszük azt a leegyszerűsítést, hogy a könyvtári kompetenciába – legalábbis első körben – a sokszorosított írásos dokumentumokat soroljuk be, akkor talán kicsivel könnyebb dolgunk lehet. Persze akkor olyan komoly filozófiai, modellezési problémákkal szembesülünk, amelyekkel nem találkozánk, ha a másolás jelenségétől eltekintենék. Az írásbeliség keretein belül létrehozunk **írásos dokumentumokat**. Az írásos dokumentum adott gondolati **tartalom** rögzítése valamilyen írástechnikával valamilyen **hordozóra**.

A dokumentum a tartalom és a hordozó egysége.

Hogy milyen és mennyi tartalmat szükséges rögzíteni ahhoz, hogy dokumentumról beszéljünk, illetve milyen hordozó esetén beszélhetünk egyáltalán írástechnikáról, nehéz kérdés, nincs egyértelmű válasz rá. Még inkább bizonytalanok lehetünk annaka kérdésnek a megválaszolásakor, hogy a dokumentumok közül miket tarthatunk (egyéni jellegű, valamilyen kreativitást megjelenítő) műveknek, és miket nem. A dokumentumok közül valamilyen társadalmi/értékelési mechanizmus révén kiválogatódhatnak azok a **művek**, amelyekhez szerzőket, kreatív alkotókat rendelhetünk, és szerzői jogokat tulajdonítunk nekik, illetve a művek azonosítása végett címetek adunk nekik. Egy következő lépésben pedig, annak érdekében, hogy ezek a művek minél több befogadóhoz eljuthassanak, sokszorosítjuk őket. Mivel ezek a sokszorosított írásművek bekerülnek abba a kulturális térbe, amelyekben a kulturális dokumentumok megőrzését, fenntartását közgyűjtemények (is) végzik, ezért ezen a ponton megjelenik az az igény, hogy az írásműveket is azonosítani tudjuk valamilyen írásműcímter segítségével. A kérdés már csak annyi, hogy pontosan mi is az, amit itt azonosítanunk kell.

Ha csak egyedi művek lennének a könyvtárakban (ahogy – kis túlzással – mondhatjuk, hogy a levéltárakban ez a helyzet), akkor könnyebb lenne megválaszolni azt a kérdést, miket és hogyan kell azonosítanunk az írásos dokumentumok világában. Minden egyes dokumentumot külön-külön kellene leírni és egyedi azonosítóval ellátni. A könyvtárak azonban sokszorosított műveket kell, hogy feldolgozzanak, leírjanak, amikor mondhatjuk, hogy a – gépi úton – sokszorosított írásművek (könyvek) nem különbözthetők meg egymástól, ezért ezeket egy csomagban kellene leírni valahogyan, nem pedig külön-külön. Természetesen a könyvtári világ már rég kitalálta a megoldást erre a problémára, hiszen a könyvtárak mindig is elkülönítették a könyvek kiadását, illetve az egyes kiadások konkrét,

individuális példányait egymástól. Ilyen módon a könyvtárban mindig is elég volt egyszer felvenni a könyv egy adott kiadásának adatait: a könyv címét, szerzőjét, kiadóját, kiadási idejét stb. Ha egy művet több kiadásban is kiadtak, akkor az nem jelentett különösebb pluszmunkát, ha az új kiadás esetében újra fel kellett venni az alapadatokat, hiszen ezekből nem volt olyan sok. Az új kiadások esetében voltak adatok, amelyek nem változtak az új kiadással (például a mű szerzője vagy címe), de a könyvtári világ számára vállalható "fölösleges teher" volt a már ismert adatok újbóli rögzítése. Azt a veszteséget is vállalhatónak tartotta a könyvtári világ hosszú időn keresztül, hogy a kiadvány-példány alapú modellben nem tudták igazán jól kifejezni a különböző kiadványok közti kapcsolatot.

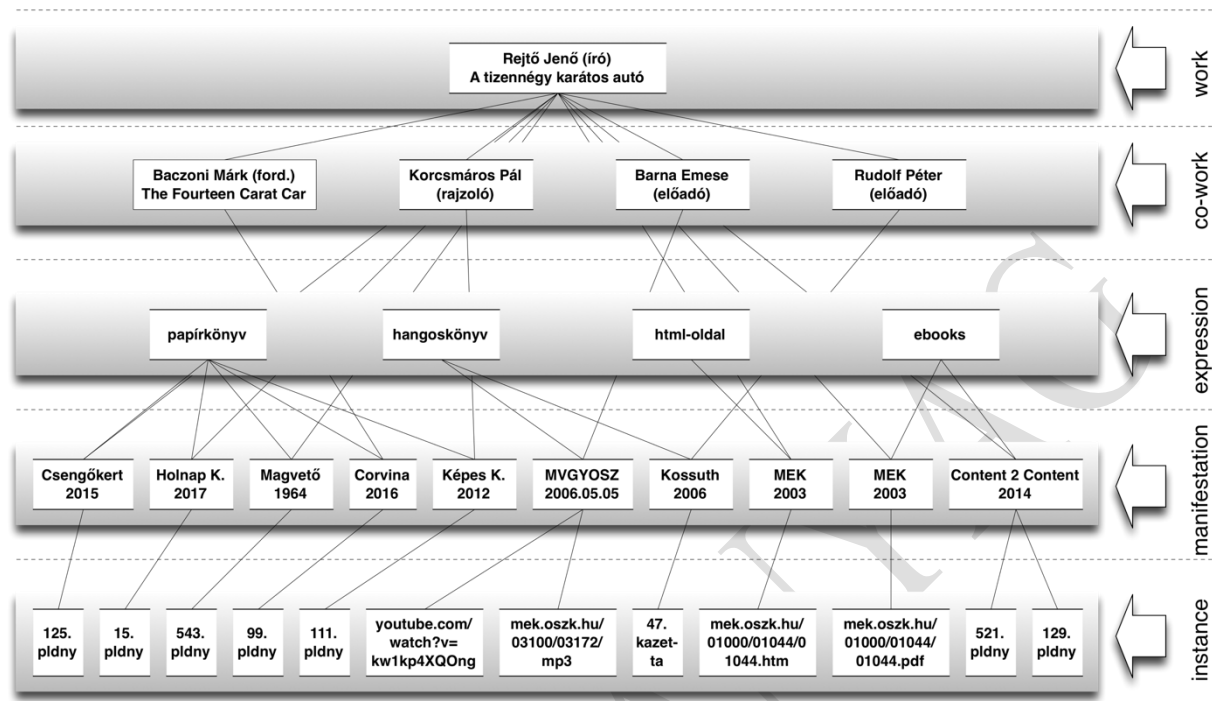
Már az 1990-es évek végén megszületett az az új elgondolás, az FRBR-modell, amely a dokumentumok többszintű leírását javasolta (FRBR 2008). Az FRBR négy szint elkülönítését javasolja, hogy egy egységes modellben lehessen kezelni a példányok (copy/instance) és kiadványok (manifestation) mellett a műveket (work) és a művek különböző kifejeződési formáit (expression). Utóbbi szint behozatalával lehet remélni azt, hogy össze tudjuk kapcsolni egy írásos mű (mondjuk: egy regény) hagyományos könyvkiadványait a mű más formátumú, más médiumokon keresztüli (például hangoskönyv vagy képregény alakban való) megjelenéseitől.

A négy szintű FRBR-modell sokkal jobban alkalmas arra, hogy megragadjuk vele a művek (tartalmak) többféle médiumon keresztüli nyilvánosságához közvetítéseket, illetve a különböző sokszorosítási eseményeken keresztüli megjelentetéseket, de azzal, hogy a modell elméletileg képes lekezelni a világ bonyolultságát, az adatrögzítés gyakorlati feladatát is bonyolultabbá teszi. Sok más tényező mellett ez a nehézség is hozzájárulhatott ahhoz, hogy a könyvtári világ sehol sem fogadta be és nem implementálta azonnal a modellt a hétköznapi munkájába. Az új, feltörekvő szabvány, a Bibframe ugyan elfogadta az FRBR-t a dokumentumok modelljeként, de csak három szintet (a work, a manifestation és az instance fogalmait) alkalmazza.

Pedig az igazán pontos modell felállításához nem hogy elvenni kéne belőle, hanem inkább még egy szintet bele kéne tenni a modellbe. Az FRBR ugyanis nem tudja elég finoman kezelni azt, hogy az eredeti műveket olykor – ugyancsak kreatív alkotói tevékenység révén – átalakítják, valamilyen értelemben más formátumra hozzák. A legnyilvánvalóbb példa erre az írásművek más nyelvre történő lefordítása és kiadása. Ilyen esetben azt sem mondhatjuk igazán, hogy az eredeti mű jelent volna meg egy új kiadványban, hiszen könnyen lehet, hogy a mű szerzője nem is tudná elolvasni, nem értené az idegen nyelvre lefordított művét, de nyilván azt sem mondhatjuk, hogy a fordítással egy új mű keletkezett volna. Ebbe a problémahalmazba tartozik az a jelenség is, amikor egy művet több fordító is lefordít egy másik nyelvre, és ilyenkor kellene tudni jelezni a kiadványok esetében, hogy melyik fordító fordításáról van szó. Ez a megkülönböztetési igény világosan jelzi a fenti problémát. Nyilván ugyanarról a műről van szó, hiszen mindkét fordítás/kiadvány esetében az eredeti mű megegyezik, de a két kiadvány mégis kétféle fordításban (valamilyen értelemben eltérő tartalommal, akár eltérő címeikkel) jelenteti meg az eredeti művet, kétféle szerzői jogi igény keletkezik a kétféle fordításért. Másik példa lehet a hangoskönyvek esete, amikor a szöveget különböző előadók olvassák fel, és ezzel másféle hozzáadott érték keletkezik. Az így kiadott hangoskönyvek nyilván eltérnek egymástól (bizonyos értelemben tartalmilag is), hiszen más színész (felolvasó) hangján szólnak meg. Nem elég tehát azt mondani, hogy ezeket hangoskönyvként elválasztjuk a papírkönyvektől (az expression szinten), és megengedjük, hogy a hangoskönyveknek legyenek egymást követő kiadásai, mert így nem tudjuk elválasztani egymástól a felolvasás révén keletkezett kreatív hozzájárulásokat, az eltérő felolvasásokat. A megoldás az lehet, hogy a mű (work) szintjéhez kapcsolódóan bevezetünk még egy második, co-work (társ-mű) szintet, ahol a származtatott, kiegészítő jellegű alkotói tevékenység eredményét is meg tudjuk ragadni (az eredeti műhöz hozzáragasztva, de egyben el is választva tőle). Javaslatunk szerint az FRBR eredeti négy szintje helyett helyesebb lenne öt szintet elkülöníteni egymástól (ez a megoldás pontosabb modellt eredményezne):

- work
- co-work
- expression
- manifestation
- instance

Az öt szintű modell értelmét (értelmességét) egy olyan ábrával szemléltetjük, ahol a megfelelő szinteken tüntetjük fel egy konkrét mű (Rejtő Jenő 'A tizennégy karátos autó' című regényének) legfontosabb adatait.



Az FRBR modell ahhoz nyújt segítséget, hogy milyen szinten kell megragadni a dokumentum fogalmát. Az írásos dokumentumokkal kapcsolatos másik fontos kérdés annak tisztázása, hogy milyen **szervezetű dokumentumról** van szó. Szervezetét tekintve a dokumentum lehet **összetett**, illetve **egyszerű** (vagy gyűjteményes), ami annyit jelent, hogy egy dokumentum (egy könyv, egy periodika adott száma) vagy tartalmaz több dokumentumot vagy nem, tehát vagy többedmagában vagy egymagában áll. Előbbi esetben összetett vagy gyűjteményes dokumentumról beszélhetünk, utóbbi esetben pedig egyszerű dokumentumról. Ha ezt a megkülönböztetést megtesszük, akkor összetett dokumentumként értelmezhetjük egy szerző tanulmánygyűjteményét, egy költő verskötetét és egyszerű dokumentumként az egyedi tanulmányokat vagy verseket, és természetesen meg kell engednünk (és a modellnek kezelnie kell tudni) az olyan eseteket, amikor egy tanulmány vagy egy vers több gyűjteményes dokumentumban is megjelenik. Ebből a szempontból tekintve a periodikumok mind gyűjteményes dokumentumoknak tekinthetők.

A harmadik dokumentumtipizálási szempont szerint arra is érdemes figyelni, hogy bizonyos dokumentumok úgy is összekapcsolhatók egymással, hogy nem egy kiadványban válnak egy összetett dokumentum részeivé, hanem időben, **sorozatként** kapcsolódnak össze egymással. A könyvsorozatokba foglalt egyedi kiadványok ilyenek, de a periodikumoknak is fontos tulajdonsága ez a sorozati jelleg. Fontos megjegyezni, hogy a sorozatiság is egyfajta gyűjteményességet jelent, csak itt az idő tengelyen kapcsolunk össze dokumentumokat egymással, míg az összetett dokumentumokban egy időben, egyszerre megjelent műveket fogunk össze. Mondhatjuk, hogy a gyűjteményesség kétféle értelme abban a különbségben ragadható meg, hogy a sorozatiság elve mentén az idő eltérő pontain megjelent dokumentumokat kapcsolunk össze, míg az összetett dokumentumok esetében az idő ugyanazon pontján kapcsolunk össze különböző dokumentumokat. A világot így értelmezve a periodikumok egyszerre összetett és sorozati dokumentumok, a gyűjteményes kötetet összetett dokumentumok, és a könyvsorozatok tipikusan sorozati dokumentumok.

Mivel a nemzeti névtér program első szakaszában nem elvárás, hogy az írásművek számára is kidolgozzunk egy modellt és implementáljunk egy önálló íráscímtér-szegmenst, ezért nem kell szembesülnünk azzal az ellenállással, nehézséggel sem, ami egy ilyen modell implementálásakor

felmerül. A személynévtér építése során azonban keletkeznek bibliográfiai adatok, amelyek írásműcímekre mutatnak, és ezek kezelésére egy írásműcímtér-prototípust érdemes felhúzni.

1.9 Köznévtér

A tulajdonnévterek mellett a névterek másik nagy típusát jelentik a köznévterek, amelyek általános fogalmakhoz kapcsolható nyelvi konstrukciókat fognak egybe. Az általános fogalmakhoz való kapcsolódás miatt a köznévterek logikája, struktúrája nagyon más a tulajdonnévterekhez képest. Amíg a tulajdonnévterek esetében szinte nem is beszélhetünk szemantikáról, addig a köznévterek esetében pont a névtér elemei közti struktúra szemantikája az igazán fontos és érdekes. Éppen ezen új minőség miatt a köznévterek tudásszervezési rendszerekként (KOS – Knowledge Organization System) is értelmezhetők, hiszen a rendszer elemei közti kapcsolatokkal mindig valamilyen ismeretterületre irányuló világtudásunkat fejezzük ki.

1.9.1 Tudásszervezési rendszerek

A tudásszervezési rendszerek abban különböznek egymástól, hogy a rendszer elemei között milyen relációkat engedélyeznek felvenni. Éppen ezért matematikai értelemben a tudásszervezési rendszerek matematikai struktúráknak tekinthetők.

Matematikailag a KOS a tárgyszavak tartóhalmazán értelmezett struktúra, amelyből természetesen többféle is lehetséges annak megfelelően, hogy milyen relációkat engedünk meg a tárgyszavak között. A struktúra az alábbi formában írható fel:

$$\text{KOS} = \langle D, R_1, R_2, \dots, R_i, R_n \rangle$$
, ahol D a tudásszervezési rendszer tárgyszavaiból álló tartóhalmaz, R_i az elemeken (tárgyszavakon) értelmezhető reláció.

A matematikai leírást azonban ki kell egészítenünk, ugyanis később látni fogjuk, hogy a tudásszervezési rendszerek különböző típusai nemcsak a tárgyszavaikban és a relációikban, hanem más jellemzőikben is eltérhetnek egymástól. Mielőtt azonban bővítenénk a tudásszervezési rendszer fogalmának összetevőit, a tudásszervezési rendszerek típusairól kell pár szót szólnunk. A különböző archívumépítési gyakorlatokban, a hálózati kultúra időszakát megelőző évszázadban, érdemben és kiterjedt módon háromféle tudásszervezési rendszert vettek használatba, majd a hálózati kommunikáció további típusok megjelenését tette lehetővé. Az alábbi öt KOS-típusról van szó:

- hagyományos archívumokban (is)
 - terminuslista
 - taxonómia
 - tezaurusz
- csak digitális archívumokban
 - ontológia
 - folkszonómia

Ezek mindegyike a rá jellemző matematika struktúrával, valamint a rendszerek gyakorlati működtetésére vonatkozó szabályrendszerrel írható le. A tudásszervezési rendszerek alaphalmaza (D) azokat a szavakat, kifejezéseket (terminusokat) tartalmazza, amelyeket a dokumentumokhoz lehet rendelni. A tudásszervezési rendszerek különbségeit az a tény határozza meg, hogy milyen – szintaktikai, szemantikai vagy pragmatikai – relációkat (R_i) engedünk meg felvenni a rendszer elemei között. Nem elég azonban csak a matematikai struktúrára figyelni, ha igazán meg akarjuk érteni a tudásszervezési rendszerek jelenségét. Azt is fel kell vennünk e rendszerek jellemzői közé, hogy van-e, és ha igen, akkor milyen felügyelet, milyen kontroll van a metaadat-hozzárendelési munka folyamatában. Előbb persze meg kell mondanunk, miért is van szükség ennek figyelembevételére. Nos, ha a tartalmi metaadatok dokumentumokhoz rendelésének az a fő funkciója, hogy egyértelműen jellemezzük általuk a dokumentumok tartalmát, akkor az egyértelműséget (vagyis a többértelműségek elkerülését) biztosítanunk kell valahogy. A többértelműségek elkerülése pedig megfelelő szakértelmet, fegyelmezett munkarendet, szakmai kontrollt, kontrollált szótárakat és tudásszervezési rendszereket

kíván. Ezt persze nem olyan könnyű formalizálni, hiszen olyan kérdésekre kell választ találnunk e szempont alapján, mint:

- $Q_1 = \{\text{támasztanak-e bármilyen szakmai feltételt, szaktudással kapcsolatos elvárást a munkát végzők személyével szemben?}\}$
- $Q_2 = \{\text{van-e bármilyen munkaszervezési szabályrendszer, ellenőrzési mechanizmus a munka menetére vonatkozóan, azaz kik, milyen jogosultságokkal vehetnek részt a munka egyes részfolyamataiban?}\}$
- $Q_3 = \{\text{kik rendelhetik a KOS-rendszer elemeit a dokumentumokhoz?}\}$
- $Q_4 = \{\text{kik szerkeszthetik, módosíthatják, bővíthetik a tudásszervezési rendszer elemeit, relációit?}\}$

A tudásszervezési rendszerek építésének és alkalmazásának kontrolljára vonatkozó fenti Q_i kérdésekre különböző válaszokat adhatunk, és ezt a feltételegyüttest vagyis az S_j társadalmi normák összefüggő rendszerét érdemes felvenni a tudásszervezési rendszer jellemzői közé. A dolgokat kissé leegyszerűsítve a következő tevékenységekre vonatkozó normákat kell rögzítenünk:

- S_1 – kinek szabad új tárgyszót létrehozni a tudásszervezési rendszerben
- S_2 – kinek szabad új relációt létrehozni a tudásszervezési rendszerben
- S_3 – kinek szabad két tárgyszót relációba állítani a tudásszervezési rendszerben
- S_4 – kinek szabad tárgyszót dokumentumhoz rendelni a katalógusban
- S_5 – csak a tudásszervezési rendszer elemeit szabad-e a dokumentumokhoz rendelni

Az öt norma közül az első három a tudásszervezési rendszerek építésével, az utolsó kettő a katalógusok bővítésével kapcsolatos. A számítógépek világában a fenti normák mind kezelhetők azáltal, hogy a digitálisan szabályozzuk, kinek van írási joga a tudásszervezési rendszer és/vagy a katalógus elemeire, illetve milyen adatokat lehet egymással összekapcsolni. Az írási, szerkesztési jogosultságokat is figyelembe véve már felírhatjuk a tudásszervezési rendszerek teljesebb formuláját, amelynek segítségével aztán majd megmutathatjuk az egyes típusok közti különbségeket is:

$KOS = \langle D, R_1, R_2, \dots, R_n, S_1, S_2, S_3, S_4, S_5 \rangle$, ahol D a tudásszervezési rendszer tárgyszavaiból álló tartóhalmaz, R_i az elemeken (tárgyszavakon) értelmezhető reláció ($i=1, \dots, n$) S_j a tudásszervezési rendszer

A fent bemutatott összetevőkkel már adott az az általános keret, amelyre támaszkodva elég pontosan megragadhatjuk a történelmileg létező, szélesebb körben elterjedt tudásszervezési rendszerek legfontosabb jellemzőit. Minden tudásszervezési rendszerben van egy közös reláció, a lexikografikus rendezés, amely a tárgyszóhalmaz elemeinek ábécé szerint való sorba állítását jelenti. A kérdés az, hogy melyek azok a relációk (és normák), amelyek alapján különbséget tehetünk az egyes típusok között. A példákat és részletes indoklásokat elhagyva az egyes tudásszervezési rendszerek az alábbiak szerint jellemezhetők (Syi 2011). A KOS-ok leírásához pontosan definiált relációkat kell használnunk, amelyeket feltüntetünk az alábbi táblában.

| | |
|---|--|
| R_1 lexikografikus rendezési reláció | R_{10} ekvivalenciareláció |
| R_2 ekvivalenciareláció | R_{11} címkegyakorisági függvény |
| R_3 hierarchikus alárendeltje (tartalmazás) reláció | R_{12} címke-címke együttjárési reláció |
| R_4 különbözőségi reláció | R_{13} felhasználó-címke együttjárési reláció |
| R_5 generikus alá- és fölérendeltje relációpár | R_{14} felhasználó-dokumentum együttjárési reláció |
| R_6 partitív alá- és fölérendeltje relációpár | R_{15} felhasználói aktivitás függvény |
| R_7 következménye-előzménye relációpár | R_{16} dokumentumcímkezési gyakorisági függvény |
| R_8 rokona (egyéb) reláció | R_{17} tetszőleges relációfogalom |

R₉ lásd és helyette szinonimareláció-pár

R₁₈ tetszőleges osztályfogalom

R₁₉ tetszőleges attribútumfogalom

A fent bemutatott keretrendszer, illetve a fenti táblázatban összefoglalt relációk alapján az egyes tudásszervezési rendszereket a következő relációk segítségével definiálhatjuk.

1.9.1.1 Terminuslista

$KOS_{list} = \langle D, R_1, R_2, S_1, S_4, S_5 \rangle$, ahol
R₁ lexikografikus rendezési reláció
R₂ ekvivalenciareláció

1.9.1.2 Taxonómia

$KOS_{tax} = \langle D, R_1, R_3, R_4, S_1, S_3, S_4, S_5 \rangle$, ahol
R₁ lexikografikus rendezési reláció
R₃ hierarchikus alárendeltje (tartalmazási) reláció
R₄ különbözőségi reláció

1.9.1.3 Tezaurusz

$KOS_{tez} = \langle D, R_1, R_5, R_6, R_7, R_8, R_9, S_1, S_2, S_3, S_4, S_5 \rangle$, ahol
R₁ lexikografikus rendezési reláció
R₅ generikus alá- és fölérendeltje relációpár
R₆ partitív alá- és fölérendeltje relációpár
R₇ következménye-előzménye relációpár
R₈ rokona (egyéb) reláció
R₉ lásd/helyette szinonimareláció-pár

1.9.1.4 Formális ontológia

$KOS_{tez} = \langle D, R_1, R_5, R_{17}, R_{18}, R_{19}, S_1, S_2, S_3, S_4, S_5 \rangle$, ahol
R₁ lexikografikus rendezési reláció
R₅ generikus alá- és fölérendeltje relációpár
R₁₇ tetszőleges relációfogalom
R₁₈ tetszőleges osztályfogalom
R₁₉ tetszőleges attribútumfogalom

1.9.1.5 Folkszonómia

$KOS_{folk} = \langle D, R_1, R_{10}, R_{11}, R_{12}, S_4 \rangle$, ahol
R₁ lexikografikus rendezési reláció

R₁₀

ekvivalenciareláció

R₁₁

címkegyakorisági

reláció

R₁₂ címke-együttjárasi reláció

1.9.2 A köznévterek és a tudásszervezési rendszerek kapcsolata

A köznévterekben döntően általános fogalmakat kezelünk, és mivel ugyanez igaz a tudásszervezési rendszerekre is, ezért tisztázni kell, hogy mit értünk a nemzeti köznévtér fogalma alatt. Olyan egyértelmű válasz sajnos nem adható erre a kérdésre, mint ahogy meg lehet válaszolni azt, hogy miként lehet elhatárolni az egyes tulajdonnévszegmenseket egymástól a nemzeti névtér egész rendszerén belül. Általános fogalmakra, köznevekre szükség van a tulajdonnevek kezelése során is, hiszen különböző módon kell tudnunk tipizálni az egyes névhordozókat, neveket, illetve a nevek, névhordozók közt levő kapcsolatrendszereket. Ezt a tipizálási igényt valamilyen köznévtérszegmens (típuslisták) segítségével lehet kielégíteni. Ennek megvalósítását a nyelvi modulban tárolt fogalmak megfelelő típuslistákba, fogalomgyűjteményekbe rendelésével biztosíthatjuk. Az e célból felhúzott általános fogalmi rendszer azonban általánosabb célokra is használható. Amennyiben az összes, a tulajdonnévterek felől érkező típuslista-kezelési elvárást meg akarjuk oldani, úgy egy olyan általános fogalmi rendszert kell felépítenünk, amely minden jelenbeni és potenciális jövőbeni általánosfogalom-kezelési igényt képes kielégíteni. Ha egy ilyen rendszert építünk, az már arra is képessé válik, hogy tetszőleges tudásszervezési rendszert is lehessen benne kezelni. Ezeket a KOS-okat azonban nem tudjuk olyan módon elválasztani egymástól, ahogy határt húzhatunk két tulajdonnévtér-szegmens közé. Egy személynévtér és egy földrajzi névtér – a bennünk tárolt adatok alapján – egyértelműen és elég könnyen elhatárolható egymástól, míg két tezauszus esetében ezt már nehezebb megtenni. Ez az elhatárolási kérdés a gyakorlatban könnyen megoldható, hiszen az általános fogalmakat tartalmazó tudásszervezési rendszereket a tartóhalmazaik, valamint a relációik révén egyértelműen megkülönböztethetjük egymástól. Praktikusán tehát könnyen megoldható a probléma, a Nemzeti Névtér projekt számára pedig nem életbevágóan fontos az elméleti kérdés végleges tisztázása.

1.10 Névterek közti kapcsolatok

A nemzeti névterek névhordozókat (személyeket, testületeket, földrajzi helyeket stb.) és a névhordozók neveit (személyneveket, testületi neveket, földrajzi neveket stb.) kezelnek abból a célból, hogy egyértelműen meg lehessen mind az individuális névhordozókat, mind azok megnevezéseit egymástól. A névhordozók egyértelmű azonosítása során sok esetben olyan adatokat használhatunk fel, amelyek a névhordozók valamilyen egymáshoz való viszonyát fejezik ki. Mivel egyfelől a névtér-használati munka során szükség lehet az ilyen adatokra, másfelől ezek a "kapcsolati" adatok ugyanolyan fontosak és érdekesek lehetnek a kulturális emlékezet szempontjából, mint maguk a névtérbe épített adatok, ezért érdemes ezeket önállóan kezelni, önállóan megjeleníteni a közönség számára.

1.10.1 Archontológia

Az archontológia a történettudomány egyik fontos segéttudománya. Ez az ismeretterület azonban – a szaktudományi használaton túl – általánosabb kulturális diskurzusokban is érdeklődésre számíthat, mert azok az adatok, amelyeket az archontológiákban tartanak nyilván, sokakat érdekelnek. Az archontológia a személyek, a testületek és a testületen belül betöltött beosztások, pozíciók időben változó kapcsolatát írja le. Klasszikus példája a kormányzati archontológia, ami adott ország minisztériumainak, minisztereinek neveit, a miniszteri kinevezések, felmentések idejét tartja nyilván. Ezt általánosítva tetszőleges testülettípus esetében megtehetjük azt, hogy a szervezeti pozíciókon, beosztásokon keresztül időben összekapcsoljuk a testületeket és személyeket egymással. Ezzel voltaképpen a munkahelyek, intézmények tartós kapcsolatrendszerét ragadhatjuk meg. Amikor az életrajzokban felsorolják, hogy a

bemutatott személy mikor, milyen szervezeti pozíciókat töltött be élete során, akkor archontológiai adatokat kezelnek.

Az archontológia két névtérsezmens, a személynévtér és a testületi névtér közti kapcsolatokat írja le úgy, hogy közben az eseménytér segítségével helyezi el az időben ezt a kapcsolatrendszert.

1.10.2 Iskolázottság

A személyek életrajzaiban másik gyakori, "kötelező" elem az emberek oktatási története: mikor, milyen formában, melyik oktatási intézménnyel volt kapcsolata valakinek. Ez a kapcsolatrendszer tekinthető archontológiának, hiszen ebben az esetben is testületek (oktatási intézmények) és diákok időben változó kapcsolatáról van szó. Ez a hasonlóság megengedi azt, hogy az adatmodell szintjén ugyanúgy kezeljük az oktatási, iskolázottsági, mint a munkahelyi adatokat, de a felhasználói felületeken érdemes ezt a két adatkört elkülönítve kezelni (mind az admin, mint a megjelenítő rétegben).

1.10.3 Társadalomföldrajz

Az emberek a földrajzi térben szétszórva élnek. A földrajzi helyek egy fontos típusát jelentik a lakott területek, a települések, és ezek fontos minősége az, hogy közigazgatásilag kiemelten kezelik őket. Ez a kitüntetett figyelem megnyilvánul abban (is), hogy településekre vonatkozóan a közigazgatás által gyűjtött, erre a területi szintre aggregált adatok állnak rendelkezésre. Ezek az adatok a földrajzi helyekhez rendelhetők, és földrajzi helyek névtéroidalain meg is jeleníthetők. A társadalomföldrajzi adatok megjelenítése a földrajzi névtér oldalain olyan lehetőséget kínál a földrajzi helyek (települések) alaposabb leírására, amit érdemes kihasználni. A névterek szempontjából tekintve a társadalomföldrajzi adatok az adott földrajzi hely, valamint személyek vagy testületek valamely – az adott területhez köthető – csoportja közti, számszerűsíthető kapcsolatot jellemzik. Ilyen adattípus lehet az adott településen működő cégek száma, az ott lakók által befizetett adó nagysága, a könyvtári kölcsönzések, az iskolák vagy a virágboltok száma.

Az ilyen esetekben a földrajzi helyek és a testületekkel vagy személyekkel valahogyan kapcsolatba hozható adattípusok időben változó kapcsolatáról van szó, és ezek nyilván a földrajzi helyek leírását teszik bővebbé, tartalmasabbá. Ezen adatkör absztrakt kapcsolatrendszere az alábbiak szerint jellemezhető:

- geoid, eventid, datatypeid, value

Mivel ezek a – névtérközi – individuális adatok egyik dimenzióban a térhez (földrajzi helyekhez), a másikban az időhöz (eseményekhez) vannak kapcsolva, többfajta megjelenítési mód elképzelhető (és kívánatos) az esetükben. Az ilyen adatokkal leírható jelenségek térhez való viszonyát azzal fejezhetjük ki a legnyilvánvalóbban, hogy a földrajzi térhez kötött adatok kiszolgálására alkalmas térinformatikai eszközöket használunk, vagyis digitális térképeken jelenítjük meg az adatokat. Az adatok viszont rendelkezésre állnak és megjeleníthetők táblázatos formában is, ami további számítások elvégzését is lehetővé teszi. A térben és időben megfelelő módon (névterek segítségével) összekapcsolt adatköröket meg lehet jeleníteni az időtengely mentén érzékeltethető változásokra fókuszálva, tehát lehet idősoros grafikonokat és/vagy táblázatokat is előállítani. A névtér egészét kiszolgáló informatikai infrastruktúrának képesnek kell lennie arra, hogy az itt jelzett funkcionalitásokat mind kezelni tudja. Ezek az adatok tipikusan külső adatforrásokból származnak, és egyedi, nagytömegű adatfeltöltés révén kerülhetnek be a rendszerbe.

1.10.4 Választás

A társadalomföldrajzi adatok egy speciális csoportját jelentik a választási adatok, amelyek a földrajzi helyekhez kapcsolhatók. A választási adatokat úgy értelmezhetjük, hogy az emberek valamely – általában az adott területen lakó – csoportja által kifejezett preferenciajelzés testületek és/vagy személyek iránt. A választási adatok mindig egy adott település (önkormányzat) szavazókörében leadott egyéni szavazatok összesítését jelentik. A szavazatoknak – névterek szempontjából tekintve – kétféle típusát különíthetjük el: lehetnek egyéni jelöltekre (tehát személyekre), valamint pártokra (tehát

testületekre) leadott szavazatok. Ezek alapján a választási adatok absztrakt kapcsolatrendszere így írható fel.

- geoid, eventid, datatypeid, agentid, value

A választási adatok modellezéséhez szükség van még arra, hogy számon tartsuk a pártok és a jelöltek közti kapcsolatot is. Ennek a kapcsolatnak a nyilvántartására alkalmas az archontológiai adatszegmens, hiszen a választási eseményekhez kapcsolódóan számon kell tartani azt, hogy adott párt (testület) és személy (jelölt) között milyen szerep (pártjelöltség) létezik. Ez felfogható archontológiai adatnak. Ezek az adatok tipikusan külső adatforrásokból származnak, és egyedi, nagytömegű adatfeltöltés révén kerülhetnek be a rendszerbe.

1.10.5 További névtérközi adatkörök

A fenti példák azt mutatják, hogy a különböző névterekbe tartozó individuumok közti kapcsolatokra fókuszálva további fontos és érdekes adatköröket lehet a névterek közé beilleszteni, amelyeket aztán a névterekben is meg lehet jeleníteni (különböző nézetekben). A névtér adminrendszerét fel kell készíteni arra, hogy lehessen további névtérközi adatkapcsolatokat definiálni a rendszerben.

2 A névterek fogalma

2.1 A névterek fontosságáról

A digitális technológiák elterjedése komoly kihívások elé állítja a közgyűjteményeket, könyvtárakat, kulturális adatszolgáltatókat. A legelső feladat természetesen az, hogy a kulturális értékeket át kell emelni a digitális hálózati térbe. Amint megvalósítjuk ezt a feladatot, a kulturális dokumentumaink elérhetővé válnak bárhol, bármikor, bárki számára. Ez a hozzáférés azonban csak elvi lehetőség marad mindaddig, amíg nem biztosítjuk azt, hogy a kulturális dokumentumokat kereső felhasználók könnyen és gyorsan megtalálhassák mindazt a digitális térben, ami érdekli őket. Ehhez meg kell megtalálni, ki kell fejleszteni és fent kell tartani a hozzáférés gyakorlati biztosítékait, technológiáit. A hálózati világban kezdetektől fogva kiemelt szerepet vittek azok a szolgáltatók, amelyek ezt tudták valahogyan biztosítani. A hozzáférés gyakorlati lehetőségét a mai napig a keresőszolgáltatások képesek felkínálni a legnagyobb sikerrel. Látni kell azonban azt, hogy a népszerű globális keresőszolgáltatók (mint például a Google vagy a Bing) elsősorban olyan információs térben működőképesek, amelyek alapvetően a dokumentumok tartalmában való keresés lehetőségét nyújtják. A közgyűjteményi világ szereplői azonban már évszázadok óta olyan információs rendszereket építenek és kezelnek, amelyekben sosem a – domináns mértékben írásos – dokumentumok tartalmában való szabadszavas keresés volt a lényeg. A közgyűjtemények (könyvtárak, levéltárak, múzeumok, zene- és filmtárak, audiovizuális archívumok) kezdetektől fogva a dokumentumok, kulturális termékek, kulturális artefaktumok metaadataiból építenek rendszereket. A strukturált adatrendszerekben való kereséssel kapcsolatban azonban még sok nyitott kérdés létezik, és még több az előttünk álló, még el nem végzett fejlesztési feladat. A nyitott kérdésekre adható válaszok tudományos kutatást, a metaadatok alapján való eligazító szolgáltatások létrehozása pedig komoly fejlesztési munkát kíván meg.

A metaadatok fontossága a nem-szóveg alapú, audiovizuális archívumok világára mutatva is látható. Amilyen mértékben kerülnek bele az audiovizuális tartalmak az interaktív hálózat világába, olyan mértékben erősödik a felhasználói igény az ilyen művek kereshetőségére. Ennek kielégítésére a szolgáltatók által megvalósított, közös metaadatkezelés a jó válasz. Ebben a kollaborációban kiemelt szerepet játszanak a névterek, amelyek révén egyértelműen azonosíthatók a keresett/kereshető művek. Hogy miért van szükség az egyértelmű azonosítására? Rossz példaként hivatkozhatunk a YouTube gyakorlatára, ahol ha rákeresünk egy filmre, akkor annyi találatot kapunk, ahányszor a felhasználók feltöltötték az adott filmet, de azt, hogy vajon ugyanarról a filmről van-e szó vagy sem, azt csak akkor tudjuk megállapítani, ha egyenként és alaposan megvizsgáljuk az összes találatot.⁴ Arra sem képes a YouTube (a látható szolgáltatási térben), hogy egy adott filmet összekapcsoljon más szolgáltatók (mondjuk az IMDb) adataival. Amennyiben a feltöltött YouTube videókhoz valamilyen névtérből hozzá lennének rendelve olyan azonosítók, amelyeket más szolgáltatók is hozzárendeltek már a saját adataikhoz, akkor e közös névtér-azonosító mentén a szolgáltatások összekapcsolhatókká válnának, és a felhasználók lehetőséget kapnának, hogy egyetlen kattintással „ugráljanak” a különböző oldalak között. A névterek fő funkciója tehát a szolgáltatások közti interoperabilitás megteremtése.

Természetesen a nagy keresőszolgáltatók régóta foglalkoznak azzal, hogy milyen módokon, milyen eszközökkel lehet meghaladni a szabadszavas keresők régóta ismert hiányosságait, gyengeségeit. Abban nem nagyon van vita, hogy a továbblépés lehetőségét a szabadszavas keresés szemantikai vakságának megszüntetése adhatja, de abban a kérdésben, hogy mindezt hogyan lehet megvalósítani, még nem nagyon alakult ki konszenzus. A **szemantikus keresés** kiemelt fontosságát fogalmazta meg már a 2000-ben meghirdetett **szemantikus web** projekt, ami az azóta eltelt közel két évtizedben még nem jelent meg senki sem átütő szolgáltatással e téren. Azt azért érdemes megjegyezni itt, hogy a Google már a kétezres évek elejétől fogva folyamatosan építi azokat a szemantikus keresőmodulokat, kérdés-válasz

⁴ Az más kérdés, hogy a YouTube – a „háttérben” zajló képi elemzések eredményeként – tudja, tudhatja, hogy két feltöltött videó ugyanannak a filmnek a két példánya. Ebből a felhasználó nem érzékel semmit.

(Q&A) rendszereket, amelyek jól behatárolható, szűk területeken már képesek szemantikus szolgáltatásokra, és ezeket a modulokat be is építi a nagy keresőszolgáltatásába, de az egységes, integrált, széles információs tartományban keresni képes szemantikus keresővel még a keresőszolgáltatások világában piacvezető cég sem tudott előállni.

A kétezres évek végétől egyre divatosabbá váló és fokozatosan terjedő, RDF-triplet logikára épülő, **linked data** alapú elképzelések a szemantikus web filozófiáját viszik tovább, miközben egyre hangsúlyosabbá válnak az ontológiák, névterek építésének szükségességét hangoztató nézetek. Ebben az új megközelítésben, ezen új technikai paradigmán belül tudományos viták során kell tisztázni a közeljövőben azt a kérdést, hogy mi a pontos fogalmi és technológiai kapcsolat a névterek és az ontológiák (illetve más tudásszervezési rendszerek) között. A névterek esetében viszont már most egyértelmű, hogy azok legfontosabb funkciója az lehet, hogy segítségükkel biztosítani lehet a hálózatra került nagy adatrendszerek közti keresztlinkek elhelyezését, és ezáltal a rendszerek közti átjárhatóságot, interoperabilitást.

A jelen projekt felfogásában a – szűkebb értelemben vett – névterek rendszerezett, egyértelműsített, egyedi azonosítóval ellátott, tipizált tulajdonnév-állományok. A közgyűjtemények, a tudományos, oktatási és társadalmi szervezetek által fenntartott kulturális adatbázisok névállományai (személynevek, testületi nevek, földrajzi nevek stb.) közösek, nagy mértékben egymást átfedők. Egy személy, egy intézmény, egy földrajzi hely sokféle módon kapcsolódhat a könyvekhez, újságokhoz, filmekhez, zenékhez, képekhez, egyszóval a kulturális dokumentumokhoz. A nevek egységes, integrált kezelését, egyértelmű azonosítását névterek segítségével lehet biztosítani. A névterek lehetővé teszik a kulturális intézmények közti kollaborációt, biztosítják a kulturális adatbázisok, szolgáltatások közti interoperabilitást, az átfedő adatszegmensekkel rendelkező információs rendszerek közti sokszálú kapcsolatrendszer kiépítését, a nevek egységes és integrált kezelésével összekapcsolják és könnyen bejárhatóvá teszik a magyar kultúra dokumentumait, lexikonjait, kézikönyveit.

Vannak már névterek a világban globális, nemzeti és lokális-intézményi szinten egyaránt, de ezek mind típusonként külön állnak, nincsenek érdemi kapcsolatban egymással, ezek a rendszerek még nem integrálódtak.⁵ A cél most az, hogy ezeket közös platformra hozni és egységes infrastruktúrával kell őket kiszolgálni – egyelőre nemzeti keretekben gondolkozva, de természetesen minden fejlesztést úgy tervezve és végezve, hogy a globális névterekhez való kapcsolódást is biztosítani tudjuk az első lehetséges alkalommal.

A névterek használata kompatibilis a szemantikus web, a linked data világával is, ahol az az elvárás, hogy a dokumentumokat, ágenseket (személyeket, szervezeteket), helyeket mind egyetlen azonosítóval (URI-val) kell leírni.

A hazai jogszabályok az Országos Széchényi Könyvtárat jelölik meg a nemzeti névterek építéséért felelős intézményként. A nemzeti névterek felépítésének munkálatai megindultak, és már a kezdetek kezdetén megmutatkoztak a közös névtérhasználat előnyei és eredményei. Bármilyen filozófia mentén is építsenek archívumot a saját maguk számára a hazai könyvtárak, levéltárak, múzeumok, audiovizuális archívumok, mindenképpen lokális névtereket kell építeniük, és érdemes csatlakozniuk a nemzeti névterek adatállományaihoz is. A nemzetközi névterekhez való kapcsolódást a nemzeti névtereken keresztül lehet megvalósítani, mert a globális kooperáció megteremtését (az érthető okokból felmerülő kontrolligény miatt) a nemzeti névtereken keresztüli együttműködések rendszerétől lehet várni (nem pedig a lokális névterek globális kollaborációjától).

A gyakorlatban működőképes globális kollaborációs lehetőségek megteremtéséhez szükség van a nemzeti (és természetesen globális) névterekre, de ezek mellett kellenek a lokális, intézményi névterek is. A háromszintű névtérrendszerre azért van szükség, mert könnyen felmerülhet az a helyzet, hogy nem kell minden adatot a nemzetközi, globális névterekben kezelni, mert ez szétfeszíthetné azok kereteit. A nemzetközi névterekbe nem érdemes minden, az adott nemzet kultúrája számára fontos névelemet felvenni, mint ahogy a nemzeti névterekbe sem érdemes (nem is lehet) minden, a lokális, intézményi névterekben tárolt adatot átemelni. Konkrét példakkal élve: egy mozgóképes nemzeti névtérbe fel lehet venni az összes olyan dokumentumfilm, kisfilmet, amelyet valamilyen szempontok szerint

⁵ A VIAF, az ISNI a személyek, testületek, helyek globális azonosítására szolgál. Az összerendelt azonosítókat, a hozzájuk kapcsolható adatokat a nemzeti könyvtárak által beküldött besorolási adataiból (authority állományokból) állítják össze. Az ISAN az audiovizuális művek nemzetközi azonosítását végzi.

megőrzendő hungarikumnak minősítünk, de ezek mindegyikét nem kell (sőt, nem is lehet) beregisztrálni a globális névtérbe, vagy egy médiaképzésben keletkezett vizsgafilm fel lehet venni a képzést végző intézmény névtérébe, de nem feltétlenül kell betenni (befogadni) a nemzeti névtérbe is.

2.2 Létező névtérhasználati gyakorlatok

A névtérek építésével kapcsolatos fogalmak, definíciók, elvárások, szabályok áttekintése előtt érdemes alaposabban feltárni, bemutatni, hogy milyen névtérek, névtérhasználati gyakorlatok léteznek szerte a világban, azokat milyen célokra használják, milyen módszerek mentén üzemelnek, mit lehet tanulni tőlük. A névtérek jelentőségét egyre jobban el- és felismerik a hálózati világ, a digitális kultúra terjedésével. Globális, nemzeti és lokális névtérek épülnek szerte világban. Mielőtt tisztáznánk, mit is jelent (és mit nem) a névtér fogalma, bemutatunk pár speciális névtérrel, hogy megnézhessük, milyen célok mentén, hogyan működnek.

2.3 Fájlnevek

Talán a legegyszerűbb, de mindenképpen a legtöbbször használt névtér a számítógép fájlrendszere, amelynek egyetlen feladata azt biztosítani, hogy a számítógépen minden állomány egyedi névvel rendelkezzen. Az operációs rendszertől függ, hogy milyen szabályok mentén lehet neveket létrehozni, de ez másodlagos a névtér működése szempontjából. Az egyedi nevek kiosztását viszont megkönnyíti az a lehetőség, hogy a neveket komponensekből lehet felépíteni. Az állományokat csoportokba (folderekbe, könyvtárakba) lehet gyűjteni, az ilyen gyűjtőmappákat egymásba lehet ágyazni, és a tárolóknak is nevet adva úgy lehet biztosítani az állomány nevének egyediségét, hogy az állományt tartalmazó gyűjtőmappák neveit is beépítjük a teljes név összeállításába. Ez egy hierarchikus névazonosítási módszer – a hierarchikus vezérlés minden előnyével és hátrányával együtt. A hierarchikus szerveződés nyelvi szempontból azt jelenti, hogy a neveket névelemekként értelmezhetjük, vagyis egy nevet mint névelemet többször is felhasználhatunk egy konkrét, fájlnev megkomponálásakor. Erre azért van lehetőség, mert az állománynevek összetett nevek, vagyis a teljes – és a teljességében egyedi – állománynév legalább annyi névkomponensből áll, ahány könyvtárba ágyazódik maga az állomány a fájlrendszer gyökeréhez képest. Ezek a névkomponensek azonban nem kell, hogy önmagukban egyediek legyenek, ezek ismétlődhetnek. Bármilyen névelemet többször is felhasználhatunk a hierarchikus névkomponálás során (minden egyes könyvtáron belül újra és újra előfordulhat ugyanaz a névelem).

Azzal a követelménnyel, hogy minden állománynak legyen neve és ez a név egyedi legyen, azt lehet biztosítani, hogy minden állomány egyértelműen azonosítható és biztosan megtalálható legyen a számítógépen. Ez az elvárás megengedi azt, hogy a fájlrendszerben egy állománynak lehessen több neve is, hiszen ha a névegyediség biztosítva van, és minden név össze van kapcsolva egy állománnyal, akkor a nevek felől mindig minden állomány elérhető (fordítva ez nem lenne igaz, de olyan igény nem jelentkezik, hogy egy állomány felől indulva kellene egy adott nevet megtalálnunk). Nyelvtani szempontokat érvényesítve azt mondhatjuk, hogy ha két név ugyanahhoz a dologhoz (itt: állományhoz) tartozik, akkor ebben a névtérben a szinonimitás megengedett, de ha egy név csak egyetlen dologhoz (állományhoz) tartozhat, akkor a homonimitás nem megengedett.

Fontos még megjegyezni, hogy a fájlrendszer, mint névtér, lokális abban az értelemben, hogy minden számítógépen van (legalább) egy fájlrendszer, tehát ilyen típusú névtérből legalább annyi van, mint amennyi számítógép. Ezek a lokális névtérek persze – hálózati kapcsolat mentén – egybeszervezhetők, és ezáltal terjedelmüket tekintve nagyobb névtérek is kialakíthatók, de az egyediség ekkor is könnyen fenntartható, csak az összekapcsolt gépeket kell ellátni egy egyedi azonosítóval. A fájlrendszer esetében egyszerű a válasz arra a – minden névtér esetében fontos – kérdésre, hogy ki kezelheti (hozhatja létre, törölheti, változtathatja meg) a neveket. A válasz (a dolgokat kissé leegyszerűsítve): a számítógép kezelője.

2.4 Domainnevek

A fájlrendszerhez hasonló elvárás, ti. a nevek egyediségének és nevekkel jelölt dolgok (számítógépes erőforrások) egyértelmű megtalálhatóságának elvárása fogalmazódik meg a webes domainnevek rendszerével mint névtérrel kapcsolatban. Akár a világháló bejegyzett domainneveire, akár a weboldalak címeinek hierarchikusan strukturált halmazára gondolunk, mindkét esetben névtérről beszélhetünk (úgy, hogy utóbbi tartalmazza az előbbit). Mi a domainnevek fő funkciója? Egyértelműen hivatkozható, ember által könnyebben memorizálható karaktersorozatokkal jelölni a weboldalak címeinek közös gyökereit (index.hu, microsoft.com, mit.edu). A domainnevek képzésére vannak szabályok (mit lehet, mit nem lehet névelemként alkalmazni, a névelemek között szeparátorokat kell alkalmazni stb.), és van egy alapvető elvárás a nevek egész halmazára vonatkozóan: minden névnek egyedinek kell lennie. Ezt azáltal lehet biztosítani, hogy a feladatot földrajzi területek szerint szétosztva, regisztrátor szervezetek kezébe teszik le a domainnév-bejegyzés jogát, akiktől elvárják az egyedi névkiosztás elvének betartását a felelősségükbe tartozó névtartományban (ez a névadási rendszer is hierarchikus szerveződésre épül).

A domainnévtér fenntartására első szinten azért van szükség, hogy egy új névbejegyzési igény esetében lehessen tudni, létezik-e az igényelt domainnév a névtérben vagy sem. Ha a névtérben már bent levő névről van szó, akkor nem lehet bejegyezni, ha az igényelt név még nincs a névtérben, akkor bejegyezhető (ha a névadási szabályoknak megfelel). Mit jelent a bejegyzés? A domainnevek és az őket bejegyző (és innentől fogva birtokló) névtulajdonosok közti kapcsolat rögzítését. A másik fontos funkciója a domainnévtérnek az, hogy egyértelmű kapcsolatot teremtsen a domainnevek és a gépek által kezelt, gépekhez tartozó IP-címek között.

A domainnévtér tehát hierarchikus rendbe szervezett, egyedi nevek, valamint a nevek és a tulajdonosok, illetve nevek és a fizikai címek közti dinamikus kapcsolatok naprakész nyilvántartása. A domainnevek egyedisége egyfelől biztosítja az általuk "jelölt" IP-címek egyértelmű beazonosíthatóságát, másfelől a regisztráció a domainnév használati jogát is egyértelműen rögzíti. Mivel a domainnévtér legfőbb funkciója a világháló összes kommunikációs csomópontjának egyértelmű azonosítása, nyilvánvaló, hogy ez egy globális névtér. Ez a névtér engedélyezi a szinonimitást, de tiltja a homonimitást, és – a fájlrendszerhez képest kisebb mértékben ugyan, de – használja az összetett név komponálásának eszközét is, hiszen a domainnevek összetevőkre bonthatók, és egyes összetevőkben lehet ugyanazokat a névelemeket is újra alkalmazni.

2.5 Emailcímek

Ahogy a távközlési szolgáltatások döntő részében, úgy az emailcímek világán belül is egy egyértelmű címzési rendszer fenntartására van szükség. Egy-két szemponttól eltekintve mondhatjuk, hogy az emailcímeket értelmezhetjük a domainnév-rendszer részeként is, hiszen az egyedi emailcímek kiosztását, illetve a levelek eljuttatását a címzethez a hierarchikus módon egyértelműsített domainnevekhez kapcsolódva lehet biztosítani, csak más protokoll mentén lehet az adatokat küldeni és fogadni a hálózaton keresztül. Az emailcím felépítéséhez egy szinttel tovább kell bővíteni a domainnevek hierarchiáját, hiszen amikor az emailcímet létrehozunk, és a domainnév elé egy '@' jelet teszünk, majd az elé még egy karaktersorozatot illesztünk, akkor ezzel az adott domain alatt hozunk létre egy új nevet egy garantáltan egyedi nevekből álló lokális névtéren belül.

```
i@syi.hu = username: 'i' + szeparátor: '@' + domain: 'syi.hu'
```

Mivel az emailcímek anoním módon is használhatók, így ez a rendszer csak azt képes biztosítani, hogy az emailcímek egyediek, ezáltal egyértelműen hivatkozhatók legyenek, de más dolgok (például emberek) azonosítására nem alkalmasak. Ennek ellenére a web világában egyre több szolgáltatás alkalmazza a felhasználók azonosítására az általuk megadott és használt emailcímet.

2.6 Telefonszámok

A telefonszámok rendszerére is ráhúzható, hogy általában egy egyértelmű címzést lehetővé tevő címteret alkotnak, amelyben két tetszőleges pont közötti privát, címzett kommunikációt pont az teszi lehetővé, hogy a számok egyediek, bár az egyedi címek kiosztásának elvét és gyakorlatát megsértő ritka kivételként hivatkozni lehet az államszocializmus idején létezett ikertelefonok jelenségére, amikor ugyanazt a telefonszámot két végpontra is bekötötték. A telefónia történelmében azonban ez csak egy ritka kivételnek tekinthető, hiszen az elvárt szolgáltatásminőség nyilván csak az egyedi címek kiosztásával volt biztosítható. A telefonszámok rendszere – általában – háromszintű, hierarchikus címzési rendszer, amelyben az országok, majd az azon belül definált régiók, körzetek, végül a végpontokba bekötött egyedi telefonszámok különíthetők el.

36 30 333 1234 = 36 (országkód) + 30 (körzetkód) + 333 1234 (telefonszám)

A címteret, a telefonszámok címterében „megbúvó” hierarchiát tovább lehet bővíteni azzal, hogy telefonmellék felvételét engedjük meg az adott telefonszám alatt, aminek elérését valamilyen továbbkapcsolási technikával biztosítunk.

A telefonszámok egyértelműen nem kapcsolatok össze a telefont használó emberrel, de a lehetőség adott (különösen a mobiltelefonok esetében), és egyre több olyan kezdeményezést, kísérletet láthatunk a világban, amelyeknek célja az, hogy a mobilszámokat személyek, illetve az általuk kezdeményezett hálózati tranzakciók azonosítására használják.

2.7 Földrajzi helyek

A földrajzi helyeket a legelvonatbabb szinten három típusra bonthatjuk: beszélhetünk pontszerű, vonalszerű és poligonszerű objektumokról. Ez a felosztás csak gyakorlati szempontból tartalmaz. Elméletileg minden lokáció poligonszerű, és csak a felbontás adott pontosságával mellett beszélhetünk pontszerű vagy vonalszerű objektumokról, a felbontás mélységét növelve mindig kiderül, hogy ezeknek is kétdimenziós kiterjedése van. A földrajzi térben való tájékozódás megkönnyítése végett az ember ősidőktől fogva nevet adott a környezetének. Mindhárom geotípus esetén találhatunk neveket, amelyeket névterekbe lehet rendezni.

2.7.1 Postai címek

A földrajzi, postai címek, valamint az ingatlanvilvántartás rendszere is értelmezhető címtérként, hiszen ezek kialakítását, fenntartását a földrajzi helyek, a lakcímek, az ingatlanok egyértelmű azonosításának igényével magyarázhatjuk. Ezekre az azonosítási, egyértelműsítési feladatokra azonban elég erős kizárólagossági igény fogalmazódik meg az államigazgatási szféra irányából, hogy ezen a terepen nem várható valódi együttműködési szándék, így nem érdemes elmélyülni ezen a területen.

2.7.2 Földrajzi térképek, földrajzi nevek

A földrajzi navigáció kiegészítő eszköze régóta a térkép, amely egyfelől geometriai adatokat, másfelől névadatokat tartalmaz. A predigitális világban a térképekhez olykor kapcsoltak földrajzinév-tárákat (gazetteereket), amelyek a földrajzi helyek kétdimenziós vizuális reprezentációját kiegészítő alfanumerikus adatokat, neveket, számokat, statisztikákat tettek elérhetővé. A térképek, földrajzinév-tárák névszerű adatkomponensei esetében nyugodtan beszélhetünk névtérről, még ha ezek kortárs használói nem is így gondoltak rájuk. A földrajzinév-tárák, a térképi segédletek a térképeken, a földrajzi térben való eligazodást a helyek és nevek egyértelmű kezelésével, azonosításával tudták igazán támogatni. Ezek a földrajzi névterek műfajilag lehetnek többfélék. Van köztük „sima” adatbázis, van tezaurusz, van térképszolgáltató.

2.7.2.1 GeoNames

A GeoNames földrajzi adatbázisban több, mint 11 millió név szerepel a világ összes országából. Az adatok letölthetők, szabadon használhatók (geonames.org). Az adatokhoz egy egyszerű keresődoboz segítségével elindított kereséssel lehet hozzájutni.



The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge.

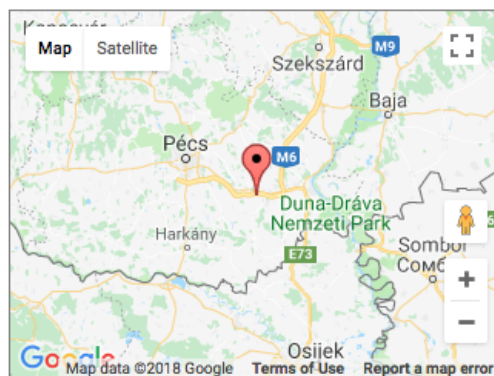
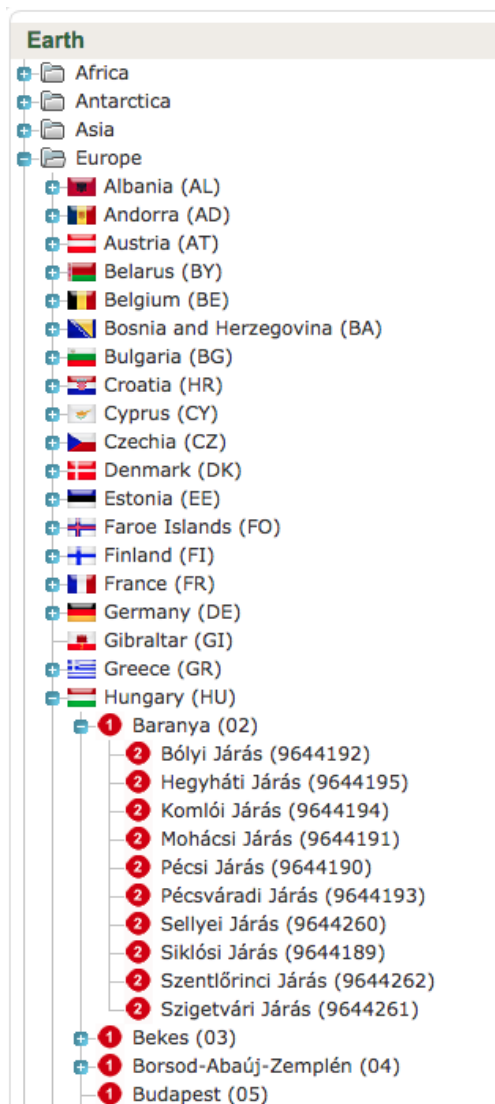
all countries

enter a location name, ex: "Paris", "Mount Everest", "New York"

| | | |
|--|---|--|
| Browse the names <ul style="list-style-type: none">CountriesPostal codesWikipediaCountry statisticsRecent modifications | Information <ul style="list-style-type: none">About GeoNamesData SourcesUser manualAmbassadors and TeamForumBlogMailing listCommercial Support and Consulting | Download <ul style="list-style-type: none">InfoFree Gazetteer DataFree Postal Code DataPremium Data Web Services <ul style="list-style-type: none">OverviewDocumentationClient LibrariesPremium Web Services |
| Sponsoring <ul style="list-style-type: none">Dealspotr Promo CodesChhabra Law - IP Attorneys | <ul style="list-style-type: none">PromoCodeWatch, LLC | <ul style="list-style-type: none">Activeinstafollowers.comDonations and Sponsoring |

a GeoNames nyitóoldala

A geoNames a találatokat sokféle formában képes megmutatni: ki tudja tenni térképes felületre, mutatni képes a kiválasztott földrajzi hely kapcsolatrendszerét, de hierarchikus nézetbe is be tudja rendezni a földrajzi helyeket azok tartalmazási kapcsolatai alapján. Hasznos szolgáltatás ezen az oldalon az, hogy az egérrel a hierarchikus lista elemeire mutatva a rendszer megjeleníti a lista mellett elhelyezett térképen a kiválasztott lokáció pozícióját a hely tágabb környezetén belül. Érdemes még azt is megjegyezni, hogy a megjelenítést mennyire képes feldobni az a megoldás, hogy az országokhoz, régiókhoz, adott esetben az alacsonyabb szintű helyekhez a nevek mellé elhelyezik a helyhez tartozó színes zászlókat, címereket.



GeoNames GeoTree hierarchikus névmegjelenítője térképes gyorsmutatóval

2.7.2.2 Getty TGN

A Getty földrajzi nevek tezaurusza (Getty Thesaurus of Geographic Names – Getty TGN) szintén a világ teljes földrajzi névkészletét igyekszik összegyűjteni. Ebben a rendszerben az elemek közti struktúrát nem a földrajzi-geometriai adatok alapján felépíthető partitív kapcsolatok adják (amit térképen lehet a legjobban reprezentálni). Ahogy a neve is mutatja, a Getty TGN tezaurusz, vagyis az egységek között többféle relációt megengedő, szabványos tudásszervezési rendszer, amiben túlsúlyban vannak a földrajzi tulajdonnevek.

Explore the Getty | Getty360 | Calendar | Blog | Connect with Us | Shop | Support Us Search

The Getty Research Institute

Exhibitions & Events | Special Collections | Library | Search Tools & Databases | Scholars & Projects | Publications | About the GRI

Print | Share

Search Tools & Databases

- Primo Search
- Getty Research Portal
- Collection Inventories & Finding Aids
- Photo Archive
- Research Guides & Bibliographies
- Digital Collections
- Article & Research Databases
- Collecting & Provenance Research
- BHA & RILA
- Getty Vocabularies

Getty Thesaurus of Geographic Names® Online

Search the TGN [? Help](#)

Find Name or ID: [Search](#)

Place Type: [Clear](#)

[Lookup](#)

Nation:

[Lookup](#)


[Pop-up Search](#) [Browse the TGN hierarchies](#)

[Have a Question?](#)

[Contact the Vocabulary Program](#)

a Getty TGN kezdőlapja

A szabványos tezaurusrelációkat nem jelentik meg az egyébként szokásos módon, bár nyilván ki lehet találni, mikor milyen reláció mentén kapcsolják össze az adataikat. A hangsúlyt a partitív relációra teszik, a keresési feltételek alapján kiadott találati oldalon a legszélesebb terjedelmű 'világ' (Föld) fogalmától kiindulva felsorolják a partitív hierarchiába tagolható egységeket, a legutolsó elemként a keresett egységet megjelénítve.



Research

[Research Home](#) | [Tools](#) | [Thesaurus of Geographic Names](#) | [Hierarchy Display](#)


Getty Thesaurus of Geographic Names® Online






Hierarchy Display

[New Search](#) | [Previous Page](#) | [? Help](#)

[Vernacular Display](#) | [English Display](#)

[View Selected Records](#) | [Clear All](#)

Click the  icon to view the hierarchy.
Check the boxes to view multiple records at once.


-  [Top of the TGN hierarchy](#) (hierarchy root)
-  [.... World](#) (facet)
-  [..... Europe](#) (continent)
-  [..... Hungary](#) (nation)
-  [..... Pécs](#) (urban county)

[New Search](#)

[Back to top](#)

a Getty TGN hierachikus névmegjelenítője 'Pécs'-re keresve


Az adott lexikai egység teljes kapcsolatrendszerét mutató oldalon látszik, hogy felveszik a földrajzi hely geometriai adatait, neveit, típusbesorolásait, a partitív herarchián belüli helyét (ebből többet is megadhatnak), az adatok forrását, olykor megjegyzéseket, további rokonsági kapcsolatokat.

Click the  icon to view the hierarchy.

[Semantic View](#) ([JSON](#), [JSONLD](#), [RDF](#), [N3/Turtle](#), [N-Triples](#))

ID: 7594106

Record Type: administrative

 **Pécs (urban county)**

Coordinates:

Lat: 46 06 00 N *degrees minutes* Lat: 46.1000 *decimal degrees*
Long: 018 15 00 E *degrees minutes* Long: 18.2500 *decimal degrees*

Names:

Pécs (preferred, C, V)
Pecs (C, V)
Pécs Megyei Város (C, V)

Hierarchical Position:

 World (facet)
 Europe (continent) (P)
 Hungary (nation) (P)
 Pécs (urban county) (P)

Place Types:

urban county (**preferred, C**)
first level subdivision (**C**)

Sources and Contributors:

Pécs..... [VP Preferred]
..... NGA, GEOnet Names Server (2008-) accessed 17 July 2014
Pécs Megyei Város..... [VP]
..... NGA, GEOnet Names Server (2008-) accessed 17 July 2014
Pecs..... [VP]
..... CIA, World Fact Book (2006-) accessed 17 July 2014

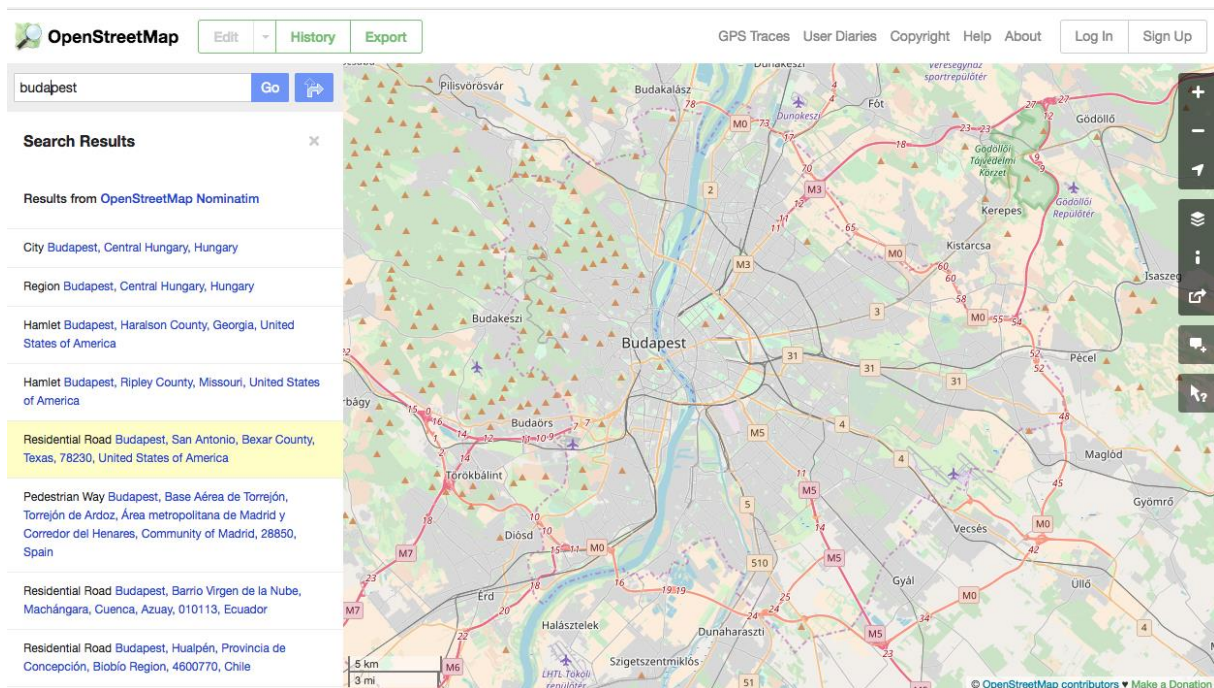
Subject: [VP]
..... NGA, GEOnet Names Server (2008-) accessed 17 July 2014

Pécs adatai a Getty TGN megjelenítő felületén

Mind a GeoNames, mind a Getty TGN alapvetően nyelvi alapon, a karakteres adatokra támaszkodva (a gazeteer mintájára) stuktúrálja a földrajzi helyekre vonatkozó adatokat. A hagyományos világ hagyományos technológiái csak ezt a rendezést tudták biztosítani. A digitális világ azonban felkínálja a rendelkezésre álló adatok geometriai alapú összerendelését és ezzel megteremti annak lehetőségét, hogy a földrajzi nevek gyűjteményét térképes alapon szervezzék egybe, amikor már a karakteres adatok lesznek a kiegészítő jellegűek. Két példát érdemes itt említeni, a Google Maps (maps.google.com) és az OpenStreetMap (osm.org) szolgáltatásait.

2.7.2.3 Google Maps

A Google Maps – az anyavállalat brutális pénzügyi és intellektuális erejére támaszkodva – rövid idő alatt piacvezető szolgáltatássá nőtte ki magát a saját műfajában. A szolgáltatás célja minél többfajta, minél alaposabb, minél megbízhatóbb navigációs segítséget adni a földrajzi térben eligazodni vágyók számára. Amennyire ebből a szempontból fontos lehet a földrajzi helyek, nevek egyértelműsítése, annyira fontos a vállalatnak a földrajzi névtér építése, de ezt – vélhetőleg – nem tekintik önmagában vett célnak, és a vállalat nem törekszik arra sem, hogy más szolgáltatókkal összerendelje saját adatállományát. Ennek ellenére hatalmas névállomány áll rendelkezésükre, amelyek a térképhez is vannak kötve. Ez nyilván kiváló alapja lehetne egy nemzetközi földrajzi névtérnek is, de hogy lenne-e bármilyen készlet egy kollaboratív alapon álló, nemzetközi földrajzi névtér felépítésére vonatkozóan a vállalat döntéshozóiban, arról semmi információ nem áll rendelkezésére, feltételezni pedig inkább azt érdemes, hogy nem lenne túl nagy céges hajlandóság erre.



az OpenStreetMap keresőfelülete

Az OSM mindegyik térinformatikai entitástípuson (pontszerű, vonalszerű vagy poligonszerű⁶) belül egyedi azonosítót rendel a földrajzi helyekhez, ami miatt lenne esély arra, hogy ezek alapján földrajzi névtérként lehessen használni, értelmezni a rendszert, de ennek a lehetőségnek útjában áll az a tény, hogy minden OSM-entitást bármikor ki lehet törölni, és persze bármikor újra lehet definiálni, vissza lehet venni a rendszerbe, de egy ilyen működésében semmi sem garantálja, hogy ugyanazok az egyedi azonosítók rendelődjenek ugyanazokhoz az entitásokhoz. Az OSM nem tudja rendszerszerűen biztosítani az azonosítók perzisztenciáját, ezért önmagában nem alkalmas arra, hogy földrajzi névtérként működjön.

2.8 Webes névazonosítási szabványok

A hálózati kommunikáció az adatok szintjén, a gépek közötti közvetlen kommunikációban mindig is címzett, pont-pont jellegű kapcsolatra támaszkodott. Az emberek számára látható, befogadható web világában a hálózaton keresztül elérhető erőforrások egyértelmű azonosítására nem lehetett időben állandó, perzisztens címzési rendszert fenntartani, mert a gyakorlatban mind a kommunikációt kiszolgáló domainnevek, mind a weboldalakra kitett dokumentumok nevei bármikor megváltozhattak. Természetesen a dokumentumok, erőforrások adott időpillanatban való egyértelmű címzését, elérhetőségét meg lehet oldani az erre szolgáló URI-technikák segítségével.

A Uniform Resource Identifier (URI) segítségével a webes erőforrásokhoz – meghatározott szabályok alapján – egyedi címet (karakter sorozatot) lehet rendelni, amely egységisége révén azonosítani tudja a címmel jelölt erőforrást, dokumentumot.

```
scheme://[user[:password]@][host[:port]]/[path][?query][#fragment]
```

Az URI maga csak egy fedőfogalom, az általános séma rögzítésén túl nincs konkrét szabályrendszer hozzákapcsolva, a tényleges azonosítást csak az URI két típusához, az URN-hez és az URL-hez kidolgozott technikák segítségével lehet megvalósítani.

A Uniform Resource Locator (URL) segítségével a domainnevek rendszerére támaszkodó webcímek a megjelölt protokollon (pl. http- vagy ftp-protokollon) keresztül azonosítják egyértelműen és teszik

⁶ Az OSM-ben a poligonokat relációknak nevezik.

elérhetővé az erőforrásokat. Ez az azonosító technika az erőforrások elérhetőségén, pontosabban az erőforrásokról szóló dokumentumok címén keresztül biztosít hozzáférést az erőforráshoz. A dokumentumok címe itt a digitális tér adott pontjára utal. Az URL egy példája lehet a következő cím.

<https://en.wikipedia.org/wiki/URL>

A másik URI-technika, az egységes forrásmegnevezés, Uniform Resource Name (URN) segítségével nem a digitális tér valamely pontjára, tehát az erőforrás helyére tudunk rámutatni, hanem az erőforrás nevét tudjuk megadni. Ehhez definálni kell egy névtérrel, és meg kell adni azt az azonosítót, amellyel egyértelműen rá lehet mutatni az erőforrásra a névtéren belül. Az URN-be bármilyen azonosítórendszer beilleszthető. Az RFC 2141 internet-szabvány értelmében minden URN azonosító három részből tevődik össze: az „urn” karaktersorból, a helynév-azonosítóból (Namespace Identifier, NID) és a helynév egyedi megjelöléséből (Namespace Specific String, NSS). Az NSS tartalmazza a felhasznált azonosítót, pl. az ISMN-t, a NID pedig a felhasznált azonosító típusának szabványos elnevezését. A könyvek világában használt ISBN-számokra az alábbi példát lehet adni.

<urn:isbn:963-05-3268-9>

Ez az URN rámutat az ISBN névtérre, és azon belül a Kenyeres Ágnes főszerkesztő vezetésével összeállított 'Magyar életrajzi lexikon' című könyv 1984-es, 4. kiadását azonosítja.

Az URN segítségével vissza lehet keresni magát a dokumentumot, annak leírását, vagy azon URL-ek listáját, amelyeken a dokumentum megtalálható. A felhasználó szemszögéből az URN használata egyszerű: a böngésző keresőablakába az URL helyett az URN-t kell begépelni. Az URN is a HTTP-protokollra és a DNS-szolgáltatásra épül.

2.9 Személynevek

A technikai példák után nézzük meg, hogy a személyneveket miért és hogyan használjuk, és hogyan lehetne személynevek esetében névtérrel építeni. Egy családon belül az eltérő keresztnév használatának nyilván ugyanaz az értelme, mint a fenti két példában: az egyedi nevek használatával az egyértelmű hivatkozás lehetőségét teremtjük meg a nevekkel jelölt dolgokra, a gyerekekre a családon belül. Egy család esetében ehhez elég abból a párezres keresztnév-állományból neveket választani, amelyek az idők során elfogadottakká váltak, de a családi közösségeken túllépve már nyilvánvalóan alkalmatlan lehet egy ilyen keresztnévtér minden gyerek azonosítására. Ebből a szempontból kicsit segíthet az összetett nevek képzése, amikor a személy teljes neve a családnév és a keresztnév kettőséből áll, de tudjuk jól, hogy ezzel a megoldással sem biztosítható az, hogy minden embernek egyedi neve legyen. A személynevek segítségével csak akkor tudjuk egyértelműen azonosítani az embereket, ha az emberek körét meghatározzuk vagy másként: pontos határvonalú megnevezési kontextust teremtünk. Mit is jelent ez? A személyek esetében szűk kontextusokban (családi, esetleg nagycsaládi, házközösségi, nemzetségi, falusi szinten) még remélhető, hogy a család- és keresztnév segítségével minden ember azonosítható legyen, de minél tágabb környezetben reméljük az egyértelmű azonosítást, annál biztosabban nem fog sikerülni, mert a homonimitás jelensége egyre gyakoribb lesz.⁷

A 'Kovács János' név már nem képes egyértelműen azonosítani egy embert, ha a faluban van több 'Kovács' nevű család, ahol kiosztják a 'János' keresztnévet valamelyik fiúknak. A példa rávilágít a megnevezési kontextus fontosságára. A családon mint névadási kontextuson belül még biztosítható a névtér működése, mindenkinek lehet egyértelmű azonosításra alkalmas nevet adni. Ezt a kontextust tágítva egy darabig még fent lehet tartani a homonimitásmentes állapotot, de a bővülés elérhet egy határt, amin túl már nem lehet biztosítani azt, hogy ugyanaz a név ne mutasson több személyre. Ekkor a névtér használhatatlanná válik, pontosabban részlegesen használhatatlanná válik, hiszen maradhatnak olyan szegmensei, amelyekben még képes jól funkcionálni, de igazából csak úgy lehet visszaállítani a névtér funkcionalitását, ha a szükséges méretűre szűkítjük a névadási kontextus terjedelmét.

⁷ A szinonimitás itt sem okoz gondot. Egy embernek lehet több neve is, ha azok egyedi, mindegyikükkel azonosítani lehet az adott személyt.

A személynévtér példája arra mutat rá, hogy a nevek önmagukban nem alkalmasak azonosításra, csak a kontextus és nevek együttes alkalmazásával remélhetjük a névterek sikeres működését. Megjegyezzük, hogy a személyek azonosítására az államigazgatás már régóta több megoldást is kidolgozott (személyi igazolvány, útlevel, személyi szám), ezekkel azonban itt nem foglalkozunk.

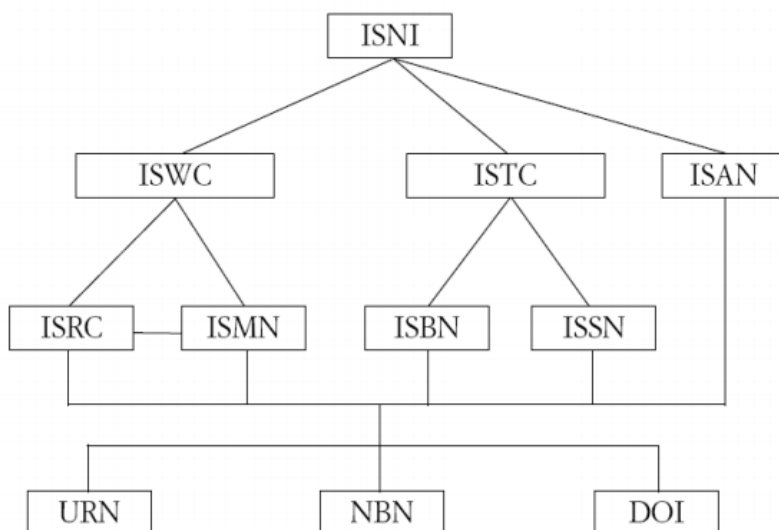
2.10 Műcímek

A közgyűjtemények, a dokumentumok világában régtől fogva komoly igény volt a nemzetközi azonosító rendszerek létrehozására a különféle dokumentumtípusok esetében. Az effajta igények létezését, erejét jelzi az a tény, hogy sok nemzetközi szabvány jött létre ezen a területen. A névtérépítkezés szempontjából nézve az alábbi szabványok lehetnek különösen fontosnak.

név- és címazonosítási nemzetközi szabványok

| | | |
|-----------|---|------|
| ISO 27729 | International Standard Name Identifier | ISNI |
| ISO 21047 | International Standard Text Code | ISTC |
| ISO 2108 | International Standard Book Number | ISBN |
| ISO 3297 | International Standard Serial Number | ISSN |
| ISO 15706 | International Standard Audiovisual Number | ISAN |
| ISO 15707 | International Standard Musical Work Code | ISWC |
| ISO 3901 | International Standard Recording Code | ISRC |
| ISO 10957 | International Standard Music Number | ISMN |
| ISO 26324 | Digital Object Identifier System | DOI |

Ezen szabványok a szöveges, az audiovizuális és az auditív dokumentumok kezelésével kapcsolatos elvárásokat rögzítik. A szabványok közti kapcsolatokat szemlélteti az alábbi ábra.



a szabványok rendszere

A szabványok fókuszja irányulhat a dokumentumok, művek létrehozójára, gondozójára, az emberre, illetve a műre mint kreatív alkotás tartalmára vagy a mű tartalma alapján készült – általában sokszorosított – dokumentumra mint kiadványra. Ennek megfelelően a fenti szabványok az alábbi három csoportba sorolhatók.

A szerzőt (létrehozót) azonosítja:

- ISNI: International Standard Name Identifier

A művet azonosítja:

- ISWC: International Standard Musical Work Code
- ISTC: International Standard Textual Work Code
- ISAN: International Standard Audiovisual Number

A kiadványt azonosítja:

- ISMN: International Standard Music Number
- ISBN: International Standard Book Number
- ISSN: International Standard Serial Number
- ISRC: International Standard Recording Code

Ezeket a szabványokat a nemzetközi szabványügyi világszervezet gondozza, és saját (ISO) számokat rendel hozzájuk. A digitális világban vannak olyan szabványok, amelyeket az ISO bocsátott ki, de ettől még nagy nemzetközi elfogadottsággal rendelkeznek. A digitális világ ilyen szabványai közül azok érdekesek számunkra, amelyek a digitális tartalmakat, forrásokat azonosítják.

- URN: Uniform Resource Name
- NBN: National Bibliography Number
- DOI: Digital Object Identifier

A hivatalos szabványosítási tevékenység egyre több területre terjed ki, és várható, hogy olyanok is megjelennek, elterjednek, amelyek érdekesek lehetnek a nemzeti névtérépítések számára is. Jelen helyzetben úgy tűnik, hogy a fenti szabványok lehetnek rövid távon is meghatározók a névterek számára, és majd az idő dönti el, hogy az alábbi listában látható, a hozzájuk hasonló szabványok közül melyikkel kell majd behatóbban foglalkozni.

azonosítással kapcsolatos, további nemzetközi szabványok

| | | |
|--------------------------|---|------|
| ISO 6166 | International Securities Identification Number | ISIN |
| ISO 7812 | Issuer identification number | IIN |
| ISO 13616 | International Bank Account Number | IBAN |
| ISO 15511 | International Standard Identifier for Libraries and Related Organizations | ISIL |
| ISO 17316 | International Standard Link Identifier | ISLI |
| ISO 17442 | Legal Entity Identifier | LEI |

2.10.1 Könyvek, folyóiratok

2.10.1.1 ISBN

Az ISBN (International Standard Book Number) a könyvek nemzetközi azonosítója. Előfordul, hogy zenei kiadványokat a könyvkereskedelmi hálózaton keresztül terjesztenek. Ha egy kiadó a könyvpiacra szánt kottát jelentet meg, az ISMN mellett ISBN azonosítóval is elláthatja. Olykor nehéz eldönteni, hogy egy kiadvány (pl. egy daloskönyv, egyházi énekeskönyv vagy egy album terjedelmes szöveges és képi anyaggal) vajon zenei kiadvány-e, "közönséges" könyv, vagy mindkettő. Ilyen esetben mind ISMN-nel, mind ISBN-nel el lehet látni a kiadványt. Ha a kiadó csak egyfajta azonosítót kíván alkalmazni, a választás az ő megítélésére van bízva, de kétség esetén az ISMN-t előnyben lehet részesíteni.

Az ISBN alkalmazásának szabályait az ISBN Útmutató tartalmazza. Az ISBN rendszer irányítását a Nemzetközi ISBN Ügynökség és a nemzeti ügynökségek végzik. Magyarországon a Magyar ISBN és ISMN Iroda. Amikor egy kiadványt ISMN-nel és ISBN-nel is ellátnak, mindkét azonosítót fel kell rajta tüntetni, és egyértelműen meg kell különböztetni őket egymástól.

Az ISBN vagy ISBN-szám (International Standard Book Number) 13 jegyű (a 2007. január 1. előtt kiadott könyveknél 10 jegyű) azonosítószám, a könyvek és egyéb monografikus jellegű művek nyilvántartására szolgáló nemzetközi szabványos számrendszerhez tartozó kód.

ISBN 0-306-40615-2



az ISBN vonalkódja

Az ISBN ötletét W. H. Smith vetette fel először 1965-ben, amikor könyvkereskedését egy számítógépekkel felszerelt raktárépületbe költöztette. 1966-ban készítették elő a Standard Book Numbering (SBN) elnevezésű rendszert, mely így 1967-ben bevezetésre kerülhetett az Egyesült Királyságban.

Az ISO tárgyalásokat kezdett 1968-ban Londonban Dánia, Franciaország, Németország, Írország, Hollandia, Norvégia, Nagy-Britannia, az Amerikai Egyesült Államok és az UNESCO megfigyelői részvételével a brit SBN rendszer nemzetközi bevezetéséről. 1969-ben a berlini és stockholmi ülések során született döntés az ISBN 1970-es nemzetközi bevezetéséről. Az ISBN lett az ISO 2108 szabvány. Ezt a rendszert használta 2007. január 1-jéig közel 150 ország, amikor az ISBN számjegyeinek számát 10-ről 13-ra emelték. Magyarország 1974 óta tagja az ISBN-rendszernek, országkódja 963. Az azonosítót az Országos Széchényi Könyvtárban működő Magyar ISBN Iroda osztja ki.

A következő kiadványok kaphatnak ISBN-t:

- Nyomtatott könyvek, brosrák;
- Braille kiadványok;
- Térképek;
- Oktatási célú video- és diafilmek, szoftverek;
- Kazettán, CD-n vagy DVD-n kiadott hangoskönyvek;
- Elektronikus kiadványok, akár fizikai hordozón, akár online hozzáféréssel;
- Multimédiás kiadványok, amennyiben a tartalom túlnyomó része szövegalapú.

Az ISBN első számcsoportja a GS1 által kinevezett 978 vagy 979 előtag; a második számcsoport a könyv származási országára vagy nyelvére utal, a harmadik a könyvkiadót jelöli, a negyedik a könyv azonosítására, az ötödik (ez csak egy számjegyből áll) az egész számsor matematikai ellenőrzésre szolgál. Az ISBN-ekhez automatikusan hozzárendelhető a hasonlóan szabványos, gépek által leolvasható vonalkód is. (A 979-es előtag 2008-as bevezetése érintette az ISMN rendszerét, amelyek a 979-0 számcsoportba kerültek.)

A könyvsorozatokot és az időszaki kiadványokat ún. ISSN (International Standard Serial Number) számmal azonosítják. Ez kétszer négy, egymástól kötőjellel elválasztott számcsoportból áll, amelyek az ISBN-nel szemben nem bírnak jelentéssel (nem utalnak országra vagy kiadóra).

Az ISBN – és sorozatok esetén az ISSN – feltüntetése a könyvben kötelező, ezek jellemzően a copyright- vagy a kolofonoldalra kerülnek. A mindig nagybetűkkel írandó ISBN/ISSN mozaikszavakat követő számcsoportokat kötőjelek vagy szóközök választják el egymástól.

Egy könyv mindegyik kiadása és változata (kivéve az utánnyomást) egyedi ISBN azonosítót kap. A szám lehet 13 vagy 10 (2007. január 1. előtti) számjegyű, és az alábbi részekből áll:

- ha 13 számjegyű ISBN, akkor egy GS1 prefix, 978 vagy 979,
- a származási ország vagy egy nyelvi területet jelölő kód,
- a kiadó,
- a termék száma, és
- egy ellenőrző karakter.

A különböző részek különböző hosszúságúak lehetnek, kötőjellel elválasztva. A jelenleg érvényes, 2006-ban megjelent ISBN útmutató szerint „az ISBN öt számcsoportból áll, melyek közül az elsőnek és

az utolsónak a hosszúsága kötött, a középső háromé változó. A számcsoportokat kötőjel választja el egymástól.”

A második számcsoport 0 vagy 1 az angol anyanyelvű országokban, 2 a francia, 3 a német, 4 a japán, 5 az orosz anyanyelvű országokban stb. Az eredeti SBN-ből hiányzott az országazonosító, de elérve egy 0-t a 9 számjegyű SBN helyes ISBN kóddá alakul át. Az ország mező 5 karakter hosszú lehet; a 99936-s Bhután számára van fenntartva.

A kiadó azonosítóját a helyi ISBN ügynökség állapítja meg. A termékszámot vagy maga a kiadó, vagy szintén az ügynökség adja ki. (Magyarországon jelenleg ez utóbbi a gyakoribb.) Nem kötelezhető minden kiadó hogy ISBN azonosítót rendeljen kiadványaihoz - mint például Kína -, mert az ISBN-re vonatkozó szabvány (MSZ ISO 2108:2005) alkalmazása önkéntes, de a legtöbb könyvesbolt csak ISBN azonosítóval ellátott kiadványt hajlandó forgalmazni.

A kiadók kiadói azonosítókat (más néven kérelmezőelemeket) kapnak. Minél rövidebb a kérelmezőelem, annál több karakter marad a kiadványok azonosítására. Így megeshet, hogy egy kisebb kiadó például 5 karakteres értelmezőelemet kap, de csak 1 karakteres termékazonosítót. Ha egy kérelmezőelem betelik, újat lehet igényelni. Ebből következően több különböző szám jelölheti ugyanazt a kiadót.

Az ellenőrző karakter

Számos irányelv létezik az ellenőrző karakter használatáról az ipar és a könyvtárak között. A kiadók olykor elfelejtik ellenőrizni az általuk kiadott kiadvány ISBN-jének érvényességét, ezzel nehéz helyzetbe hozva a könyvtárakat, könyvesboltokat és az olvasókat. A legtöbb könyvtár és könyvesbolt a kiadó által helytelenül megadott ISBN-t használja. A Library of Congress egy listát tart nyilván a hibás ISBN-nel megjelent könyvekről. Néhány könyvrendelő rendszer, mint például az Amazon.com, nem kereshető érvénytelen ISBN-nel.

Számítása a 10 számjegyű ISBN esetén

Hivatalos kézikönyvében a Nemzetközi ISBN Ügynökség a 10 számjegyű ISBN ellenőrző karakterét - ami a kód utolsó karaktere - a számjegyek 10-től 2-ig történő súlyozásával számolja modulo 11. Ez azt jelenti, hogy mind a kilenc számjegyet megszorozzák egy 10 és 2 közötti szorzóval, majd ezek összegét kiegészítik a legkisebb olyan számmal, ami maradék nélkül osztható 11-gyel. A kiegészítés lesz az ellenőrző karakter.

A fenti módszerre példa, ha az első kilenc karakter "0-306-40615":

$$\begin{aligned} & 10 \times 0 + 9 \times 3 + 8 \times 0 + 7 \times 6 + 6 \times 4 + 5 \times 0 + 4 \times 6 + 3 \times 1 + 2 \times 5 \\ & = 0 + 27 + 0 + 42 + 24 + 0 + 24 + 3 + 10 \\ & = 130 \end{aligned}$$

11 legközelebbi többszöröse: $12 \times 11 = 132$

$$132 - 130 = 2$$

Tehát az ellenőrző karakter a 2, és a teljes számsor az ISBN 0-306-40615-2.

A másik módszer szerint a számjegyeket megszorozzák a számjegy pozíciójával a számsorban (ez egy 1-9 közötti érték), majd ezek összegét kiegészítik a legkisebb olyan számmal, ami maradék nélkül osztható 11-gyel. A kiegészítés lesz az ellenőrző karakter.

A fenti módszerre példa, ha az első kilenc karakter "0-306-40615":

$$\begin{aligned} & 1 \times 0 + 2 \times 3 + 3 \times 0 + 4 \times 6 + 5 \times 4 + 6 \times 0 + 7 \times 6 + 8 \times 1 + 9 \times 5 \\ & = 0 + 6 + 0 + 24 + 20 + 0 + 42 + 8 + 45 \\ & = 145 \\ & = 13 \times 11 + 2 \end{aligned}$$

Tehát az ellenőrző karakter a 2, és a teljes számsor az ISBN 0-306-40615-2.

Látható, hogy a 10 számjegyű ISBN ellenőrző karakterét a számjegyek skaláris szorzataként számoljuk modulo 11. $(1,2,3,4,5,6,7,8,9) \times (x,x,x,x,x,x,x,x,x) \text{ mod } 11 = \text{ellenőrző karakter}$.

A két leggyakoribb hiba az ISBN használatakor (például gépeléskor vagy íráskor) a hasonló vagy egymás mellett álló számjegyek felcserélése. Mivel a 11 prímszám, az ISBN ellenőrző karakter számítási módszere minden esetben jelzi az említett két hibát. Ha a hiba a kiadónál keletkezik, a könyvet hibás ISBN azonosítóval is kiadhatják.

Számítása a 13 számjegyű ISBN esetén

A Nemzetközi ISBN Ügynökség hivatalos 2005-s kézikönyve tartalmazza a 2007. január 1-jétől használandó 13 számjegyű ISBN azonosítók ellenőrző karakterének számítási módszerét.

Az ellenőrző karakter számításához először mind a 12 számjegyet (13 karakteres ISBN, kivéve az utolsó ellenőrző karakter) eggyel vagy hárommal kell megszorozni. Ezeket összegezve modulo 10, majd a maradékot kivonva 10-ből áll elő az ellenőrző karakter. Ha 10 az eredmény, akkor 0 használandó. Például, a 978-0-306-40615-__ számítása:

$$9 \times 1 + 7 \times 3 + 8 \times 1 + 0 \times 3 + 3 \times 1 + 0 \times 3 + 6 \times 1 + 4 \times 3 + 0 \times 1 + 6 \times 3 + 1 \times 1 + 5 \times 3 \\ = 9 + 21 + 8 + 0 + 3 + 0 + 6 + 12 + 0 + 18 + 1 + 15 \\ = 93$$

$$93 / 10 = 9 \text{ maradék } 3$$

$$10 - 3 = 7$$

Tehát az ellenőrző karakter a 7, és a teljes számsor az ISBN 978-0-306-40615-7.

A 978-0-356-42615-__ számítása:

$$9 \times 1 + 7 \times 3 + 8 \times 1 + 0 \times 3 + 3 \times 1 + 5 \times 3 + 6 \times 1 + 4 \times 3 + 2 \times 1 + 6 \times 3 + 1 \times 1 + 5 \times 3 \\ = 9 + 21 + 8 + 0 + 3 + 15 + 6 + 12 + 2 + 18 + 1 + 15 \\ = 110$$

$$110 / 10 = 11 \text{ maradék } 0$$

$$10 - 0 = 10$$

10 cseréje 0-ra

Tehát az ellenőrző karakter a 0, és a teljes számsor az ISBN 978-0-356-42615-0.

Szemben a 10 számjegyű ISBN-nel, ez a rendszer nem minden esetben képes jelezni a helycserélési hibákat.

2.10.1.2 ISSN

Az ISMN és ISBN mellett egy kiegészítő azonosító számrendszer is működik a folytatódó kiadványok (korábban: időszaki kiadványok) számára: az International Standard Serial Number (ISSN). A folytatódó kiadvány egyik típusa az időszaki kiadvány. Ide tartozik – definíciója szerint – minden olyan kiadvány, tekintet nélkül hordozójára, amely egymást követő részegységekben jelenik meg, általában számozással vagy kronologikus megjelöléssel van ellátva, és előre meg nem határozott időtartamra tervezik a folytatását. Az időszaki kiadványokat meg kell különböztetni a többkötetes kiadványoktól, amelyek meghatározott számú kötet megjelenésével válnak teljessé.

Az időszaki kiadványok közé tartoznak a periodikumok és a sorozatok. Kotta csak elvétve jelenik meg periodikumként, de sorozatként elég gyakran. Mindkét kategóriában ISSN-nel kell azonosítani az időszaki kiadvány címét (amely a periodikum összes részegységén, illetve a sorozat minden kötetén változatlan marad), és sorozat esetében ISMN-t kell kiutalni a sorozatban megjelenő minden önálló műnek. (Ha a sorozat címe megváltozik, új ISSN megállapítására van szükség.)

A kiadók nem végezhetik maguk az ISSN azonosítók kiutalását. Magyarországon az időszaki kiadványokat megjelentető kiadóknak ISSN-ért a Magyar ISSN Nemzeti Központoz kell fordulniuk. Amikor egy kiadványt ISMN-nel és ISSN-nel is ellátnak, mindkettőt fel kell tüntetni rajta, és egyértelműen meg kell különböztetni őket egymástól.

Az ISSN (International Standard Serial Number) az időszaki kiadványok nemzetközi azonosítója. Az időszaki kiadványok (periodikumok) lehetnek folyóiratok vagy sorozatok. (Sorozatnak minősül az időszaki kiadvány akkor, ha mindig önálló, monografikus művek jelennek meg az egyes részegységekben, és e műveket különböző szerzők írják.) Alkalmazását az ISO 3297 nemzetközi szabvány, illetve annak honosított változata, az MSZ ISO 3297:2000 írja elő.

Az ISSN feladata

Minden ISSN egy adott időszaki kiadványt azonosít. Az ISSN azonosítószámhoz elválaszthatatlanul kapcsolódik az ISSN nemzetközi számítógépes nyilvántartási rendszerének előírásai szerint szabványosított címforma, azaz a kulcs cím. E szoros kapcsolat miatt, ha az időszaki kiadvány címében bekövetkezett változás folytán a kulcs cím megváltozik, a megváltozott címet egy másik ISSN-nel kell azonosítani.

Amennyiben az időszaki kiadvány címe, illetve kulcs címe változatlan marad, miközben megjelenési helye vagy kiadója, közreadója, megjelenési gyakorisága, netán külalakja vagy belső tartalma megváltozik, mindezen változások mellett ugyanaz az ISSN fogja azonosítani a kiadványt.

Saját ISSN-nel rendelkeznek az önálló, periodikusan megjelenő mellékletek, az ágazatok, az alsorozatok stb. Az időszaki kiadvány különféle kiadásváltozatait (másnyelvű kiadásváltozatokat, helyi vagy időbeli

kiadásváltozatokat, eltérő hordozón, pl. kazettán, mikrofilmlapon, mágnes- vagy kompaktlemezen, avagy a világhálón megjelenő kiadásváltozatokat) külön ISSN-nel kell azonosítani. Az egyszer már felhasznált ISSN nem alkalmazható újra egy másik címmel összefüggésben!

Az ISSN felépítése

Az ISSN nyolc számjegyből álló numerikus kód, melynek elemei - a könyveket azonosító ISBN-től eltérően - semmiféle jelentést nem hordoznak, csupán azonosító funkciót töltenek be. A nyolcadik számjegy ellenőrző számjegy, mely az előző hétből számítható ki, és 0-tól 10-ig terjedhet. A 10-es számjegyet római szám helyettesíti, azaz: X.

Az ISSN feltüntetése két, egyenként négy számjegyet tartalmazó adatszoportban történik, melyeket kötőjel kapcsol össze, és az ISSN betűkód előz meg.

ISSN 0251-1479

Az ISSN hálózat

Az ISSN nemzetközi hálózata jelenleg közel 90 nemzeti központból áll. Minden évben újabb nemzeti központok kezdik meg működésüket a világ különböző pontjain – folyamatosan karbantartott címjegyzékük elérhető a következő címen: <http://www.issn.org/the-centre-and-the-network/members-countries/the-iss...>

A nemzeti központok a saját országukban kiadott időszak kiadványok ISSN-nel történő azonosításáért és a nemzetközi rendszer előírásainak megfelelő nyilvántartásáért felelősek.

A hálózatot összefogó ISSN Nemzetközi Központ Párizsban működik. E központ végzi a nemzetközi szervezetek, valamint a nemzeti központtal nem rendelkező országok időszak kiadványainak ISSN-nel való ellátását és nyilvántartásba vételét.

A hálózat célja, hogy a tudomány valamennyi területén világszerte megjelenő időszak kiadványokról – tekintet nélkül azok megjelenési formájára, fizikai hordozójára – nemzetközi, számítógépes nyilvántartást vezessen. E szervezet tagjai felelősek az ISSN kiutalásáért, nyilvántartásáért és használatának ellenőrzéséért, valamint a világon megjelenő időszak kiadványokat magába foglaló Regiszter létrehozásáért és karbantartásáért.

Az ISSN regiszter

A párizsi ISSN Nemzetközi Központ a kezelése alatt álló nemzetközi adatbázis alapján közreadja az ISSN regisztert (ISSN Register), amely a világszerte kiutalt valamennyi ISSN azonosító szám hiteles nyilvántartását foglalja magába. Az ISSN regiszter tartalmazza az ISSN hálózat által feldolgozott, karbantartott és továbbított bejelentéseket, melyek tájékoztatnak az időszak kiadványok azonosításához és leírásához szükséges leglényegesebb bibliográfiai adatokról.

A közel 2 millió rekord 180 országban, 144 nyelven megjelent időszak kiadványról nyújt információt. A nemzetközi adatbázis évente 60-70 ezer új tétellel bővül, miközben közel 170 000 leírás frissítésére kerül sor, karbantartása folyamatos.

Az ISSN nemzetközi regiszter online elérését az ISSN Portál szolgáltatja. Hozzáférhető továbbá Z39.50 és OAI-PMH protokollal, valamint ISO 2709 csereformátumú nyers adatfájlként. A rekordok formátuma igény szerint lehet ISSN MARC 21, ISSN UNIMARC és MARC XML.

A regiszter 1996-ig mikrofilmlapokon jelent meg, 2004-ig mágnesszalagon, és 2006-ig CD-ROM kiadásban, ISSN Compact címmel. A világhálón ISSN Online néven volt ismert 1998 és 2005 között. 2005 márciusától kezdve az ISSN nemzetközi adatbázis az előfizetők számára felhasználói névvel és jelszóval, vagy IP cím azonosításával érhető el, az ISSN Portálon keresztül.

A nyílt hozzáférésű tudományos forrásokot leíró rekordok 2014 decembere óta leválogatásra kerülnek az ISSN regiszter előfizetéses adatbázisából a ROAD program keretében szabadon elérhető adatbázis számára: <http://road.issn.org/en/>.

Szintén szabadon letölthető az ISSN-L tábla, amely a különféle fizikai hordozókon megjelenő kiadásváltozatos ISSN azonosítóit teszi visszakereshetővé az ún. „kapcsoló ISSN” („linking ISSN”, azaz ISSN-L) segítségével.

Az ISSN jövője

Az ISSN világrendszer stratégiai célkitűzéseinek meghatározása arra a felismerésre épül, hogy - figyelembe véve a jelenlegi és a prognosztizálható trendeket - az ISSN jövője attól függ, mennyire terjed el használata állandó azonosítóként az Interneten. Az ISSN csak akkor őrizheti meg szerepét és hitelét,

ha a hálózat rövid időn belül jelentős mennyiségű elektronikus kiadvány azonosítását tudja felvállalni. E felismerésnek fontos következményei vannak az egész hálózatra nézve - egyrészt az erőforrások biztosításának vonatkozásában, másrészt az új feladatok pontos definiálása, a hálózati központok tevékenységének összehangolása terén.

Az ISSN alkalmazási körének szélesítése gyakorlatilag azt jelenti, hogy a hagyományos értelemben vett, egymást követő részekben megjelenő időszakos kiadványokon kívül a világhálón elterjedt integráló típusú bibliográfiai források is azonosíthatók ISSN-nel (meghatározott formai kritériumok megléte mellett). Ez utóbbiak között lehetővé válik az adatbázisok ISSN-nel történő ellátása is: célszerűnek látszik az előfizethető adatbázisok nyilvántartásba vétele, majd a bibliográfiai és cikkadatbázisok regisztrációja, valamint a kutatók számára fontos, egyes tudományterületekre szakosodott adatbázisok számbavétele, végül a rendszeresen frissített cím- és cégadatbázisok leírása is. Az elektronikus és azon belül a távoli hozzáférésű bibliográfiai források számbavételének folyamata már a 20. században, a 90-es évek közepén elindult, és a helyi lehetőségek, kapacitás és támogatottság függvényében halad azóta előre, egyre inkább gyorsuló ütemben. 2015-re mintegy 155 ezer folytatódó online forrás rekordját tartalmazza a nemzetközi ISSN adatbázis – köztük 1600 rekordot a Magyar ISSN Nemzeti Központ feldolgozásai nyomán. A MARC rekord 856-os mezőjében megadott távoli elérési címek révén az ISSN adatbázis felhasználói, illetve előfizetői világszerte tudomást szereznek a világhálón publikált időszakos kiadványokról.

2.10.1.3 ISTC

Az International Standard Text Code (ISTC) egyedi azonosító szöveg alapú művek számára. Az ISO 21047:2009 szabványt a TC 46/SC 9 testület dolgozta ki, 2009 márciusában publikálták. A szabvány alkalmazásáért felelős testület az International ISTC Agency. Az ISTC-kódokat, a hozzájuk tartozó metaadatokkal együtt egyetlen adatbázisban rögzítik, és ezt az adatbázist teszik kereshetővé, szűrhetővé.

Az ISTC-azonosítókat be lehet írni a művek bibliográfiai rekordjaiba, a webáruházak katalógusaiba, bárhová, ahol a mával kapcsolatos információt tárolnak, és ezek után a közös ISTC-kód alapján a szétszórt adatok összekapcsolhatók egymással. Az ISTC-azonosító segíthet az azonos című, de más tartalmú művek megkülönböztetésében, vagy épp olykor a más cím alatt megjelenő, ugyanazon művek összekapcsolásában. További előnyök származhatnak az ISTC-kódok alkalmazásából a származtatott művek összekapcsolásakor. Ha például a lefordított mű bibliográfiai rekordjába beleteszik az eredeti mű ISTC-kódját is, akkor ezek automatikusan összekapcsolódnak, és így további szolgáltatásokat lehet rájuk építeni (például keresések során mindkét irányból rá lehet mutatni a másik műre).

ISTC-kód 16 hexadecimális számból áll, amelyeket négy blokkra tagolva kell kitölteni. Az első blokkban kell megadni a regisztrációs ügynökség kódját, a másodikban a regisztráció évét, a harmadikban a mű egyedi azonosítóját, a negyedikben pedig egy ellenőrző számot kell elhelyezni, amely a ISO 7064 szabványban definiált MOD 16-3 algoritmus szerint végez vizsgálatot. A harmadik blokkban levő műazonosító az adott éven belül és adott ügynökséghez kötődően garantálja az egyediséget. E két adat adja az azonosító kontextusát.

ISTC 0A9-2002-12B4A105-6

2.10.1.4 DOI



A DOI-t (Digital Object Identifier = digitális objektumazonosító) az (elsősorban online megjelenésű) elektronikus kiadványok azonosítására hozták létre; segítségével az adott tartalom elektronikus úton terjeszthető. Hatékony eszköz a (bármilyen formában megjelenő) digitális tartalom hatékony kezeléséhez, összeköti a vásárlót a tartalomszolgáltatóval, megkönnyíti az elektronikus kereskedelmet és lehetővé teszi a szerzői jogok automatikus kezelését.

Mivel a DOI rendszer tagszervezetei nem földrajzi egységeket, hanem bizonyos dokumentumtípusokat és tevékenységeket fednek le, DOI igénylésekor célszerű a központhoz fordulni. A DOI kompatibilis az

ISMN-nel: a szerkezetébe akár az ISMN-t, akár más, a kiadók által használt azonosítót be lehet építeni. A DOI perzisztens azonosító.

<https://doi.org/10.1016/j.jtbi.2013.12.011>

2.10.1.5 NBN

Az URN (Universal Resource Name) egy olyan azonosító, amely az interneten található elektronikus dokumentumok hosszútávú azonosítására szolgál. Ennek egy altípusa az NBN (National Bibliography Number). Az URN azonosító típust azzal a céllal hozták létre, hogy kiküszöböljék az URL-ek használatából fakadó hátrányokat. Mivel az URL-ek gyakran változnak, a hivatkozások folyamatos karbantartása nagy erőfeszítést igényelne. A problémát még tovább növeli, hogy a változást közölnünk kell mindazokkal, akik az általunk közzétett forrásra hivatkoznak. Az URN azonosítók (természetesen az URL-el együtt) ezekre a nehézségekre próbálnak megoldást nyújtani.

<urn:nbn:de:bvb:19-146642>

2.10.2 Zenezámok

A zenezámok azonosítására több szabvány is létezik. Ennek az az oka, hogy a zenezámokat másként kezelik attól függően, hogy magát a művet vagy annak valamelyik felvételét akarják azonosítani, illetve más rendszert használnak a zenezámok auditív vagy vizuális megjelenési formáinak, a hanganyagoknak és a kottáknak az azonosítására.

2.10.2.1 ISMN

Az ISMN (International Standard Music Number = Nemzetközi Szabványos Kottaazonosító) numerikus kód, mely bármilyen formában megjelenő kották azonosítására szolgál. Alkalmazását az ISO 10957 nemzetközi szabvány írja elő.

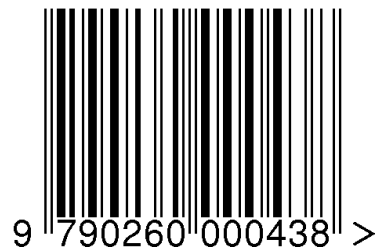
Miután az ISBN-ről bebizonyosodott, hogy széles körben alkalmazható, hasznos eszköz a könyvkereskedelemben, a zeneműkiadók is hasonló szabványos azonosító bevezetését javasolták a kották számára. A tárgyalásokat lassította az afölötti vita, hogy vajon elegendő-e egy egyszerű azonosító szám, vagy inkább bibliográfiai kódra van szükség, amely jelzi például a partitúra és a szólamok közötti kapcsolatot. Amikor a Zenei Könyvtárosok, Archívumok és Dokumentációs Központok Nemzetközi Szövetsége (International Association of Music Librarians, Archives and Documentation Centres = IAML) a Nemzetközi ISBN Ügynökséghez fordult egy tíz számjegyből álló, bibliográfiai kódok nélküli azonosító bevezetésének javaslatával, a tervezet megjelent az ISBN Review-ban, és lelkes fogadtatásra talált. Az ISO TC 46-os munkacsoportja befogadta a tervezetet, és komoly viták után, egy Ottawában rendezett találkozón az európai és észak-amerikai kiadók és szakértők megegyezésre jutottak. A fő érv a bibliográfiai kódra vonatkozó eredeti elképzelés elvetése mellett az ISBN-nel kapcsolatos tapasztalat volt, és az a lehetőség, hogy a 10 jegyű számot az ISBN mintájára be lehet építeni a 13 jegyű nemzetközi vonalkódrendszerbe.

Az ISMN szabványt rekordidő alatt dolgozták ki, és 1993 végén meg is jelent. Az ISO-szabványt a Magyar Szabványügyi Testület magyar szabványként 1997 júliusában adta közre. Az ISMN alkalmazása teljes körű ésszerűsítési lehetőségeket kínál a kottakiadás és-kereskedelem, valamint a zenei könyvtárak számára.

Az ISMN eredetileg 10 jegyű volt (1. a 10. fejezetet), 2008-ban viszont 13 jegyűvé alakult. Az eredetileg az ISBN-nek fenntartott 979 előtaggal kezdődik, melyet a 0 karakter követ; ez utóbbi került a 10 jegyű ISMN M betűje helyére. Ellentétben az ISBN-nel, az ISMN nem tartalmaz nyilvántartásicsoport-azonosítót (elterjedtebb nevén országazonosítót), a zene nemzetközi jellegéből adódóan.

Amikor egy 10 jegyű ISMN-t 13 jegyűvé alakítunk, az ellenőrző szám változatlan marad. (Ezzel szemben, ha egy 10 jegyű ISBN-t alakítunk át 13 jegyűvé, újra kell generálni az ellenőrző számot.)

ISMN 979-0-2600-0043-8



az ISMN vonalkódja

Az ISMN kottakiadványokat azonosít, amelyek kereskedelmi forgalomba kerülnek, bérbe vehetők vagy ingyenesen kerülnek terjesztésre, vagy csak szerzői jogi célokra hozzáférhetők. Az ISMN nem alkalmazható hang és videofelvételekre (kivéve az alábbiakban jelzett ritka eseteket), mivel ezeket más számok vagy kódok azonosítják. A hangfelvételeket a hangfelvételek nemzetközi szabványos kódjával, az ISRC-vel (International Standard Recording Code), a videofelvételeket az ISAN-nal (International Standard Audiovisual Number) kell ellátni.

Egy kiadvány minden önállóan hozzáférhető alkotórészét saját ISMN-nel kell ellátni. ISMN-nel azonosítandó kiadványok lehetnek:

- Partitúrák
- Zseb- vagy kispártitúrák (tanulmánypartitúrák)
- Vokális partitúrák
- Szólamgarnitúrák
- Önállóan kapható (hozzáférhető) egyéni szólamok
- Könnyűzenei fóliánsok
- Antológiák
- Más hordozók (kísérő anyagok), amelyek a zenei kiadvány szerves részét képezik (pl. olyan hangfelvétel, amely egy zenemű egy szólama)
- Nyomatott zeneművel megjelentetett dalszövegek vagy versek (amennyiben önállóan kaphatók/hozzáférhető)
- Nyomatott zeneművel megjelentetett magyarázó szövegek (amennyiben önállóan kaphatók/hozzáférhető)
- Daloskönyvek (nem kötelezően!)
- Zenei kiadványok mikroformában
- Zenei kiadványok Braille-írással
- Elektronikus formában terjesztett kották

A következő kiadványok nem kaphatnak ISMN számot:

- Zenei tárgyú könyvek, kivétel, ha kottázott zenei műveket tartalmaz
- Önálló hang- vagy videofelvételek (beleértve a számítógépes formában terjesztett felvételeket)
- Periodikumok és sorozatok egésze, ellentétben a sorozatok egyes köteteivel

Az ISMN felépítése

A zeneművek nemzetközi azonosító száma (ISMN) a 979-0 előtagból és az azt követő kilenc számjegyből áll. Nyomatásban vagy írásban az azonosítót az ISMN betűkód előzi meg. (Olyan

országokban, ahol nem latin betűket használnak, a helyi írásrendszerben alkalmazott rövidítés is szerepelhet az ISMN mellett.)

Az ISMN négy részből áll, ezek közül kettő változó hosszúságú. Nyomtatásban vagy írásban az ISMN egyes részeit kötőjel vagy szóköz választja el egymástól.

ISNM 979-0-2600-0043-8

A négy rész értelmezése a következő.

Előtag A 979-0 előtag különbözteti meg az ISMN-t az ISBN-től (a könyvek nemzetközi azonosítójától). A GS1 nemzetközi vonalkódrendszerében a 979 általánosságban az ISBN-t jelöli; ennek a számkapacitásnak az egytizedét, a 979-0-val kezdődő tartományt tartották fenn az ISMN számára.

Kiadói azonosító Ez az elem az adott zenei kiadvány kiadóját jelöli, 3-7 karakterből állhat. A nagyszámú kiadványt megjelentető kiadók rövid kiadói azonosítókat kapnak; a kevés zeneművet megjelentető k hosszabbakat.

Kiadványazonosító A harmadik elem a zenemű egy kiadását vagy annak önállóan terjeszthető, hozzáférhető vagy beszerezhető alkotórészét azonosítja; pl. a teljes partitúrát, a zseb- vagy kispartitúrát, a fűvós szólamok csoportját, az oboa szólamot stb.

A kiadványazonosítót rendszerint a kiadó állapítja meg egy meghatározott mű számára a rendelkezésére álló számtartományból (lásd még az 5. Nem közreműködő kiadók szakaszt). A kiadványazonosító hossza az előtte álló kiadói azonosító hosszától függ: ez a két elem együtt nyolc számjegyet tesz ki. Az ISMN-t már a kiadvány tervezése és előkészítése során ki lehet utalni, így belső azonosítóként is szolgálhat.

Ellenőrző számjegy Az ellenőrző számjegy 10-es modulussal számítható ki az 1-es és a 3-as váltakozó súlyszámokkal balról jobbra haladva. Ez azt jelenti, hogy az ISMN első 12 karakterének mindegyikét – azaz, magának az ellenőrző számjegynek a kivételével minden számjegyét – meg kell szorozni váltakozva 1-gyel és 3-mal balról jobbra haladva, és az így kapott szorzatok és az ellenőrző számjegy összegének maradék nélkül oszthatónak kell lennie 10-zel.

2.10.2.2 ISRC

Az ISMN nem vonatkozik a hang- és videofelvételekre (kivéve az olyan eseteket, amikor egy kottakiadvány részeként jelennek meg). Az International Standard Recording Code, ISRC (nemzetközi szabványos felvétel kód) azonosítja egy mű egy adott felvételét (de gyakran nem a fizikai egységet), tekintet nélkül arra, hogy milyen környezetben vagy hordozón jelenik meg. Az ISRC rendszer irányítását az IFPI: International Federation of Phonogram and Videogram Producers (Hangfelvétel és Videofelvétel Gyártók Nemzetközi Szövetsége) végzi. Az ISRC rendszernek nincs magyarországi tagszervezete. A CISAC magyarországi tagszervezete az Artisjus.

Az ISRC-kód négy részből áll:

- az első blokk az ország kódja,
- a második az országon belüli regisztrációs szervezet kódja,
- a harmadik a kibocsátás évét jelöli,
- a negyedik a regisztrátor által kiadott kód az adott zeneszámmra.

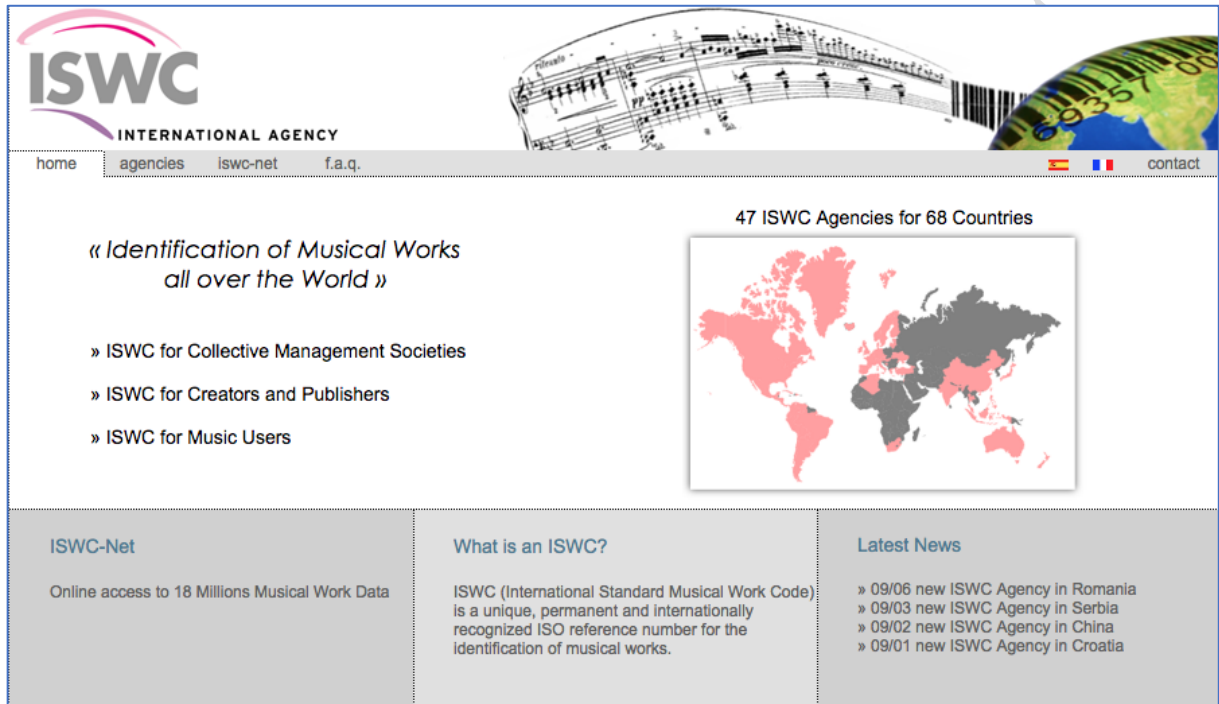
US-Z03-99-32476

2.10.2.3 ISWC

A könyvekhez hasonlóan a zeneművek is rendelkeznek nemzetközi azonosító számmal, ez az ISWC (International Standard Musical Work Code – nemzetközi szabványos zeneműkód). Az ISWC-kód

magát a zeneművet azonosítja, függetlenül annak megjelenési formájától. Alapvető fontosságú a szerzői jogok kezelésében. Minden ISWC egy "T" betűvel kezdődik, ezt követi egy 9 jegyű szám, majd egy ellenőrző számjegy. Az ISWC nemzetközi szinten azonosítja az adott művet. Minden zenemű, függetlenül attól, hogy megjelenik-e vagy sem, illetve hogy újonnan keletkezett vagy korábban hozták létre, kaphat ISWC-t. Ugyanez vonatkozik a zeneművek átdolgozásaira is.

A kód segítségével a külföldi jogvédő társaságok könnyedén beazonosíthatják a magyar műveket, az utánuk befizetett jogdíjakat pedig az Artisjuson keresztül elszámolhatják a magyar jogosultaknak. A zeneszámok Youtube-ra való feltöltésekor is érdemes megadni ezt a kódszámot a mű egyértelmű beazonosításához. Így az ebből a forrásból beérkező jogdíjak biztosan a jogosultakhoz fognak eljutni. Gyakran külföldi szerzői pályázatok, felhasználói szerződések kötelező eleme az ISWC kód.



ISWC
INTERNATIONAL AGENCY

home agencies iswc-net f.a.q. contact

« Identification of Musical Works
all over the World »

- » ISWC for Collective Management Societies
- » ISWC for Creators and Publishers
- » ISWC for Music Users

47 ISWC Agencies for 68 Countries

ISWC-Net
Online access to 18 Millions Musical Work Data

What is an ISWC?
ISWC (International Standard Musical Work Code) is a unique, permanent and internationally recognized ISO reference number for the identification of musical works.

Latest News

- » 09/06 new ISWC Agency in Romania
- » 09/03 new ISWC Agency in Serbia
- » 09/02 new ISWC Agency in China
- » 09/01 new ISWC Agency in Croatia

az ISWC nyitóoldala

Egy ISWC-kód felvételéhez minimum az alábbi információt meg kell adni.

- a zenemű címét,
- a zeneszerzők, a hangszerelők, alkotók nevét, szerepét a mű létrehozásában egy központi foglalkozási szereplista alapján), valamint az alkotók CAE/IPI-számát,
- a mű besorolási kódját (CIS-kód),
- azt a művet, amelyből a mű származik (ha átdolgozásról van szó).

Újra fontos itt jelezni, hogy az ISWC-kód a művet, nem a felvételt azonosítja. A felvételt az ISRC-számmal lehet azonosítani.

ISWC Regional Agencies

| Agency | Countries | Contact |
|------------|--|---------------------------------|
| APRA | Australia, New Zealand, Fiji | showland[aj]apra.com.au |
| CASH | People's Republic of China, Hong Kong, Taiwan | apple.tam[aj]cash.org.hk |
| ECAD | Brasil | gloria_braga[aj]ecad.org.br |
| Latinautor | Dominican Republic, El Salvador, Guatemala, Honduras, Nicaragua | mrevolorio[aj]aei-guatemala.org |
| MIS@Asia | India, Indonesia, Philippines, Singapore, Thailand | raymondtan[aj]compass.org.sg |
| Nord-ic | Denmark, Estonia, Finland, Greenland, Iceland, Latvia, Lithuania, Norway, Sweden | johan.hammarstrom[aj]stim.se |
| Song Code | Canada, United States | songcode[aj]ascap.com |

ISWC Agencies per Country


| Country | Agencies |
|----------------------------|------------|
| Algeria | ONDA |
| Argentina | SADAIC |
| Australia | APRA |
| Austria | AKM |
| Belgium | SABAM |
| Bolivia | SOBODAYCOM |
| Bosnia and Herzegovina | SNQ |
| Brasil | ECAD |
| Canada | Song Code |
| Chile | SCD |
| People's Republic of China | MCSC |
| Colombia | SAYCO |
| Costa Rica | ACAM |
| Croatia | HDS-ZAMP |
| Cuba | ACDAM |
| Czech Republic | OSA |
| Denmark | Nord-ic |
| Dominican Republic | Latinautor |
| Ecuador | SAYCE |
| El Salvador | Latinautor |
| Estonia | Nord-ic |
| Fiji | APRA |
| Finland | Nord-ic |
| France | SACEM |
| Georgia | SAS |
| Germany | GEMA |
| Greece | AEPI |
| Greenland | Nord-ic |
| Guatemala | Latinautor |
| Honduras | Latinautor |
| Hong Kong | CASH |

ISWC Local Agencies

| Agency | Countries | Contact |
|-------------|------------|-----------------------------|
| AACIMH | Honduras | |
| ACAM | Costa Rica | e.serrano[aj]acam.cr |
| ACDAM | Cuba | oliva[aj]acdama.cu |
| ACUM | Israel | odedf[aj]acum.org.il |
| AEI | Guatemala | |
| AEPI | Greece | |
| AGADU | Uruguay | |
| AKM | Austria | waltraud.schmidt[aj]aume.at |
| APA | Paraguay | apa[aj]quanta.com.py |
| APDAYC | Peru | |
| ARTISJUS | Hungary | avarkonyi[aj]artisjus.com |
| BUMA-STEMRA | Netherland | info[aj]bumastemra.nl |
| CASH | Hong Kong | apple.tam[aj]cash.org.hk |
| GEMA | Germany | jroehr[aj]gema.de |
| HDS-ZAMP | Croatia | Eugen Sugovic |
| IMRO | Ireland | declan.rudden[aj]imro.ie |

az ISWC regionális és országos ügynökségei

Az ISWC-szám kiosztását Magyarországon az Artisjus végzi. A kód kiosztásának feltétele a mű bejelentése az Artisjushoz.



| Működ | Műcím | Műbejelentés dátum | Műfaj | Időtartam | ISWC |
|------------|----------------|--------------------|-------|-----------|-------------|
| 400007875 | FASLYTALE | 2005.09.15 | 0-01 | 3:19 | T0070012345 |
| 400551314 | BELEVE | 2011.11.23 | 0-01 | 3:46 | T0070012346 |
| 4001980309 | THE VOICE | 2013.06.03 | 0-03 | 3:02 | T0070012342 |
| 4002049827 | ONLY TEARDROPS | 2013.11.11 | 0-01 | 2:30 | T0070012347 |
| 4002049987 | MY NUMBER ONE | 2013.11.11 | 0-01 | 3:57 | T0070012341 |

zenei művek bejelentése az Artisjushoz

A művek azonosítása mellett mód van a szerzők azonosítására is. Erre szolgál az IPI-kód (IPI – Interested Parties Information), amellyel a szerzőt, a jogtulajdonost lehet regisztrálni. A művek mellett a személyekre is van Artisjus-azonosító.

Szerzői ü
Tel.: (1) 4
ügyfészol

artisjus
Szerzői
Információs
Rendszer

Művek Forgalmak Értesítők

Artisjus azonosító: 1234567 IPI azonosító: 12345678

Személyi adatok Pénzügyi adatok

Az IPI szám egy nemzetközi azonosító, mely zenei jogosultak (szerzők, zeneműkiadók) pontos beazonosítását segíti. Az IPI azonosító kiosztását Magyarországon az Artisjus végzi, amit automatikusan eljuttatunk a külföldi jogvédő társaságokkal közösen használt nemzetközi adatbázisokba is.

a zenei művel kapcsolatba hozható személy IPI-azonosítója

2.10.3 Filmek, audiovizuális tartalmak

2.10.3.1 ISAN

Az International Standard Audiovisual Number (ISAN), mint a nevéből is látszik, az audiovizuális műveket azonosítja. Audiovizuális műnek számít minden olyan alkotás, amely mozgóképekből tevődik össze, hangkísérettel vagy anélkül. Az ISAN magát a művet azonosítja, nem a fizikai hordozón megjelenő kiadványt vagy az adást. Lehetővé teszi a szerzői jogok birtokosai (szerzők, előadók, gyártók) számára, hogy egyértelműen azonosítsák az audiovizuális művet.

Az ISAN Audiovizuális művek és verziók azonosítására szolgáló, önkéntes számozási rendszer, amelyet az ISO (Nemzetközi Szabványügyi Szervezet) egyik csoportja (a Műszaki Bizottság 46, 9. albizottság) dolgozott ki. Az ISAN projektet az AGICOA (az Audiovizuális művek Nemzetközi Kollektív Igazgatásának Szövetsége), a CISAC (a Szerzők és Zeneszerzők Szövetségeinek Nemzetközi Szövetsége) és a FIAPF (a Filmgyártók Szövetségeinek Nemzetközi Szövetsége) irányította, mindhárman számos audiovizuális gyártással foglalkozó nemzetközi és nemzeti szervezetet képviselve. A nemzetközi szabványt, amely az ISAN rendszer alapját képezi 2002-ben adták ki mint ISO 15706-1: 2002 Információ és Dokumentáció - Nemzetközi Szabvány Audiovizuális Szám (ISAN) - 1. rész: Az Audiovizuális mű azonosítója. 2007-ben a ISAN szabványt kiegészítették, hogy verzió azonosító támogatást nyújtson az ISO 15706-2: 2007 Információ és Dokumentáció - Nemzetközi Szabvány Audiovizuális Szám (ISAN) - 2. rész: Verzió Azonosító. Együtt, egy teljes Audiovizuális Tartalomazonosítási rendszert írnak le.

Az ISAN egy olyan önkéntes számozási rendszer, amely nem kapcsolódik semmilyen szerzői jogi nyilvántartáshoz. Az ISAN az ISO 15706 nemzetközi szabványnak felel meg.

Az ISAN változatlan marad a műmegosztás minden szakaszától függetlenül, beleértve minden új verziót.

Az ISAN olyan konkrét leíró információkat (metaadatokat) foglal magában, mint a cím, a rendező, a típus, a időtartam stb. Ez minden műfajra vonatkozik, beleértve a játékfilmeket, a sorozatokat, a dokumentumokat, az élő közvetítéseket, a szórakoztató programokat, a multimédiát stb., és azok sok változatát (szinkronizálás, hangalámondás, közvetítés, Blu-ray), beleértve a szorosan kapcsolódó tartalmakat (pl. előzetesek) vagy szorosan kapcsolódó elemeket (pl. filmzene, poszter).

Bár az ISAN gyökér szegmense ugyanaz marad az Audiovizuális mű minden változatánál, az ISAN verziószegmens lehetővé tesz vékonyabb szemcsés azonosítást verziószinten is, úgymint kiadásokat, nyelvi változatokat és médiafigyelést, mint például a hagyományos filmszalag, digitális mozi, a Blu-ray, DVD, mágnesszalagok, digitális fájlok, digitális streamek (adatfolyamok). Az ISAN kompatibilis a legfontosabb audiovizuális szabványokkal, például MPEG, Windows Media, DVB ATSC, MXF stb. Az ISAN beépíthető fizikai adathordozókba, digitális streamekbe (adatfolyamokba), digitális fájlokba, valamint vonalkódokba, kiadványokba, csomagolásba, reklámba és audiovizuális megállapodásokba. Az Audiovizuális Tartalmakhoz (művekhez vagy verzióhoz) hozzárendelt ISAN állandó, független a tulajdonjogi változásoktól.

Az ISAN felhasználási módjait úgy alakították ki, hogy azok megfeleljenek az audiovizuális iparág különböző igényeinek. Erőssége ebből az „egyetemességből” fakad.

Az ISAN-szabvány az erős nemzetközi kulturális sokféleség miatt az ISAN Regisztrációs Ügynökségek hálózatán alapul, amelyek a nemzetközi ISAN Rendszer és a helyi ipar közötti kapcsolódási pontként működnek. A regionális ISAN Regisztrációs Ügynökségeket az ISAN-IA nevezi ki, miután támogatást kaptak a helyi ipartól. A regisztrációs ügynökségek alkalmazott ISAN Szolgáltatásokat és szaktudást nyújtanak a helyi nyelven és az ISAN közösségek felé jelentik a helyi elvárásokat.

Az ISAN csak az ISAN Rendszeren keresztül kinevezett Regisztrációs Ügynökségeken keresztül adható ki.

Az ISAN hálózat a Nemzetközi Ügynökség (ISAN-IA), amelyet az AGICOA, a CISAC és a FIAPF alapított, támogatásával üzemel.

Az ISAN-t már több mint 730 000 Audiovizuális Tartalom-hoz rendelték hozzá.

Az ISAN Regisztráció online történik, függetlenül attól, hogy egy vagy több műre vonatkozik. A tömeges ISAN-allokációk egyszerre végrehajthatók XML-fájl alapú folyamatokon keresztül. Egy ötvennél több ISAN Webes Szolgáltatás csomagos készlet teszi lehetővé az automatikus regisztrációkat és a lekérdezéseket harmadik féltől származó alkalmazásokhoz.

Az ISAN az alkotásokat, azok teljes kereskedelmi ciklusán keresztül azonosítja, mindenféle változatban, különböző nyelveken, vágásokban, formátumokban, médiumokban; mindenkor, minden módon és a terjesztés minden szakaszában.

Az ISAN hasznára van a szerzőknek, a műsorszolgáltatóknak, a gyűjteménytársaságoknak, a vállalatoknak, a gyártóknak, a jogtulajdonosoknak és még sokan másoknak.

Az ISAN előnyei

Minden egyes ISAN célja, hogy egyedülálló nemzetközi azonosítóként szolgáljon minden Audiovizuális mű („AV/mű”), és annak bármely verziója számára. Az Audiovizuális művek és a hozzájuk kapcsolódó verziók ugyanazt a gyökérszegenst tartalmazzák. Az ISAN leíró információkkal (metaadatokkal) van összekapcsolva, amelyek mindegyike a Központi Adattár-ban található.

Az ISAN kódok egy egyedi Központi Adattár-ból származnak. Minden AV/mű vagy verzió egyetlen egyedi ISAN-al rendelkezik. Minden ISAN csak egy AV/művet vagy egy verziót azonosít. Az ISAN egy permanens és állandó azonosító.

Az ISAN minden típusú Audiovizuális műre és azok mindenféle verziójára vonatkozik, úgymint nyelvi változatok, kiadások, kivonatok, klippek, média változatok, digitális kódolás, valamint az ezekhez szorosan kapcsolódó tartalmak vagy tételek, úgymint filmzenék, feliratozás, dobozolt készletek stb.

Az ISAN úgy lett kialakítva a Nemzetközi Szabványügyi Szervezet (ISO) által, hogy megfeleljen a digitális forradalom, tartalomtulajdonosok, közvetítők, a médiavállalatok és a szabványügyi szervezetek által meghatározott követelményeinek.

Az ISAN és a hozzá kapcsolódó leíró információk mindenkor elérhetők online a Felhasználók (Regisztrálók és Olvasók) által. Az API-k és a tömeges folyamatok lehetővé teszik a meglévő folyamatok hatékony integrálását. Az ISAN segíti az ISAN regisztrálók hatékony készletgazdálkodását, különösen a digitális környezetben. Azáltal, hogy megköveteli, hogy az ISAN adásnaplóban szerepeltetését a használat nyomon követését könnyíti meg. Egy teljesen digitális környezetben az ISAN magába a műbe integrálódik.

Kollektív igazgatással történő adminisztráláskor az ISAN használatával könnyebbé válik a jogok nyilvántartása, ami jelentős idő- és költségmegtakarítást eredményez. Ez lehetővé teszi a gyűjteménytársaságok gyorsabb, megbízhatóbb és hatékonyabb jogelosztását.

Az ISAN megkönnyíti a különböző külső adatbázisok közötti elektronikus információcserét. Az ISAN elősegíti a mű használatának jobb nyomon követését a digitális világban, és ezáltal segít a kalózkodás ellen. Az ISAN kódok könnyen alkalmazhatóak, és felhasználhatóak az audiovizuális e-kereskedelmi környezetben.

Az ISAN az ISO tulajdona. Az ISO szerződéses kötelezettséget vállal az ISAN-IA és az ISAN Regisztrációs Ügynökségek ISAN Szabvány költség-visszatérítési alapon való működtetéséhez és az ISAN Szolgáltatásainak ésszerű díjért történő kiszámlázására.

Az ISAN Felhasználási Feltételek lehetővé teszik az ISAN kódok korlátozások nélküli felhasználását kereskedelmi és nem kereskedelmi célokra egyaránt.

Az ISAN azonosító általános szerkezete

Az ISAN-t úgy tervezték, hogy emberek olvassák és információs rendszerek dolgozzák fel, egy 24 bites hexadecimális számként (0-tól 9-ig és A-tól F-ig) vagy 96 bites bináris számként. Az ISAN szerkezetét úgy alakították ki, hogy kielégítse az audiovizuális ellátási lánc különböző igényeit, és a számozási sémát úgy tervezték meg, hogy milliárd különálló cím azonosítását tegye lehetővé, és így az audiovizuális termelési igényeket legalább egy évszázadra kiszolgálja.



ISAN azonosító

Az ISAN három szegmensre oszlik:

- az első 12 számjegy jelöli a gyökér szegmenst,
- a következő 4 számjegy jelöli az epizódot/ egy sorozat részt,
- az utolsó 8 számjegy a verziószegmenst jelöli.

Az ISAN gyökér- és epizódsgzmenst jelölő első 16 számjegy a művek bármely verziójától függetlenül azonos marad - csak a legutolsó 8 számjegy változik, ahogy a verziók készülnek.

Az ISAN 17. és 26. számjegyei az ellenőrző karakterek, amelyek az ISAN helytelen átírásából eredő hibák ellen védenek. Az ellenőrző karaktereket egy számítógépes *algoritmus* rendeli hozzá automatikusan.

Amikor az ISAN-t olyan műhöz rendelik hozzá, amely nem Epizód, vagy Audiovizuális sorozatmű része, akkor az epizódsgzmenst 4 nullát (0000) fog tartalmazni.

Amikor az ISAN-t egy Epizód-hoz vagy egy Audiovizuális sorozatmű-höz rendelik hozzá, az epizódsgzmenst egy sorozathoz tartozó bejegyzett epizód azonosítására szolgál. Ebben az esetben egy epizódsgzmenst soha nem tartalmaz 4 nullát (0000).

0 értéket (azaz 0000-0000) akkor rendelnek az ISAN Verziósgzmenst-éhez, ha nincs verzió.

Az ISAN Verziósgzmenst, ahol az első 4 bit 1111 (azaz az F hexadecimális érték) privát használatra van fenntartva, és soha nem hivatkoznak rá az ISAN Központi Adatbázisában. Az ISAN Verziósgzmenst-ének ez a tartománya kizárólag belső alkalmazásokhoz van fenntartva (pl. a nyilvános szerkesztéshez készült verzió elkészítése során a belső szerkesztések nyomon követésére), ahogy azt egy ISAN "Verzió Regisztráló" meghatározza, aki az ISAN Verziósgzmenstét önállóan hozzárendelheti nem-regisztrált alapon. Az ebből a belső tartományból származó ISAN nem terjeszthető külsőleg, mivel automatikusan más alkalmazások számára érvénytelennek minősülnek.

Nyomatott formátum

A nyomtatott formátumban megjelenített ISAN mindig tartalmazza a két ellenőrző karaktert. Az ISAN pontos átírásához a következő szabályok érvényesülnek:

- az „ISAN” szavat, amelyet szóköz követ, meg kell előznie minden számjegynek;
- az első 16 számjegyet 4 db 4 számjegy hosszúságú egységre kell felosztani, ahol minden egyes egységet szóköz vagy kötőjel választ el egymástól, és amelyeket az ellenőrző karakter követ;
- azután a 8 számjegyet 2 db 4 számjegy hosszúságú egységre kell felosztani, ahol minden egyes egységet szóköz vagy kötőjel választ el egymástól, valamint az ellenőrző karakter követ;

Ezek a szabályok kizárólag arra szolgálnak, hogy vizuális segítséget nyújtsanak az ISAN pontos átírásához. Semmiféle jelentést, egységet vagy karaktert sem rendelnek hozzá semmilyen számjegyhez..

Példa:

ISAN AEF0-1000-6721-0022-X-0000-9034-1

Bináris formátum

Amikor bináris formában jelenítik meg, akkor az ISAN-t 96-bites bináris számként balról jobbra kódolják. A 96 bitet kódolják és előjelmentes egész számként továbbítják. Az ellenőrző számjegyeket figyelmen kívül hagyják.

MSB

LSB

RRRRRRRRRRRRREEEEEVVVVV VVV

Msb = legjelentősebb bit

Lsb = legkevésbé jelentős bit

XML formátum

Amikor XML dokumentumokban jelenik meg, akkor az ISAN-t a <http://www.isan.org/ISAN/isan.xsd> linken található séma szerint kódolják. Az "ISAN" betűket elhagyják, és a két ellenőrző karakter opcionális az ISAN XML-ábrázolásában. Az <http://www.isan.org/ISAN/isan.xsd> által meghatározott XML séma korlátozásain túl az XML kódolás még a következőképpen kerül korlátozásra:

- ha a check1 attribútum jelen van, akkor az episodeOrPart attribútumnak jelen kell lennie;
- ha a verzió attribútum jelen van, akkor az epizódOrPart attribútumnak jelen kell lennie;
- ha a check2 attribútum jelen van, akkor a check1 attribútumnak jelen kell lennie;
- hacsak az XML-kódolást nem szigorúan gépek között használják, minden attribútumra szükség van.

Nézzünk meg néhány példát!

```
<ISAN root ="1881-66C7-3420" episodeOrPart ="6541" version ="9F3A-0245" />
<ISAN root ="1881-66C7-3420" episodeOrPart ="6541" Check1 ="X" version ="9F3A-0245"
check2 ="Y" />
<ISAN root ="1881-66C7-3420" episodeOrPart ="6541" />
<ISAN root ="1881-66C7-3420" />
```

Megjegyzés: A 4. példa nem teljesen definiálja az ISAN-t, de speciális körülmények között, ahol csak az ISAN gyökérkomponense ismeretes, alkalmazható.

URN formátum

Amikor URN formában jelenik meg, a reprezentáció hasonló az emberi, olvasható formátumhoz, és az "URN: ISAN:" előre definiált, mint a következő példában látható.

URN: ISAN: 0000-0000-D07A-0090-Q-0000-0000-X

Az ISAN alkalmazása audiovizuális alkotásoknál és verzióknál

Audiovizuális művek és változataik

A rendszer számára az ISAN egy Audiovizuális művet, úgy határoz meg mint olyan "művet, amely kapcsolódó képek sorozatából áll, kísérő hanggal együtt vagy anélkül, amelyet mozgó képként jelenítenek meg eszközök használatával, a kezdeti vagy későbbi rögzítés módjától függetlenül ". Ez semmiképpen sem tekinthető azonban audiovizuális alkotások jogi meghatározásának.

A művek legtöbb típusa sorozatként vagy nem sorozatként regisztrálható. A művek legtöbb típusa az ISAN bevezetésének korai szakaszában használható. A művek legtöbbjének lehetnek verziói (kifejezések és megnyilvánulások), amelyek egy különálló ISAN Verzió-val azonosíthatók

Az AV/művek típusai, amelyekhez az ISAN hozzárendelhető, a következők például:

| Kód | Leírás | Meghatározás | Lehet sorozatszám? | Lehet összetett? |
|----------|----------|--|--------------------|------------------|
| HIRDETÉS | Reklám / | Művek, amelyek terméket, szolgáltatást, ötletet, személyt vagy | | |

| | | | | |
|----|--|--|-------|------|
| | Reklámfilm | céget, beleértve a fogyasztók számára szóló promóciós anyagot, színházi szolgáltatásokat, infomercial-t digitális feliratozást reklámoznak. Igen Igen | Igen | Igen |
| | | előzetesek (promóciós reklám egy film vagy televíziós produkció számára) NEM szerepelnek ebben a kategóriában, és azokat, mint az előzetes által azonosított mű egy verziójaként kell azonosítani. | | |
| FF | Játékfilm | Mű, amelyet eredetileg színpadra vagy közvetlenül videóra készült | Nem | Igen |
| TF | TV film vagy Tévjáték | Televíziós műsorszórásra szánt mű. | Igen* | Igen |
| DO | Dokumentumfilm | Egy mozgókép, amely tényalapú archívumot hoz létre vagy emberekről, állatokról, zöldségekről és növényekről, eseményekről vagy ipari folyamatokról tudósít. | Igen | Igen |
| ED | Nevelési | Az ilyen műveket elsősorban oktatási mintsem szórakoztatási célokra szánták.. | Igen | Igen |
| SE | Drámai & Komédia sorozat Szériák, "szappanoperák", | Egy fikciós sorozata gyakran ugyanazokkal a főszereplőkkel működik vagy kapcsolódó témákról szóló nyílt végű televíziós dráma vagy vígjáték, ahol a történet epizódról epizódra folytatódik. | Igen | Igen |
| SH | Rövidfilm | Fikciós mű, amely rendszerint eredetileg színházi előadásra vagy közvetlenül videóra adtak ki amelynek eredeti hossza kisebb, mint a játékfilmé. Magába foglal rövid animált alkotásokat . | Igen* | Igen |
| MM | Multimédia | Audiovizuális mű interaktivitással. | Igen | Igen |
| NE | Hírek | Olyan audiovizuális produkció, amely közérdeklődésre számot tartó témákról számol be. Nem-fikciós programok,riportok és kommentár aktuális eseményekről, amelyeket rendszerint különféle tudósítóktól származó riportokba szerkesztenek | Igen | Igen |

| | | | | |
|----|---|--|------|-------------------------|
| LV | Élő esemény | Minden nem-fikciós mű, amely élő eseményeket rögzít, kivéve olyan sporthoz vagy előadásokhoz kapcsolódóakat, amelyeknek saját típusuk van. Például politikai események széles körű tudósítása, élő beszélgető műsorok, politikai viták, választások, úrbéli közvetítések házasság, temetés stb. élő rögzítése. | Igen | Igen |
| PF | Performansz | Művek amelyek előadásokat dokumentálnak, ideértve koncerteket, tánc, zene, opera, színpadi produkció, recitálás, bűvész-show, cirkuszi, stand-up komédia, bórleszk, vagy más színpadi előadás. | Igen | Igen |
| SP | Sportesemény, Sporthoz kapcsolódó program | Nem-fikciós tudósítás sporteseményekről. | Igen | Igen |
| TE | Egyéb televízió szórakozás | Televíziós sugárzásra szánt programok, beleértve reality-show-kat, magazinokat, vetélkedőket, kivéve máshova kategorizáltakat. | Igen | Igen |
| VC | Zenei videoklip | Rövid audiovizuális nem élő produkció, amely egy adott hangfelvételhez kapcsolódó képeket mutat. | Igen | Igen |
| CO | Összeállítás | 1 - Egyetlen kiemelt mű több kapcsolódó művel egy csomagban. (pl. a hagyományos DVD bónusz funkciókkal) - mindegyik mű a gyűjteményben saját felvételel rendelkezik. 2 - Egy dobozos készlet, amely két vagy több kiemelt művet tartalmaz egy csomagban. | Nem | Igen (mindig összetett) |

Egy ISAN Verzió-t bármely audiovizuális mű adott példányához (vagy verziójához) lehet alkalmazni, amely elemek bizonyos összesítéséből eredve egy ISAN-al már azonosított Audiovizuális mű tartalmát befolyásolja. Változások, mint például kiadások, a feliratok hozzáadása, a szinkronizálás, a hangsáv módosítása, a tartalom rögzítése egy adott médiumon (filmnyomatás, Blu-ray, digitális fájl vagy adatfolyam) a műszaki tulajdonságok (HD / SD, 2D / 3D, kódolási formátum) megváltoztatása mind új ISAN Verzió kiadásához vezethet.

Néhány példa művek módosítására, amelyek audiovizuális alkotások új verzióit eredményezve ISAN Verzió kiadásához vezethetnek:

- Egy adott nyelvi sáv megváltoztatása vagy nyelvi sávok elrendezése (nyelvi változat)
- A feliratozás változásai (feliratozott verziók)

- Változások a konkrét kép vagy hang formátumokban, mint például: széles képernyős vs. „pan & scan”, normál felbontású vs. nagy felbontású, 2D vs 3D, „kiváló minőségű” és „rossz minőségű” (pl D1 szalag átmásolása VHS kazettára, minőségromlással történő átkódolás)
- Átkódolás (pl. váltás AC3-ről MP4 hangkódolásra)
- Az Audiovizuális mű fizikai hordozójának megváltozása
- Egy Audiovizuális mű meghatározott célra történő szerkesztése, például: hosszúsági szerkesztések, a kereskedelmi forgalomba hozatalhoz szükséges módosítások, tartalomszerkesztési célú változtatások.
- A technikai adatfolyamban bekövetkezett változás, amely befolyásolja a mű tartalmát (pl. egy másik szoftver, amely háttérképeket generál, vagy a függőleges üres intervallum a feliratozáshoz).
- Az Audiovizuális mű formátumának és/vagy időtartamának változása (pl. televíziós közvetítés, légitársaságok számára történő használat vagy hasonló változások egy adott felhasználási mód követelményei szerint)
- A mű digitalizált verziója eredetileg analóg formátumban
- A rendezői vágás;
- Jelentősen rövidített és újratervezett verzió;
- Egy eredetileg fekete-fehérben létrehozott Audiovizuális mű "színesített" verziója, vagy egy eredetileg színesben előállított Audiovizuális mű "fekete-fehér" verziója;
- Az Audiovizuális mű más jelentősen átdolgozott változatai.

Az ISO 15706-2 felhatalmazza továbbá az ISAN-Verzió alkalmazását olyan más tartalmakra is, amelyek az Audiovizuális művek verzióiból származnak, vagy ahhoz szorosan kapcsolódnak, ha ezt a tartalmat tömeges terjesztésre szánják, például műsorszórás céljából, és amikor szükséges a tartalom azonosítása egy létező ISAN vagy ISAN Verzió alkalmazásának szélesebb körű kontextusában. Ilyen kapcsolódó tartalom lehet:

- a sugárzott Audiovizuális mű-ből nyert hangsáv
- leíró hangsáv
- feliratozó sáv,
- több rendszeres hangsáv
- egy Audiovizuális mű hangsávjából kivont csengőhang
- egy kivonat ujjlenyomata vagy a teljes Audiovizuális mű
- egy kivonat audió ujjlenyomata vagy teljes hangsávja.
- az Audiovizuális mű plakátja
- az Audiovizuális művek dobozos készlete
- stb ...

A következő módosítások nem minősülnek különálló Audiovizuális mű-nek vagy ISAN Verzió-nak:

- változás az Audiovizuális mű jogaiban vagy tulajdonosában
- másolás azonos formátumú adathordozók között (pl. szalag-szalag másolat)
- változás a művet leíró metaadatokban
- az Audiovizuális mű használatával kapcsolatos árak, vagy díjak módosítása

Az ISAN Művek és a Verziók közötti kapcsolat

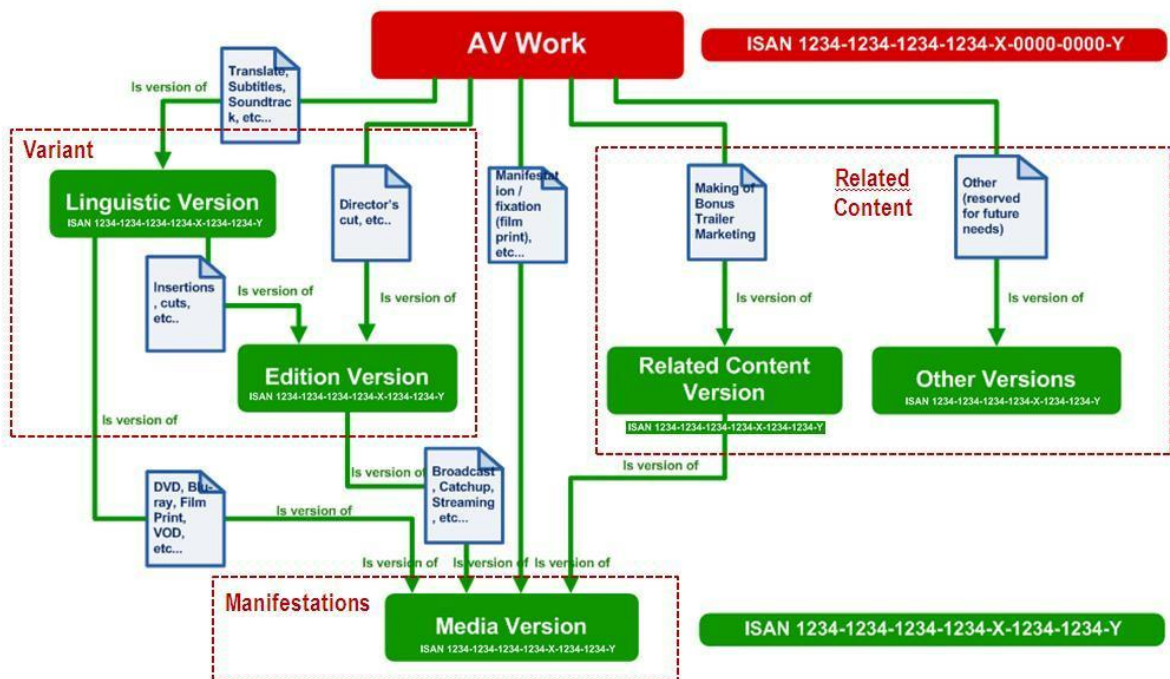
Az ISAN azonosító egy 24 számjegyű szám, amely két fő részből áll:

- az első 16 számjegyet az ISO 15706-1 szabvány határozza meg Audiovizuális mű azonosítójaként;
- a következő (és az utolsó) 8 számjegyet az ISO 15706-2 mint Verzió Azonosító kiterjesztést határozza meg.

A mű és a verzió közötti kapcsolat natív módon szerepel az ISAN Azonosítóban; tulajdonképpen minden változat ugyanazon az első 16 számjegyen osztozik csak az ISAN utolsó 8 számjegye különbözik egyik verzióról a másikkra.

Az Audiovizuális művek verziói közötti különbségeket, valamint az azonos művek verziói közötti kapcsolatot (ha vannak) az egyes verziók metaadataiban definiálják.

Ugyanazon Audiovizuális művek különböző verziói összekapcsolhatók az FRBR alapú modellekkel, ill. más hierarchiák engedélyezéséhez, ahogy az például az alábbi sémában látható.



audiovizuális művek verzióinak összekapcsolása FRBR-alapon

Audiovizuális sorozatművek és epizódok

Az ISAN-rendszer céljaira az Audiovizuális sorozatművek és az epizódok tág értelemben kerülnek meghatározásra. Következésképpen az epizódok között semmilyen nyilvánvaló szekvenciális allokációk sem jelzik azt a sorrendet, amelyben az epizódokat előállították vagy továbbították. Egy Audiovizuális sorozatművet különálló epizódokban vagy olyan részekben állítanak elő amelyek kapcsolatban állnak egymással. Általában egy közös cím van az egész sorozat számára. Példák Audiovizuális sorozatművekre:

- minden olyan televíziós sorozat, műfaji megkötés nélkül, amelyet rendszeres időközönként sugárzott epizódokban állítanak elő;
- napi / heti talk show, hírközlés, sport esemény, stb ...;
- napi vagy rendszeres hírműsorok;
- Ezek NEM Audiovizuális sorozatművek:
- egy film folytatása (pl „Rocky IV”);
- egy Audiovizuális mű részeinek különálló sugárzása;
- évenkénti esemény éves tv-közvetítése (például egy szilveszteri műsor, vagy az Oscar Díj közvetítése).

Összetett művek és összetevőik

Egy összetett mű, olyan audiovizuális alkotás, amely teljesen más Audiovizuális műveket vagy Audiovizuális művekből származó kivonatot tartalmaz, és amelyek mindegyike lényegtelen az egész összetett alkotáshoz képest.

Egy összetett Audiovizuális mű saját ISAN-al rendelkezik, függetlenül bármely olyan ISAN-tól, amely az összetevői számára lett kibocsátva.

Példák az összetett művekre:

- Egy olyan film, amely animációs rajzfilmeket vagy más filmklippeket tartalmaz;
- Egy olyan televíziós műsor, amely korábban előállított filmekből vett részleteket tartalmaz;
- Egy televíziós epizód, amely más tévésorozat epizódjait tartalmazza.

A gyűjtemények mindig összetett művek.

Példák NEM összetett művekre:

- Egy bizonyos jelenet egy játékfilmen belül, kivéve, ha az adott jelenetet önálló terjesztésre szánják;
- Egy képmontázs.

Összetett művek esetén a regisztrálónak meg kell adnia, az összetett művet alkotó minden egyes Audiovizuális mű rendelkezésre álló ISAN számát.

Az ISAN AV művekhez történő hozzárendelése

Az ISAN-t permanensen kell hozzárendelni AV/mű mesterpéldányához annak érdekében, hogy az minden ilyen AV /mű egész életciklusa alatt fennmaradjon.

Digitális formátumú audiovizuális tartalmak esetén, az ISAN magába a műbe kódolható. Sok digitális formátumban alkalmazott audiovizuális szabvány ISAN kompatibilis, úgymint MPEG, Windows Media Format, DVB, ATSC, Blu-ray (AAC3) UltraViolet, Digital Cinema (DCI), MXF, BXF, Electronic Program útmutatók (Open EPG, TV Anytime), stb

Analog formátumú audiovizuális tartalmak esetében, ajánlott, hogy az ISAN:

- Legyen hozzárendelve a mesterpéldányhoz vagy permanensen hozzákapcsolva a mesterpéldány tárolásával megbízott intézmény nyilvántartáshoz.
- Legyen láthatóvá téve az AV/mű fő- és /vagy végfelsorolásában.
- Javasolt továbbá, hogy minden egyes AV/művel együtt forgalmazott valamennyi kereskedelmi és műszaki dokumentáció tartalmazza az ISAN -t.

Amikor az ISAN ki van nyomtatva vagy másképpen ábrázolva, akár egy Audiovizuális Tartalom-on, címkéken, műszaki és kereskedelmi dokumentáción, és a csomagoláson vagy máshol, akkor azt az „ISAN” betűknek, kell megelőzniük, hogy világosan megkülönböztethető legyen minden más termék számtól. Még ha nem is római számokat használnak, az „ISAN” szónak latin betűkkel kell szerepelniük.

Az ISAN pontos átírását elősegítendő, a következő szabályokat kell alkalmazni:

- az „ISAN” szónak, plusz egy üres helynek meg kell előznie minden számjegyet;
- a 16-jegyű ISAN-t fel kell osztani 4 egységnyi 4 számjegyre ahol minden egyes egység szóközzel vagy kötőjellel van elválasztva, amit az ellenőrző karakter követ;
- majd a 8 számjegyű verzió-hosszabbítást kell felosztani 2 egységnyi 4 számjegyre ahol minden egység szóközzel vagy kötőjellel van elválasztva, amit az ellenőrző karakter követ;

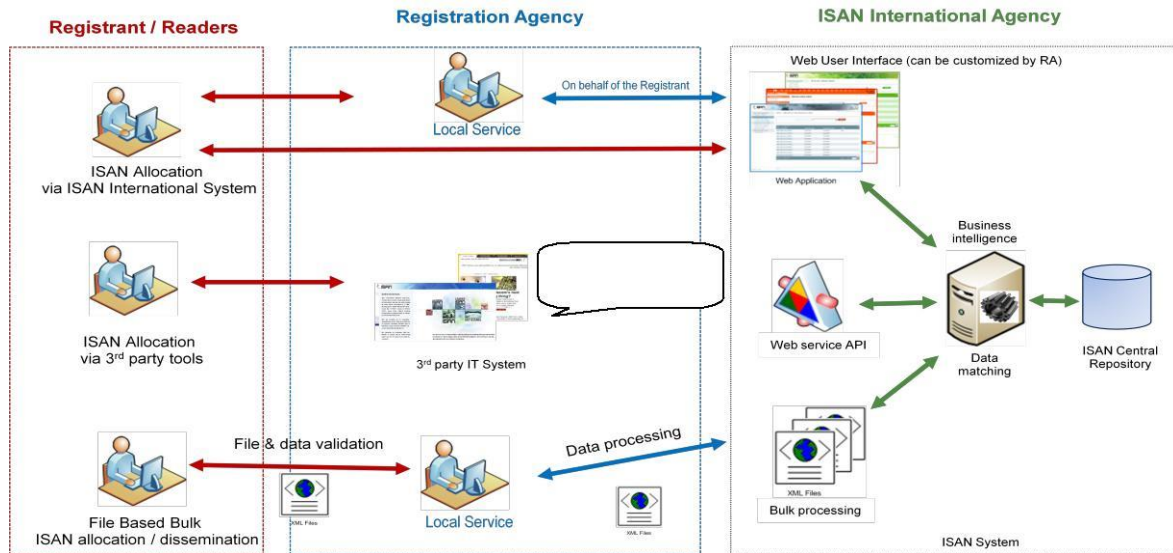
Ezek a szabályok kizárólag vizuálisan támogatják az ISAN pontos átírását; semmilyen jelentés sincs bármelyik számjegyhez, egységhez, vagy a karakterekhez rendelve. Az ISAN-al kompatibilis jelentős szabványokra és alkalmazásokra vonatkozó utalások megtalálhatók itt: www.isan.org

Az ISAN rendszerhez való hozzáférés

Az ISAN-t és a kapcsolódó metaadatokat a ISAN Központi Adatbázis tárolja, amely éjjel-nappal online elérhető, regisztrációt, konzultációt és kereséseket tesz lehetővé. Az ISAN rendszer, az infrastruktúra és az alkalmazások a svájci Genf egy ultramodern létesítményében üzemel, és nyújt biztonságos környezetben szolgáltatásokat, úgymint 24x7x365 ellenőrzést, teljesen redundáns biztonságos internet-kapcsolatot és teljes heti ill. napi back-up-ot. Az ISAN alkalmazás üzemeltetését és karbantartását az ISAN-IA mérnökei és külső szolgáltatók végzik.

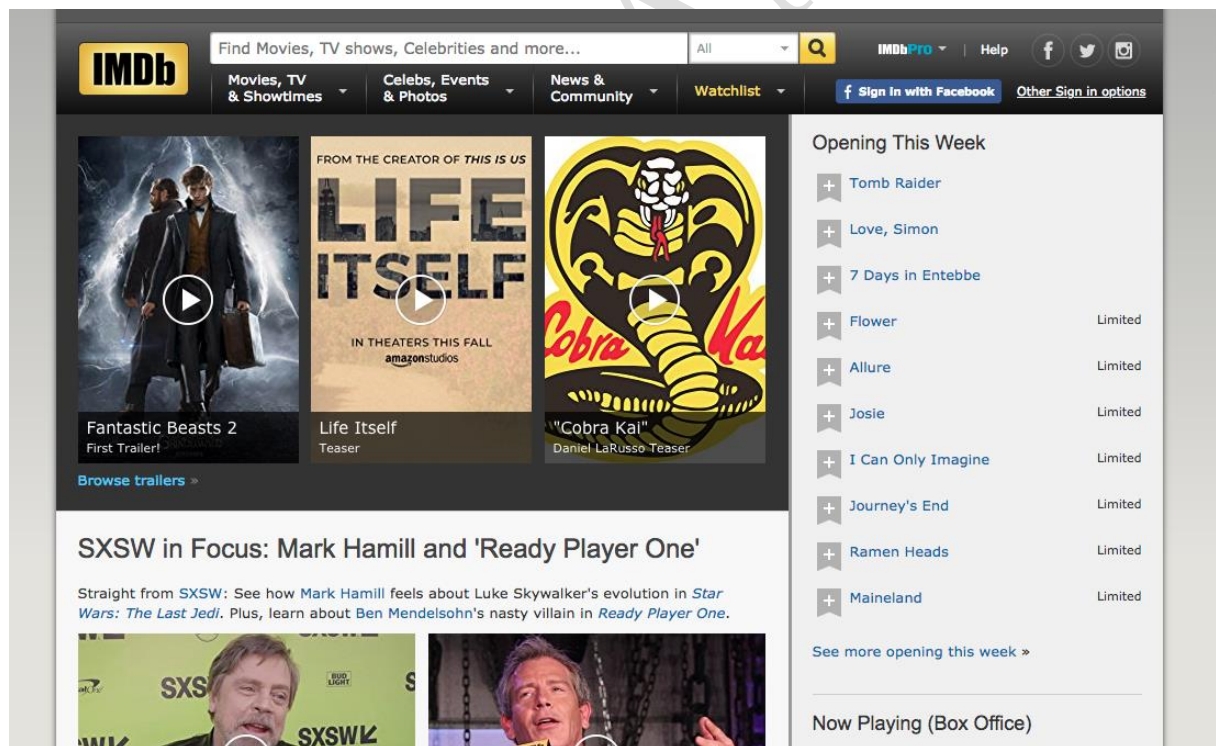
Az ISAN adatbázis a regisztráláshoz és az interneten keresztüli lekérdezésekhez/keresésekhez a következőket kínálja.

- Webszolgáltatások, amely lehetővé teszik harmadik fél (mint például műsorszolgáltatók, stúdiók, digitális platformok, stb), informatikai rendszerének felhasználói felületének összekapcsolását a ISAN adatbázissal.
- Az XML fájlon keresztül történő tömeges feldolgozás nagy katalógusok számára is hatékony regisztrációt tesz lehetővé, drága műszaki beruházások, mint pl. az API alkalmazása, nélkül..
- AZ ISAN Adatbázis helyi másolatai (Oracle mentés) a Támogatói Olvasók rendelkezésre áll.



2.10.3.2 IMDb

Az Internet Movie Database (IMDb), melynek tulajdonosa az Amazon.com, egy online információs adatbázis filmszínészekről, filmekről, filmsorozatokról, egyéb televíziós műsorokról, reklámokról és videojátékokról. A regisztrált felhasználók értékelhetik a filmeket és szerkeszthetik a cikkeket. Az IMDb a legismertebb és legnagyobb méretű filmes internetes adatbázis.



az IMDb kezdőoldala ma

Az IMDb hatalmas mennyiségű információt tesz elérhetővé az alkotásokról, többek között alapvető adatokat, mint a színészek és rendezők neve, a stáb tagjai, a történet, kritikák, valamint kevésbé ismert adatokat, mint a háttérinformációk, baklövéses, zenei kiadványok listája, képarány és alternatív verziók. A színészeknek, rendezőknek, íróknak és egyéb résztvevőknek is van saját adatbázis információjuk,

felsorolva a filmeket és adásokat, melyekben részt vettek, továbbá a legtöbb esetben életrajz is olvasható róluk.

Az IMDb több, mint egy film és videojáték adatbázis, mivel napi film és TV híreket közöl, valamint kiemelt információkat tesz elérhetővé nagyobb filmes eseményekről, például az Oscar-díj átadásról. Az IMDb egy aktív közösséggel is rendelkezik. Minden filmhez tartozik egy szál a lap alján, továbbá általános vitalapok is vannak. Kibővítették a szolgáltatásokat a hivatásos szakemberek számára, így létrehozva az IMDbPro adatbázist, ahol a filmiparban szereplő személyek elérhetőségét, esemény naptárat és sok más szolgáltatás érhető el. Az IMDbPro nem a nyilvánosság számára készült, és így a tartalom sem ingyenes.

| FILMS IN PRODUCTION (2 TITLES) | YEAR | BUDGET | STATUS |
|---|------|--------|-----------|
| Billy Lynn's Long Halftime Walk – Denise Lynn | 2016 | | Post-Prod |
| Thirsty – Doris Townsend | 2015 | | Post-Prod |

| PAST FILM & VIDEO (14 TITLES) | YEAR | BUDGET | U.S. GROSS |
|----------------------------------|------|--------|------------|
| Lionhead – Detective Lundgren | 2013 | \$2MM | |
| Killing Vivian (Short) – Toaster | 2013 | | |
| Bad Teacher – Sasha's Mother | 2011 | \$20MM | \$100MM |

az IMDbPro felülete

Mindenki, aki egy e-mail címmel és egy cookie-kat támogató böngészővel rendelkezik, regisztrálhat az IMDb-re, majd feltölthet információkat, véleményezheti a filmeket, rangsorolhatja őket. A lekérdezések eredménye letölthető (tömörített) szöveges fájlként és a tartalmuk kinyerhető (jellemzően parancssoros eszközökkel). Az ellenőrzés hiányosságai miatt az adatbázisba téves információk is bekerülhetnek (például amatőr, botcsinálta vagy akár teljesen önkényes címfordítások), így a tartalom mindig fenntartással kezelendő.

1989-ben Col Needham és mások jelen voltak a rec.arts.movies Usenet hírcsoportban, ahol filmekről beszélgettek és információkat osztottak meg egymással. GYIK (Gyakran ismételt kérdések) listákat hoztak létre színészekről, színésznőkről és rendezőkről, adatokat és bibliográfiát közöltek a filmkészítőkről, akik elhunytak. 1990 végén közel 10 000 filmről és televíziós sorozatról volt már GYIK. 1990. október 17-én Needham Unix héjprogramokat küldött be, amik képesek voltak keresni a négy GYIK listában, és így megszületett az IMDb. Abban az időben azonban rec.arts.movies filmadatbázisként volt ismert.

az IMDb arcuata 25 éve

1993-ra az adatbázist kiegészítették háttérinformációkkal, bibliográfiákkal és tartalmi összefoglalókkal, és egy központosított e-mail-felülettel a lekérdezések egyszerűsítésére. Még ebben az évben áttették World Wide Webre. Az adatbázis a walesi Cardiff Universityn található szervereken fut. Rob Hartill volt az eredeti webes felület alkotója. 1994-ben az e-mail-felületet továbbfejlesztették, hogy képes legyen fogadni információkat, így már nem kellett a lista karbantartónak elküldeni a változásokat. Az évek során tartalom tükrök hálózata alakult ki a világ minden táján – adományozott sávszélességgel. Minden közreműködő, aki tartalmat hoz létre az adatbázisban, szerzői jogokkal rendelkezik az általa létrehozott tartalomra, de teljes jogot ad másolásra, módosításra, valamint hozzáférést biztosít magának az IMDb-nek is. Az IMDb nem teszi lehetővé mások számára írásos engedély nélkül, hogy felhasználják a film összefoglalókat vagy a színészek életrajzeit. A reklámszűrő szoftverek használata az oldalon kifejezetten tilos. A filmográfiák csak egy kis részét lehet idézni, és azokat is csak nem kereskedelmi weboldalakon. Az utóbbi adatfelhasználási megkötés nem érvényesíthető az Amerikai Legfelsőbb Bíróság ítélete értelmében (Feist v. Rural), mivel a tények gyűjtése nem tartozik a szerzői jog hatálya alá.

Az IMDb tipikus crowdsourcing jelenség, így magán viseli azok szabályozatlanságát, kontrollálhatatlanságát. Ennek ellenére ma már megkerülhetetlen tényezőnek számít a filmek, mozgóképek világában, és egy audiovizuális nemzeti névtér számára természetesen adódó nemzetközi kapcsolódási pontot jelent.

2.11 Személyek azonosító rendszerei

A személyek azonosítása minden területen lehet valamiért fontos, de a kreatív szakmákban sokszor szükség lehet arra, hogy megtaláljunk kutatókat, tudósokat, művészeket, írókat stb., akik valamilyen tevékenységük révén bekerülnek a nyilvánosság valamelyik szegmensébe. Mivel régóta tudjuk, hogy a személynevek önmagukban nem alkalmasak arra, hogy rajtuk keresztül megtaláljuk a minket érdeklő embereket, ezért régtől fogva komoly igény mutatkozott a személyek azonosítására képes rendszer bevezetése iránt.

2.11.1 Tudományos kutatók és cikkek azonosító rendszerei

A tudomány nemzetközi jellege miatt nem meglepő, hogy az elmúlt évtizedekben a tudományos világban több kezdeményezés is elindult a kutatók egyértelmű azonosítása érdekében. Ezt a célt nyilván csak egyetlen nemzetközi rendszerrel lehet csak elérni, de azt ma még nem lehet látni, hogy a jelenleg versengő rendszerek közül melyik lesz majd a „befutó”.

2.11.1.1 MTMT

A Magyar Tudományos Művek Tára (MTMT) a magyarországi szerzők hiteles, tudományos publikációs adattára. A publikációk és hivatkozások adatait ide fel kell tölteni, melyek hitelességét könyvtárosok ellenőrzik. Ezek alapján történik a publikációs teljesítések igazolása a fokozatszerzések, habilitáció, pályázatok esetén. Az MTMT létrehozásának célja az volt, hogy legyen egy – többféle célra is hasznosítható – nemzeti tudományos bibliográfiai adatbázis Magyarországon. A Magyar Tudományos Akadémia, a Magyar Akkreditációs Bizottság, a Magyar Rektori Konferencia, az OTKA és az Országos Doktori Tanács 2008 második felében kezdte meg az egyeztető tárgyalásokat. Az öt alapító intézmény, továbbá a koncepciót támogató OKM és NKTH egyetértett abban, hogy a hazai tudományos kutatás hiteles nyilvántartását és bemutatását kell megvalósítani az MTMT, illetve a rendszerhez csatlakozó szervezetek segítségével.

Az MTMT használatából származó előnyök a következők.

- Országosan egységes rendszerben vehet részt az intézmény;
- Fejlesztése és üzemeltetése lényegesen olcsóbb, mint a helyi fejlesztés és működtetés;
- Van kipróbált előzménye, a KPA, amit sok kutató ismer. Használták az ODT, az OTKA, az MTA igényeinek megfelelően;
- Az adatok sok célra felhasználhatók;
- Alkalmas belső igények kielégítésére (habilitáció, belső pályázatok, doktori képzés, statisztika, vezetői információs rendszer (VIR, AVIR) egyes adattípusaihoz);
- A külső minőségbiztosítás beépül a rendszerbe;
- Széles szakértői bázison működik, képzési lehetőséget biztosít;
- Nagyobb szervezetek adatigényeivel kompatibilis;
- Non-profit szervezet működteti.

Az MTMT használata előnyös a kutatók számára is. Miért jó a kutatóknak?

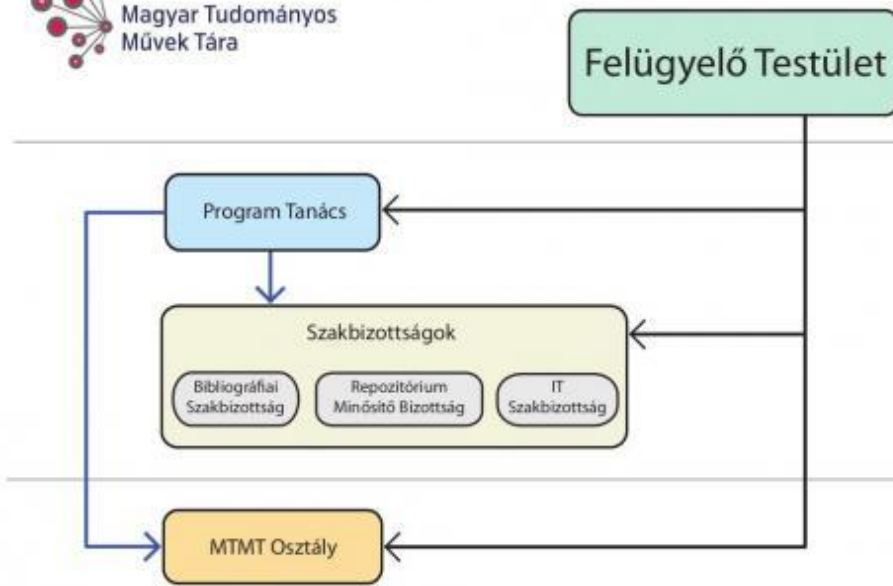
- Személyes tudományos bibliográfia összeállítása egyszerű;
- Kis munkaigénnnyel karbantartható;
- Segíti pályázatok és jelentések készítését;
- Széles körben ismert formátum;
- Személyes honlapra beilleszthető;
- MTA, ODT, MAB, OTKA elfogadja, preferálja.

Miért jó a könyvtáraknak, könyvtárosoknak?

- Kiegészítő könyvtári funkciókat nyújt (folyóiratcikk nyilvántartás, idézetek nyilvántartása);
- Egységes folyóirat publikációs és idézési nyilvántartás;
- Ügyfélszolgálat (probléma megoldás, szoftverhasználati segítség).

Az együttműködés jogi alapja és szervezeti kerete

Az együttműködés a rendszerhez önkéntesen csatlakozó intézmények és az MTMT-t működtető MTA Könyvtára (a továbbiakban: MTAK) közötti szerződés alapján valósul meg. A felhasználók összességét az MTMT Program Tanács (PT) képviseli. A Felügyelő Testület (FT) ellenőrzi a tevékenységet, és dönt a hatáskörébe tartozó kérdésben.



a MTMT szervezete

Az MTMT operatív működését az MTA Kutatásszervezési Intézete (KSZI) szervezeti egységként létrehozott, majd annak jogutód nélküli megszűnése után az MTAK szervezeti keretein belül működő MTMT Osztály (korábban Koordinációs Iroda) biztosítja. Az Osztály folyamatos kapcsolatot tart azokkal a könyvtárosokkal és/vagy más munkatársakkal, akik az egyes intézmények adatbázisát karbantartják.

Az MTMT szolgáltatásai.

- Modern informatikai eszközökkel hiteles tudományos eredmény-nyilvántartást biztosít, és bemutatja az eredményeket, magyar és részben angol nyelven.
- Nyilvános, hiteles, átvehető és hivatkozható publikációs listákat ad. Ezek felhasználhatók az intézményi akkreditációban, a doktori iskolákkal kapcsolatos követelmények teljesítése során, MTA, OTKA és NKTH pályázatoknál.
- Segíti az akadémikus tagajánlásokat, az "MTA doktora" pályázatok beadását és azok értékelését.
- Bibliográfiai adatok révén átmenetet biztosít a teljes szövegű közleménytárhoz (repozitórium).
- Az adatot szolgáltató intézmények tudományos teljesítményéről összesített képet ad. Ez felhasználható az adott intézmények önértékeléséhez és/vagy külső értékeléséhez.
- A rendszer automatikus adatfeltöltést és megfelelő előkészítéssel adatkonverziót is végez, meglévő adattárakból és könyvtári rendszerekből.

A szolgáltatások korszerűsítésén az MTMT Osztály a források figyelembe vételével folyamatosan dolgozik.

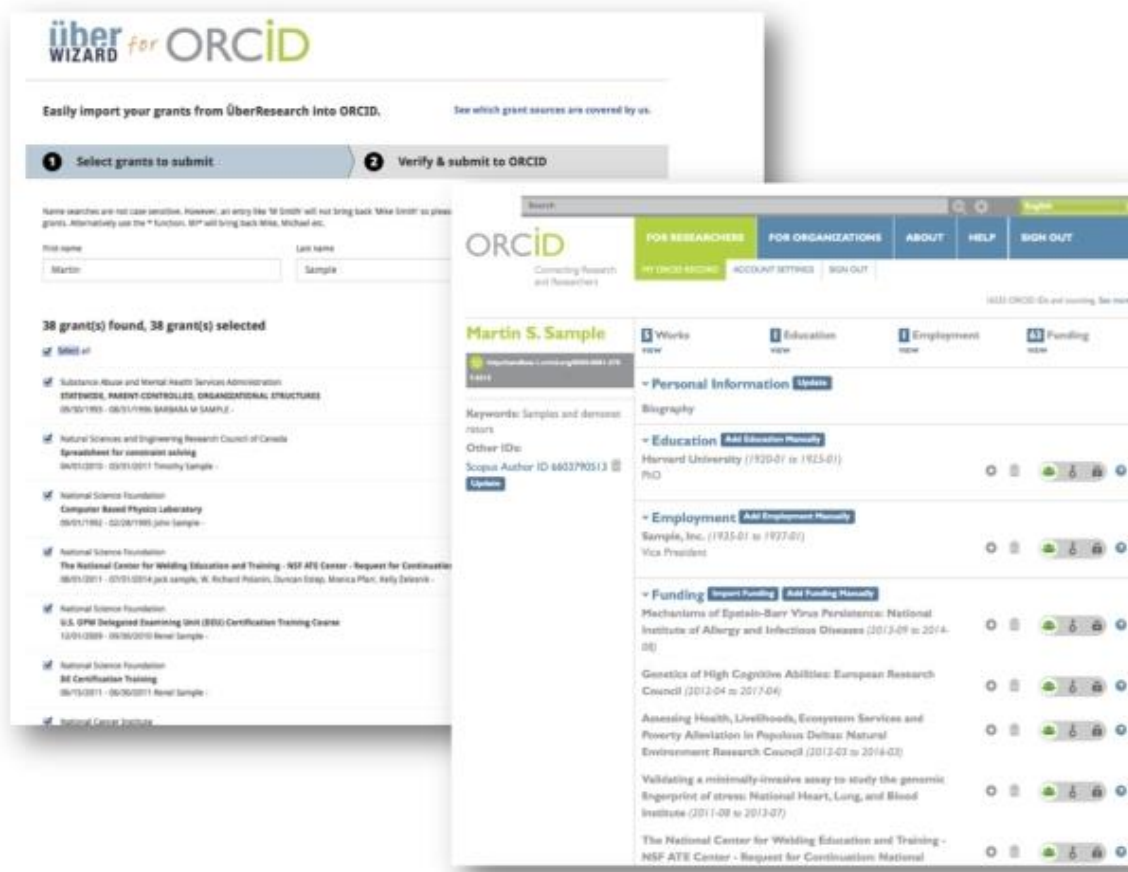
| Személyes adatok | |
|---|--|
| |  |
| név | Szakadát István |
| születési év | 1958 |
| intézmény neve | Eötvös Loránd Tudományegyetem Pécsi Tudományegyetem Budapesti Corvinus Egyetem |
| doktori iskola | ELTE Szociológia Doktori Iskola (oktató) PTE Nyelvtudományi Doktori Iskola (oktató) BCE Társadalmi Kommunikáció Doktori Iskola (témavezető) |
| doktori képzéssel kapcsolatos munkájának megoszlása | ELTE Szociológia Doktori Iskola 50% BCE Társadalmi Kommunikáció Doktori Iskola 50% |
| adott-e már oktatóként valamely doktori iskolát működtető intézménynek akkreditációs nyilatkozatot? | Budapesti Műszaki és Gazdaságtudományi Egyetem |
| Elérhetőségek | |
| drótpostacím | i@syi.hu |
| telefonszám | +36 1 458-0633 |
| | saját honlap |
| | saját honlap (angol) |
| Fokozat, cím | |
| tudományos fokozat, cím | PhD |
| fokozat megszerzésének éve | 1999 |
| fokozat tudományága | szociológiai tudományok |
| fokozatot kiadó intézmény neve | MTA |
| Jelenlegi munkahelyek | |
| 1997 - | Budapesti Műszaki és Gazdaságtudományi Egyetem egyetemi oktató |

személyi adatlap az ODT oldalán

2.11.1.3 ORCID

A tudományos publikációk szerzőinek automatikus azonosítása hozták létre az egyre szélesebb körben elfogadott ORCID (open researcher and contributor ID) szerzői azonosítási rendszert. A rendszer terjedését nagyban segíti az a tény, hogy az ORCID-ban rögzített személyes adatokat egyre több folyóirat, kiadó automatikusan áttemeli saját rendszerébe (például egy kézirat beküldésekor). Az ORCID-ban a kutatók publikációk adatai is tárolhatók, és ezek szinkronizálhatók más adatbázisok (pl. Scopus, ResearcherID, Crossref) listáival.

Az ORCID a kutatók azonosítását egy 16 számjegyből álló egyedi azonosítóval oldja meg. A számok egy központi regiszterben tárolódnak, ahonnan sokféle irányban lehet elvinni az egyértelműen beazonosítható kutatók adatait.



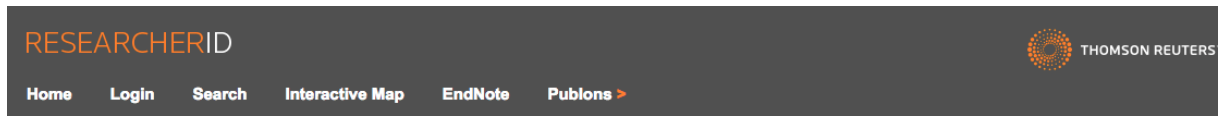
keresési eredmények és szerzői profil az ORCID oldalon

Az ORCID összefűli a ResearcherID, a Scopus Author ID-nk vagy a CrossRef által indexelt cikkek adatait, s onnantól automatikusan lehet nyomon követni a kutató publikálási tevékenységének és cikkeinek hatását. Az ORCID-felhasználók száma 2013 óta folyamatosan nő, 2018 tavaszán az élő ORCID-azonosítóval rendelkező kutatók száma meghaladta a 4,5 milliót (<https://orcid.org/statistics>). Az ORCID 37 műtípust támogat, a folyóiratciktől kezdve a táncelőadásig bezárólag, a tudományos és művészeti produktumok széles skáláját képes kezelni. Az ORCID kapcsolatot kínál más kutatói szolgáltatáshoz. Össze lehet kötni az ORCID profilt olyan weboldalakkal, mint a Web of Science (<http://wokinfo.com/researcherid/integration/>), Figshare (<https://figshare.com/>) vagy az Impactstory (<http://impactstory.org/>). Az összekapcsolások eredményeként könnyedén áramolhatnak az információk minden irányban, így nem kell ugyanazt az információt több helyre is feltölteni. Nap mint nap új szolgáltatások kapcsolódnak az ORCID-hez, így egyre növekvő számú platform, repozitórium, pályázati alap között áramlik az információ. Folyóiratok, alapítványok és intézetek mozdulnak el az ORCID irányába. Több mint ezer folyóirat – többek között a PLOS, Nature és Elsevier által kiadottak – tette lehetővé a szerzőik számára, hogy az ORCID használatával kezeljék az információkat a kézirat feltöltő rendszerekben. Az ORCID képes arra, hogy a szerző megjelent publikációit összegyűjtse és a cikkekre vonatkozó metrikák aggregálásával szerzői szintű tudományometriai adatokat mutasson. Az ORCID szolgáltatása nyílt forráskódú, szabadon lehet az API-ját használni, lehet az adataiban bányászni.

2.11.1.4 ResearcherID

A ResearcherID a Thomson Reuters Web of Science adatbázisával összekapcsolt, szerzői azonosítót biztosító szolgáltatás. Létrehozható benne szerzői adatbázis, szinkronizálható az ORCID adatbázissal. A ResearcherID szerzői profil létrehozása és használata külön felületen, a <http://www.researcherid.com/> címen történik. Az oldalon a regisztrációt követően lehetőség van egyéni publikációs listák és

hivatkozási riportok létrehozására, valamint a szerzőnek a Web of Science adatbázisban szereplő publikációinak magához rendelésére.



Szakadát, István [Get A Badge](#) [ResearcherID Labs](#)

ResearcherID: H-3517-2012
Other Names: Syi
URL: <http://www.researcherid.com/rid/H-3517-2012>
Subject: Mathematical Methods In Social Sciences; Sociology
ORCID: <http://orcid.org/0000-0002-1960-191X>

My Institutions (more details)
Primary Institution: Budapest University of Technology and Economics, BME
Sub-org/Dept: Department of Sociology and Communication
Role: Researcher (Academic)

My URLs: <http://syi.hu>

My Publications

My Publications (8)
[View Publications](#)
[Citation Metrics](#)

ResearcherID labs
[Create A Badge](#)
[Collaboration Network](#)
[Citing Articles Network](#)

My Publications: View

This list contains papers that I have authored.

8 publication(s) Page 1 of 1 Go Sort by: Publication Year Results per page: 10

- Title: Hidden Patterns of Reciprocity
Author(s): Syi
Source: Journal of Theoretical Biology Volume: 345 Pages: 92-98 Published: 2014
added 19-Feb-14
- Title: MEO ontology infrastructure
Author(s): Szakadát, István; Szots, Miklos; Gyepesi, Gyoergy; et al.
Source: Advances in Information Systems Development, Vol 1 Pages: 369-377 Published: 2007
Times Cited: 0
DOI: 10.1007/978-0-387-70761-7_32
added 25-Sep-12
- Title: Dynamics of information access on the web
Author(s): Dezso, Z.; Almaas, E.; Lukacs, A.; et al.
added 25-Sep-12

kutatói adatlap a ResearcherID oldalon

A ResearcherID felületére belépve a My Researcher Profile menüpontban érhetőek el a szerző adatai és publikációs listája. Az oldal felső részében megtalálható a szerző ResearcherID azonosítója és a profil URL címe, amely a szerző adatlapjára vezető linkként más weboldalakon is feltüntethető. Az oldal ORCID mezője lehetőséget ad a szerző ResearcherID és ORCID szerzői profiljának összekapcsolására. A szerző a My Researcher Profile oldal My Publications nevű részében állíthatja össze publikációs listáját. A Web of Science felületén megtalálható saját publikációk az adatbázisban történő keresést követően a közlemények szerzőhöz rendelésével vehetők fel a listára, de az adatbázis által nem indexelt publikációk is hozzáadhatók. A Web of Science heti rendszerességgel frissíti a ResearcherID azonosítókhoz rendelt adatbázisban szereplő közleményeket a My Publications listák alapján, azoknál a szerzőknél, akik nyilvánossá tették profiljukat. A saját publikációkat regisztráló lista gyakori frissítése és pontos vezetése tehát fontos ahhoz, hogy a Web of Science felületén az azonosító alapján egyértelműen megtalálhatóak legyenek az adott szerző adatbázisban szereplő közleményei.

Az Add vagy Add Publications gombra kattintást követően megjelenő új oldalon a Web of Science adatbázisból történő adatátteleléshez és a közlemények szerzőhöz rendeléséhez juthatunk. További lehetőségekként az EndNote bibliográfiakezelő program webes felületéről történő adatimportálás, illetve saját RIS formátumú fájlként publikációs lista feltöltése érhető el. A Web of Science adatbázisban való keresés és az adatok importálása, azonban csak az adatbázis hozzáférési jogosultságával rendelkező felhasználóknak érhető el.

publikáció hozzáadása a My Researcher Profile oldalon

A Search Web of Science Core Collection opció kiválasztása után megjelenő oldalon a szerző vezetéknévének és a keresztnév kezdőbetűjének megadása kötelező ahhoz, hogy a Search gombbal közleménykeresést lehessen indítani az adatbázisban. A pontosabb eredményhez azonban érdemes további kritériumokat is beállítani.

A találati listán a szerző által írt közlemények egyenként vagy csoportosan is kijelölhetők, majd az Add felírra kattintva hozzáadhatók a szerző ResearcherID profiljának My Publications listájához. A közleményekkel együtt az idézettségi adatok is átemelésre kerülnek. A saját közlemények egyértelmű beazonosításához érdemes a címre kattintva a részletes adatokat is előhívni.

ResearcherID azonosítót csatolni lehet az MTMT-ben a szerzői adatlaphoz.

ResearcherID azonosító csatolása szerzői adatlaphoz MTMT-ben

Az MTMT felületén a szerző a személyi adatlapjának Azonosítók, hivatkozások gombjára kattintva tud a profiljához szerzői azonosítókat csatolni. A linkként működő ResearcherID hozzáadásával az MTMT felületéről egy kattintással elérhetővé válnak a szerző Web of Science adatbázisban megtalálható publikációi és azok idézettségi adatai. Az azonosító hozzáadásához az Azonosítók, hivatkozások felírra kattintva megjelenő Típus mező legördíthető menüjéből a ResearcherID lehetőséget kell választani és az Azonosító mezőbe be kell másolni a ResearcherID azonosítót. Ezt követően a Hozzáad gombra kattintva az azonosító megjelenik a szerző személyi adatlapján.

Budapest: Akadémiai Kiadó, 2013. 73 p.

ISBN: 978-963-05-9449-3

Nyelv: Angol

Könyv/Konferenciakötet/Tudományos [2454434] kézi felvitel (Forrásév: 2013) (2013-11-13 [jövöhagyva])

Duplum vizsgálat Idézők szerkesztése Fejezetek, konferenciatickek listázása

Duplumnak jelölöm

Külső azonosítók

Konferencia helye, ideje: Budapest, Magyarország, 2013.11.05-2013.11.07.
Budapest: Akadémiai Kiadó, 2013. 73 p.

ISBN: 978-963-05-9449-3

Nyelv: Angol

Link(ek): WoS

Könyv/Konferenciakötet/Tudományos [2454434] kézi felvitel (Forrásév: 2013) (2013-11-13 [jövöhagyva])

Duplum vizsgálat Idézők szerkesztése Fejezetek, konferenciatickek listázása

**Külső azonosítók
kezelése, hozzáadása**

**Hozzáadott külső
azonosító**

Bezár

[sn]
(szerk.)
2nd Conference of Cereal Biotechnology and Breeding: Book of Abstracts
Konferencia helye, ideje: Budapest, Magyarország, 2013.11.05-2013.11.07.
Budapest: Akadémiai Kiadó, 2013. 73 p.

ISBN: 978-963-05-9449-3

Nyelv: Angol

Könyv/Konferenciakötet/Tudományos [2454434] kézi felvitel (2013-11-13 10:45:26) (xPuskás Lajosné (ATK MGI 4-es admin))

Típus: **Indexelő adatbázis** | **Azonosító** | **WoS** | **30033876610002** | Open Access | Embargó lejárata

Hozzáad **Mégsem**

Típus Azonosító OA Embargó Felvivő Érvényesítő

Külső azonosító típusának meghatározása

**Web of Science
azonosító beillesztése**

**Külső azonosító
hozzáadása**

ResearcherID azonosító szerkesztése az MTMT rendszerében

Az MTMT rendszerében a lehetőség van WoS azonosítók feltöltésére vagy szerkesztésére. A Web of Science azonosító hozzáadásához a Típus mezők legördíthető menüiből az Indexelő adatbázis és a WoS lehetőségeket kell választani. Ezt követően az Azonosító mezőbe beilleszthető az azonosító számsor, majd a Hozzáad gombbal csatolható a közleményhez. A Bezár gombra kattintva az azonosító megjelenik a közlemény alatt „Link(ek): WOS” formában, amire kattintva elérhető a publikáció Web of Science rekordja. Ha a Web of Science adatbázisból történő importálással került be az MTMT rendszerébe a publikáció, akkor az azonosító automatikusan áttemelésre került.

2.11.1.5 Web of Science

A Web of Science rendszeren frissülő hivatkozáskereső bibliográfiai adatbázis, amely számos tudományterület rangos szakfolyóiratait, szakkönyveit és konferenciaanyagait szemlézi. Lehetővé teszi ezáltal a publikációkra történt hivatkozások megtalálását, valamint a kutatási trendeknek és az intézmények tudományos életben betöltött pozíciójának a vizsgálatát.

Az adatbázis felületén egy szerző publikációi és azok adatai a Basic Search felületen szerző szerinti kereséssel érhetőek el.

Publikáció részletes adatai

Web of Science azonosító

Document Information

Document Type: Article

Language: English

Accession Number: WOS **000338766100002**

ISSN: 0006-3606

eISSN: 1744-7429

Journal Information

Impact Factor: Journal Citation Reports

Publikáció részletes adatai

Web of Science azonosító

Document Information

Document Type: Article

Language: English

Accession Number: WOS **000338766100002**

ISSN: 0006-3606

eISSN: 1744-7429

Journal Information

Impact Factor: Journal Citation Reports

publikáció és publikáció azonosító a Web of Science oldalon

A Web of Science adatbázisban minden publikáció rendelkezik egyedi azonosítószámmal, amely az MTMT rendszerében úgynevezett külső azonosítóként használható. Az MTMT-ben szereplő közleményekhez csatolható külső azonosítók linkként működnek, így általuk lehetővé válik az adott mű adatainak vagy teljes szövegének elérése, ellenőrzése.

2.11.1.6 Google Scholar

A Google Tudós (Google Scholar) a Google által kifejlesztett és fenntartott tudományos keresője, amely különféle témájú és formájú tudományos publikációk között keres. 2004. november 18-án indult, mára szinte minden online elérhető referált folyóiratban keres. 2006. február 20-ától a magyar könyvtárakban is keres a MOKKA-n keresztül.

A Google Scholar indexelő algoritmus egy-egy publikáció fontosságának meghatározásakor azt veszi elsősorban figyelembe, hogy más cikkek milyen gyakran hivatkoznak rá. A keresések a publikációk teljes szövegében történnek, szűkíthetők szerző, év vagy megjelenési hely alapján, a találatoknál a kereső megjeleníti a cím és az esetleges online elérhetőség mellett azt is, hogy milyen más tudományos munkák hivatkoznak rá, és milyen könyvtárakban található meg. Utóbbihoz a WorldCat adatbázist, és egy saját, Library Links nevű szolgáltatást használ.

A Google Scholar tevékenységével szemben megfogalmazódott az a kritika, hogy nem csak nem nyújt, de nem is akar teljes körű szolgáltatást nyújtani, mert csak a neten elérhető adatokra támaszkodik, márpedig a weben eddig a tudományos szakirodalomnak csak kisebb része jelent meg. Természetesen a hagyományos világ digitalizálásával, a retrospektív feldolgozó munkák befejeztével ez a kritika érvényét vesztheti.

a Google Scholar keresőfelülete

A Google Tudós esetén lehetséges szerzőként regisztrálni, ami után a szerző közleményeit és ennek hivatkozásait automatikusan a szerzői adatbázishoz rendeli. A gépi kereséseknek köszönhetően ez az összerendelés nem mindig pontos, de van lehetőség az adatok kézi szerkesztésére. Automatikus értesítést lehet kérni a cikkek új idézeteiről.

A Google Scholar többfajta mérőszámot is számol a kutatók tudományos hatásának mérésére.

2.11.1.7 Scopus

A 2004 novemberében indított Scopus szolgáltatás a lektorált szakirodalom legnagyobb absztrakt és citátum adatbázisa, amely intelligens eszközökkel kiegészítve lehetővé teszi a kutatások nyomon követését, elemzését és megjelenítését. A több mint 5000 nemzetközi kiadótól származó, több mint 19000 kiadvány mellett a SciVerse Scopus gyors, könnyű és átfogóan használható eszközt nyújt a kutatóknak, amely jól megfelel a kutatási igényeknek a természettudományok, a műszaki tudományok, az orvostudomány, a társadalomtudományok, és az utóbbi időben már a humán tudományok területein is.

The screenshot shows the Scopus search results page for the author last name 'smith'. The page displays 83,944 author results. The search criteria are 'Author last name "smith"'. The results are sorted by 'Document count (high-low)'. The table below shows the top three results:

| | Author | Documents | Subject area | Affiliation | City | Country/Territory |
|--------------------------|---|-----------|---|---|----------|-------------------|
| <input type="checkbox"/> | 1 Davey Smith, George Smith, George D. Smith, George Davey Davey-Smith, George | 1654 | Medicine ; Biochemistry, Genetics and Molecular Biology ; Agricultural and Biological Sciences; ... | University of Bristol | Bristol | United Kingdom |
| <input type="checkbox"/> | 2 Smith, James G. Smith, Jim Smith, J. Smith, J. G. | 1301 | Physics and Astronomy ; Mathematics ; Materials Science; ... | University of Colorado at Boulder | Boulder | United States |
| <input type="checkbox"/> | 3 Smith, Richard D. Smith, Richard O. Smith, R. D. Smith, R. | 1105 | Biochemistry, Genetics and Molecular Biology ; Chemistry ; Medicine; ... | Pacific Northwest National Laboratory | Richland | United States |

keresési eredmények a Scopus oldalon

A SciVerse Scopus fókuszában 2012 májusában több mint 19,500 kiadvány állt. Ezek közül 18,500 lektorált folyóirat (ezen belül 1,800 Open Access folyóirat), 400 szakmai kiadvány, 340 könyvsorozat volt. Az adatbázisban a 47 millió rekord közül 26 millió rekord 1996-ig visszamenő referenciákkal (melyekből 78% tartalmaz referenciát) rendelkezik, 21 millió rekord 1996 előttről, egészen 1823-ig visszamenően dolgozza fel az adatokat.

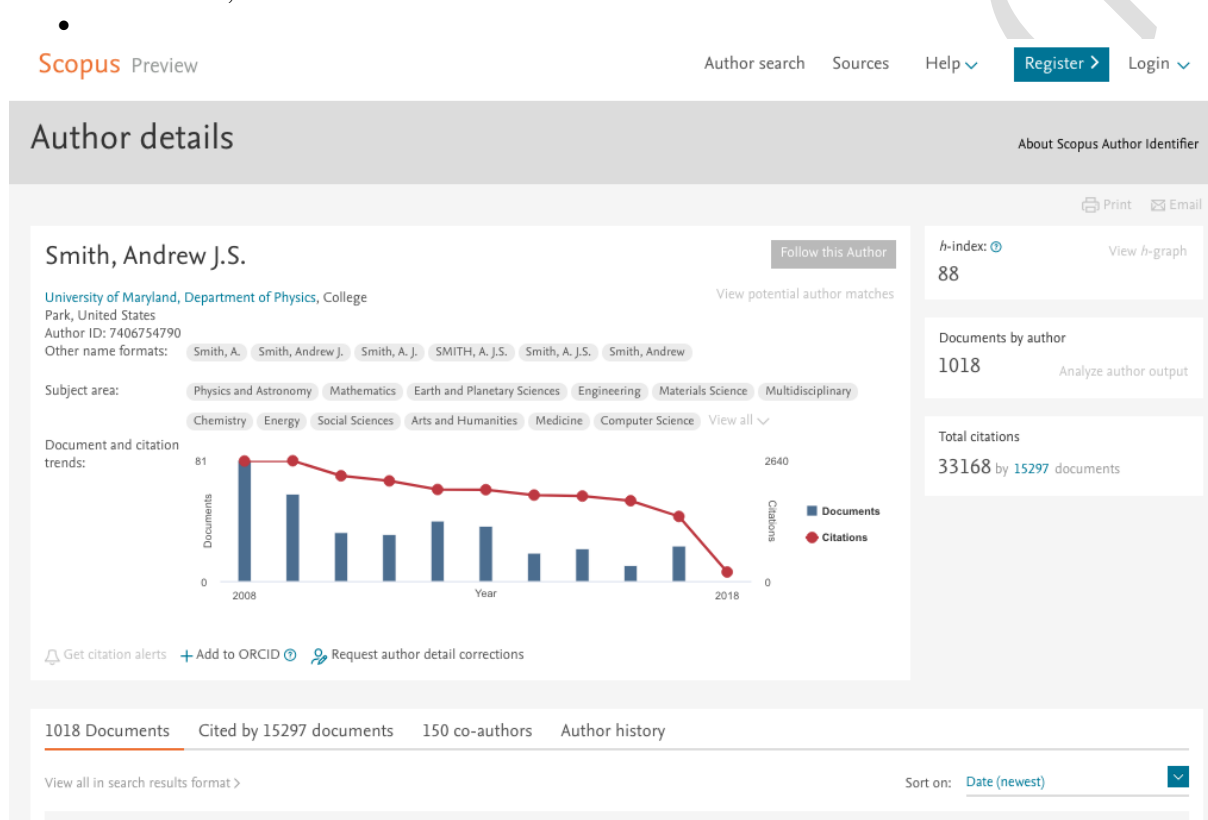
A folyóiratok megoszlása tudományterületek szerint:

- kémia, fizika, matematika és mérnöki tudományok: 4500 folyóirat;
- élet- és orvostudományok: 5900 folyóirat (benne a teljes Medline adatbázis);
- társadalomtudományok, pszichológia és közgazdaságtudomány: 2700 folyóirat;
- biológia, agrártudományok és környezetvédelem: 2500 folyóirat;
- általános tudományok: 50 folyóirat.

A Scopus adatbázis folyóiratainak 52%-a európai kiadású.

A Scopus szolgáltatásai között több olyan van, amely valamiféle névtér-kezdeményezésnek is tekinthető.

- Linkeket biztosít a teljes szövegű cikkekhez és egyéb könyvtári forrásokhoz.
- Mindenkinek biztosít egy azonosítót (Author Identifier), amelynek révén automatikusan hozzá lehet rendeni a szerző publikált kutatásaihoz a h-indexet (Hirsch indexet) a tudományos kutatás hatásának mérésére.
- Az idézet-nyomkövetővel (Citation Tracker) egyszerűen lehet keresni, ellenőrizni, követni a cikkek hivatkozásait valós időben.
- A rendszer biztosít egy intézményi azonosítót is (Affiliation Identifier), amely révén automatikusan azonosítani lehet, és az intézményhez lehet rendelni az összes kutatási eredményt.
- A Journal Analyser segítségével könnyen képet lehet kapni a folyóiratok teljesítményéről.
- Képzett kapcsolat van a SciVerse ScienceDirect, Reaxys, a ProQuest CSA Illumina irányába.
- Az adatokat könnyű exportálni a bibliográfia-kezelő alkalmazásokkal, mint például a RefWorks, EndNote and BibTex.



személyi adatlap a Scopus oldalon

A Scopus rendszerét szerzői adatbázisként is lehet használni, ahol a cikkek a Google Scholarhoz hasonlóan automatikusan a szerzőhöz rendelődnek hivatkozásaikkal együtt. Előfordulhat téves vagy hiányzó közleményhozzárendelés, illetve az is, hogy ugyanazon szerzőt több szerzőként tartja nyilván a rendszer - ilyen esetekben korrekciót lehet kérni.

Az innen exportált cikkek és hivatkozások adatai az MTMT-be importálhatók. Szinkronizálható az ORCID adatbázissal. Automatikus értesítést lehet kérni a szerzői cikkek új idézeteiről, a cikkek adatbázisba kerüléséről.

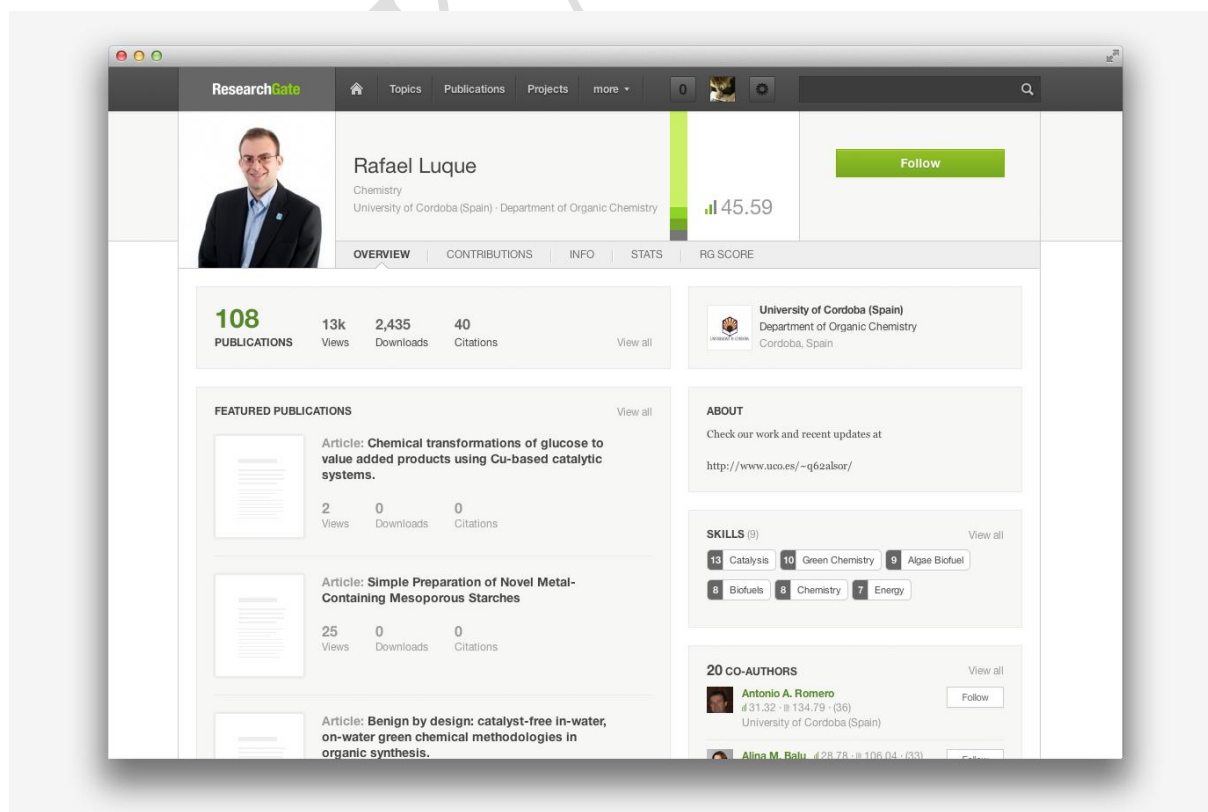
2.11.1.8 ResearchGate

A ResearchGate egy díjmentes közösség háló alapú internetes oldal, amely a különféle tudományágzatokban tevékenykedő kutatók számára lett kifejlesztve. Az oldal szabad hozzáférést kínál a különböző szakfolyóiratok cikkeihez, melyek egy több mint 30 millió bejegyzést tartalmazó, és folyamatosan bővülő adatbázisban találhatóak.



a ResearchGate oldala

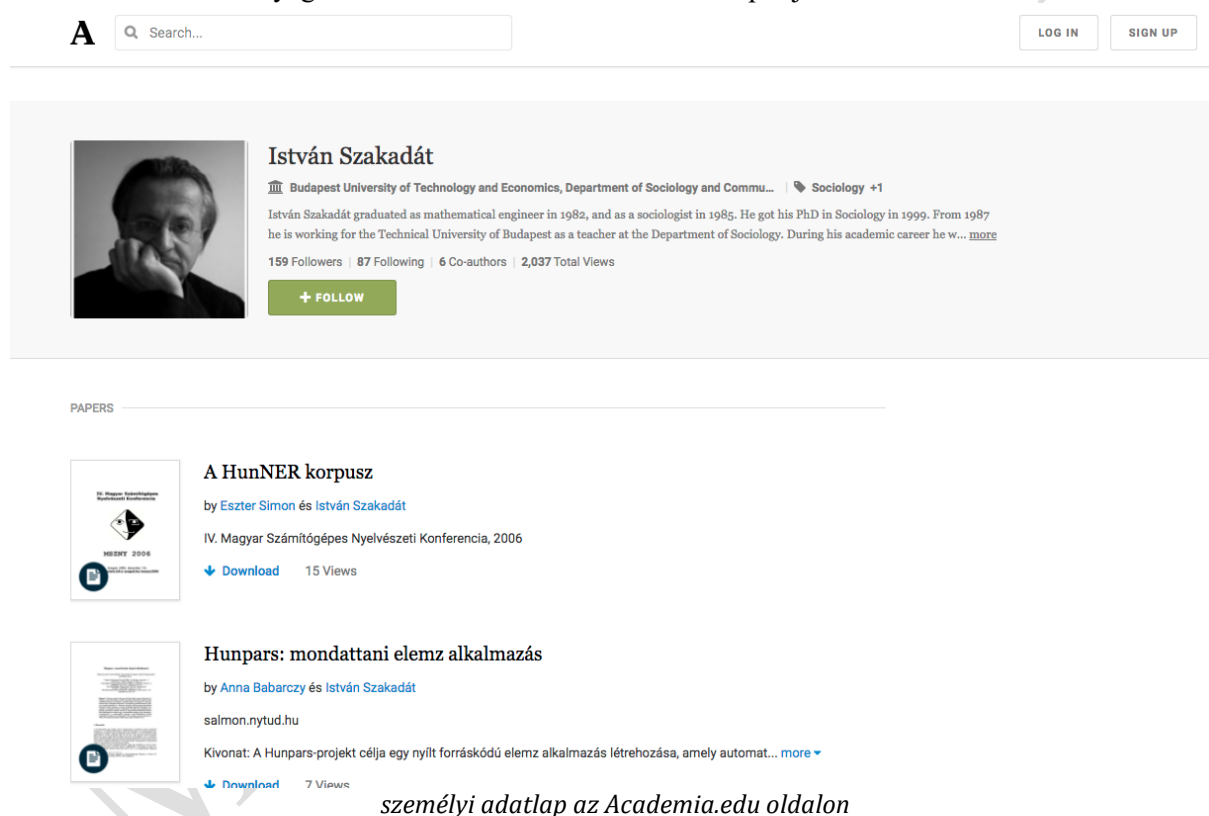
A keresés kezdetekor a felhasználó keresőbe beírt, például „orvosi folyóiratcikkek” kulcsszó megadása után, máris megtekintheti a hasonló témájú, rendelkezésre álló releváns, időrendi sorrendben megjelenő tanulmányokat. Ugyanezzel a funkcióval kereshetők azok a kollégák a világ minden tájáról, akik az adott tudományágban dolgoznak. A ResearchGate 2009 óta a tagoknak nyílt hozzáférést biztosít a szabadabb publikáláshoz. A feltöltött dokumentumok bekerülnek a belső források közé, melyeket a tagok a szabadon letölthetnek, megtekinthetnek. Szolgáltatásai közé tartozik továbbá a fájlmegosztás, fórumok, módszertani beszélgetések, vitacsoportok stb. Ezen funkciók használatának kiinduló feltétele egy személyes kutatóprofil létrehozása. A platform tartalmaz egy úgynevezett állásbörzét, mely a világszerte kínálgató tudományos munkakörű állásokat gyűjti össze. A munkakeresés során megadhatók választási szempontok, mint kulcsszó, pozíció, munkaterület, ország stb.



Sok tudományos szervezet, mint az International Academy of Life Sciences (IALS), az European Science Foundation, a ResearchGate-et használják közös találkozási pontként, ahol a tagok és résztvevők megoszthatják az adatokat, lefolytathatják a szakmai vitákat, megbeszéléseket. A platform ezért kifejlesztett egy úgynevezett subcommunities-t is, melyen belül csak az adott intézmény tagjai kommunikálhatnak. Számos nemzetközi egyetem professzorai garantálják az Advisory Board (tanácsadó szolgáltatás) minőségét és függetlenségét. A ResearchGate, melynek központja Bostonban és Berlinben található, 2011. májusi megalapítása óta a világ 192 országában van jelen több mint 1400000 taggal.

2.11.1.9 Academia.edu

Academia.edu egy amerikai nonprofit közösségi háló tudósok számára, amely prémiumszolgáltatásokat is nyújt fizetés ellenében. A platformon lehetőség nyílik a publikációk megosztására, az impact faktor monitorozására, és az adott tudományterületen zajló kutatások nyomon követésére. A honlap 2008 szeptemberében indult, és 2017 decemberére 36 milliós egyedi látogatót ért el, és addigra a felhasználók több mint 20 millió anyagot töltöttek fel. Az Academia.edu alapítója Richard Price.



A Search... LOG IN SIGN UP

István Szakadát
Budapest University of Technology and Economics, Department of Sociology and Commu... | Sociology +1
István Szakadát graduated as mathematical engineer in 1982, and as a sociologist in 1985. He got his PhD in Sociology in 1999. From 1987 he is working for the Technical University of Budapest as a teacher at the Department of Sociology. During his academic career he w... [more](#)
159 Followers | 87 Following | 6 Co-authors | 2,037 Total Views
[+ FOLLOW](#)

PAPERS

A HunNER korpusz
by Eszter Simon és István Szakadát
IV. Magyar Számítógépes Nyelvészeti Konferencia, 2006
[Download](#) 15 Views

Hunpars: mondattani elemz alkalmazás
by Anna Babarczy és István Szakadát
salmon.nyud.hu
Kivonat: A Hunpars-projekt célja egy nyílt forráskódú elemz alkalmazás létrehozása, amely automat... [more](#)
[Download](#) 7 Views

személyi adatlap az Academia.edu oldalon

A honlapon a felhasználók létrehozhatják a saját személyes oldalukat, feltölthetik a publikációikat, megadhatják saját kutatási, illetve érdeklődési területeiket, és böngészhetnek a hasonló érdeklődésű személyek profiljaiban. Az oldalnak 2017 decemberében 58 millió felhasználója volt világszerte.

2.11.2 További azonosító rendszerek személyekre, helyekre

A személyek, testületek, helyek azonosítására természetesen nem csak a kutatók esetében van szükség, más szakmai területeken is alkalmaznak ilyen rendszereket, sőt, a közgyűjtemények világában már jóval előbb belekezdtek az azonosító rendszerek kiépítésébe.

2.11.2.1 VIAF

A VIAF (Virtual International Authority File, Nemzetközi Virtuális Besorolási Állomány) egy, az OCLC (Online Computer Library Center, Online Számítógépes Könyvtári Központ) által üzemeltetett, nyilvános katalógus rendszer. Célja a különböző nyelvterületek eltérő katalogizálási szabályai alapján kialakított névanyagok egész világon elérhető virtuális katalógusba rendezése.

A VIAF az országokként, nyelvekért, hagyományokként eltérő névvariánsokat, és személyekhez tartozó adattartalmakat kapcsolja össze.

szerzői oldal a VIAF katalógusában

A VIAF a személyeken túl további entitásokat (pl. testületeket, földrajzi helyeket, műveket, kiadványokat) is felvesz az azonosítottak közé, és az azonosítókat nem különíti el az entitástípusok szerint. A VIAF előnye az, hogy nemzetközi azonosító rendszer, és komoly nemzeti közgyűjtemények csatlakoztak hozzá, amelyek folyamatosan küldik be saját authority-rekordjaikat. A VIAF weboldalán lehetséges a teljes állományban is keresni, de mód van arra is, hogy a csatlakozott intézmények, illetve az entitástípusok szerint lehessen szűrni.

Az IFLA-n belül már 1978-ban felvetődött egy közös, nemzetközi besorolási állomány létrehozásának az igénye, a megvalósításra azonban még 2003-ig várni kellett. Ekkor indította el a VIAF-ot a Német Nemzeti Könyvtár (Die Deutsche Bibliothek), a Kongresszusi Könyvtár (Library of Congress) és az OCLC (Online Computer Library Center). A VIAF-ra indulásakor a katalogizálás segédeszközeként tekintettek, ám az elindulása után más funkcióra is használni kezdték. Az OCLC felmérése szerint a Linked Data alapú adatsere-folyamatokban a VIAF a leggyakrabban használt adatforrás. Az új szerepből következnek az új elvárások, amelyek értelmében a VIAF-nak – LD-szolgáltatóként – hiteles és megbízható adatokat kell tartalmaznia.

A VIAF-azonosító egy szám, és a különböző entitásokhoz rendelt számokat sorban osztják ki, nem különítik el egymástól őket. Az alábbi VIAF-szám Budapest azonosítója.

VIAF: 154759119

A VIAF – a partnerintézmények által beküldött adatoknak köszönhetően – folyamatos frissül, azonban ez egyfajta függést is jelent, ráadásul a könyvtári forrásadatbázisokkal való folyamatos kapcsolat

veszélyezteti a teljes adatállomány stabilitását, perzisztenciáját, ami felveti ellenőrző mechanizmusok beiktatásának szükségességét.

Az Országos Széchényi Könyvtár a MEK-ben 2013. július 16-tól folyamatosan felveszi az egyes művek szerzőinek a VIAF azonosítóját.

| CÉDULA | CÍMKÉS | HUNMARC | MARC21 | XML |
|--------|--------|---------|--------|-----|
|--------|--------|---------|--------|-----|

CÉDULA FORMÁTUM

MEK-13611

Parlamenti képviselői helyek megoszlásának becslése közvéleménykutatási adatok alapján / Mészáros József, Szakadát István

tanulmány(ok) ; magyar

(Tárki társadalompolitikai tanulmányok ; 6.)

eredeti kiadvány: Parlamenti képviselői helyek megoszlásának becslése közvéleménykutatási adatok alapján / Mészáros József, Szakadát István Budapest : Tárki, 1998
(Tárki társadalompolitikai tanulmányok, ISSN 1418-0839 ; 6.)
ISBN 963 7869 12 3
OSZK: <http://nektar.oszk.hu/hu/manifestation/327913>

Magyar belpolitika, Kormányzat, parlament, Politológia parlamenti választás, választási rendszer, közvélemény-kutatás, társadalomstatisztika, prognosztika, Magyarország 🇹🇷 🇵🇸, 1990-es évek

MEK-be került: 2014-12-17
URL: <http://mek.oszk.hu/13600/13611>

Jogi közlemény: Jogvédett.

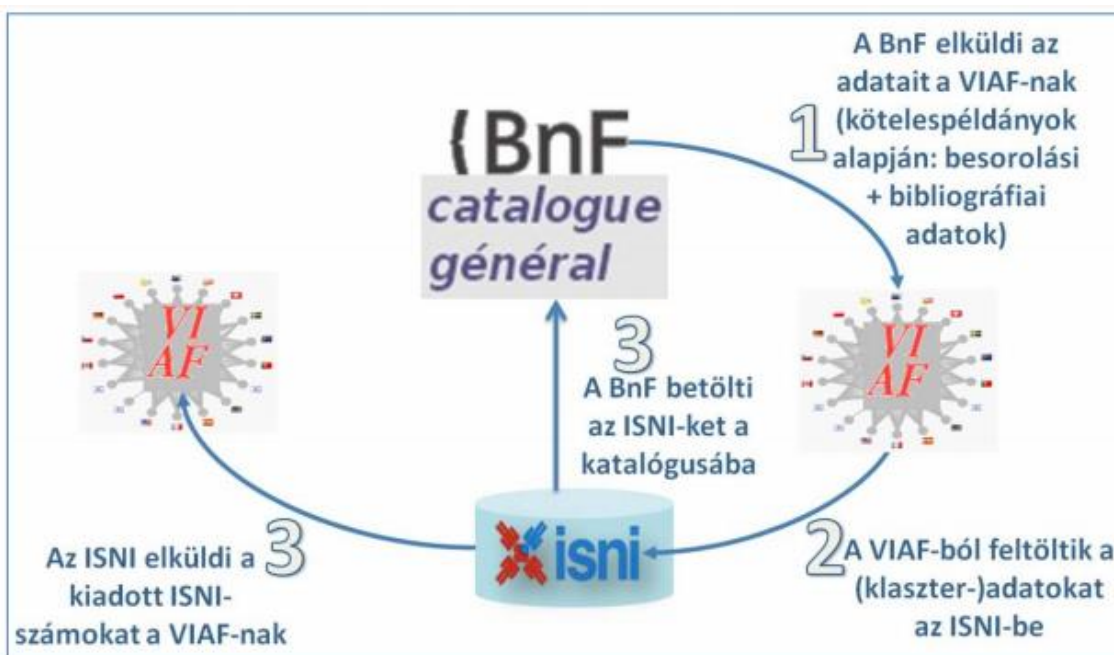
a szerzők neve mögött a VIAF kék azonosító ikonja a MEK katalóguscéduláján

2.11.2.2 ISNI

A könyvtári világban már ismert és elterjedt VIAF-azonosító mellett létezik egy másik nemzetközi szabványos megoldás is, az ISNI-kód (nemzetközi szabványos névazonosító – ISO 27729:2012 Information and documentation – International standard name identifier). Az ISNI célja hatékony eszközt biztosítani az alkotók, a produkciós iparban és a tartalomszolgáltatásban szereplők pontos, egyértelmű azonosításához. Az ISNI nem ad közvetlen hozzáférést az adott személyekről szóló információkhoz, de kapcsolatot teremt a közismert névváltozatok között, egyértelműsíti, hogy kiről van szó, és ezzel lehetővé teszi a kereshetőséget az olyan rendszerekben, amelyekben a személyekről széles körű információk találhatóak.

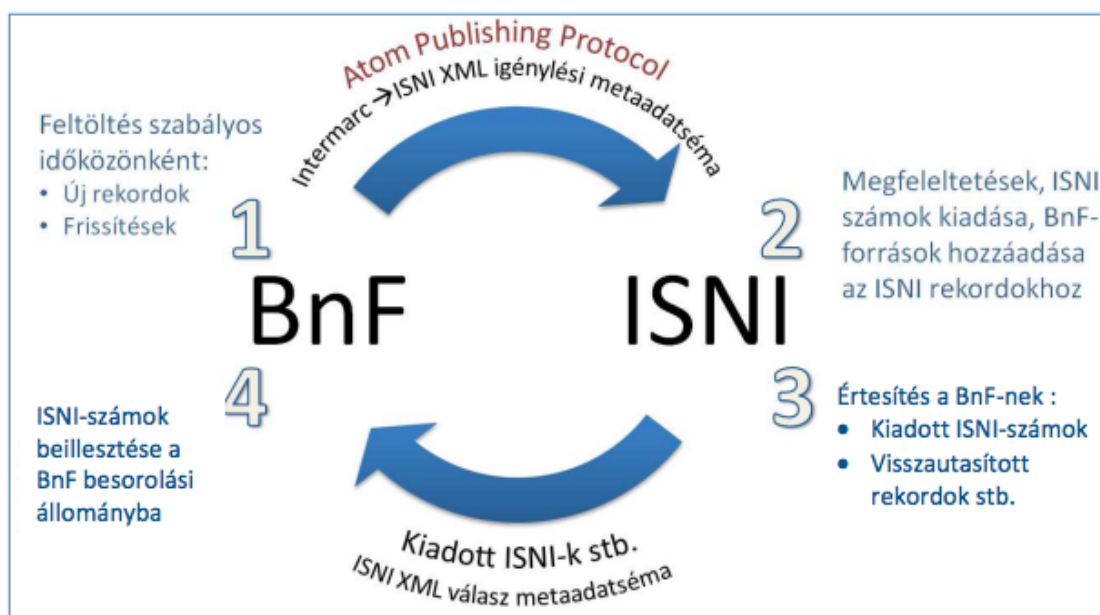
Az ISNI névazonosító kód két különböző problémakörre is megoldást nyújt: egyrészt a különböző információs rendszerek egymástól eltérő azonosítási szisztémát alkalmazhatnak, amelyeket kívánatos lenne összhangba hozni egymással, másrészt pedig a nyilvánosság által ismert nevek, illetve a nevekkal jelölt személyek egyértelműsítését, beazonosítását is segíti.

A VIAF működésével kapcsolatos gondok megoldásaként javasolták sokan a nemzetközi szabványos névazonosító, az ISNI használatát. Utóbbi épít a VIAF-ra, ugyanakkor humán erőforrást (egy ún. „Quality Team”) biztosít az állomány karbantartására, a névazonosságokból fakadó hibák kiküszöbölésére. Az ISNI nemzetközi szabvány, nemzetközi névtérnek tekinthető, ami miatt bevezetése, alkalmazása kérdéseket vet fel közgyűjtemények, könyvtárak munkafolyamatainak újratervezésében. Példaként érdemes áttekintenünk, hogy milyen módon alakított ki kapcsolatot a francia nemzeti könyvtár (BnF) az ISNI-vel annak érdekében, hogy elmozduljon az adatok újrafelhasználását támogató automatizált eljárások irányába, összehangolva az újszerű technológiák alkalmazásából fakadó előnyöket a nemzeti könyvtári feladatokkal és minőségi elvárásokkal (Dancs 2017). A BnF-ben nem a könyvtárosok munkájának teljes kiváltására, sokkal inkább emberek által ellenőrzött automatizmusok, azaz félautomatizmusok implementálására törekednek. Az alábbi ábrán az látható, ahogyan a BnF kapcsolatba kerül a VIAF-fal, ISNI-vel.



BnF-ISNI adatcsere

A könyvtáron kívüli adatfolyamok becsatornázása az ISNI-számigénylés megvalósításával történik. Fontos szerepet játszik ezen a ponton a könyvterjesztésben alkalmazott ONIX-szabvány. A félautomatizált katalogizálás azt jelenti, hogy a bibliográfiai leírásokat készítő munkatársak építhetnek a könyvipari értéklánc más pontjairól (pl. kiadók, könyvterjesztők) érkező rekordokra.

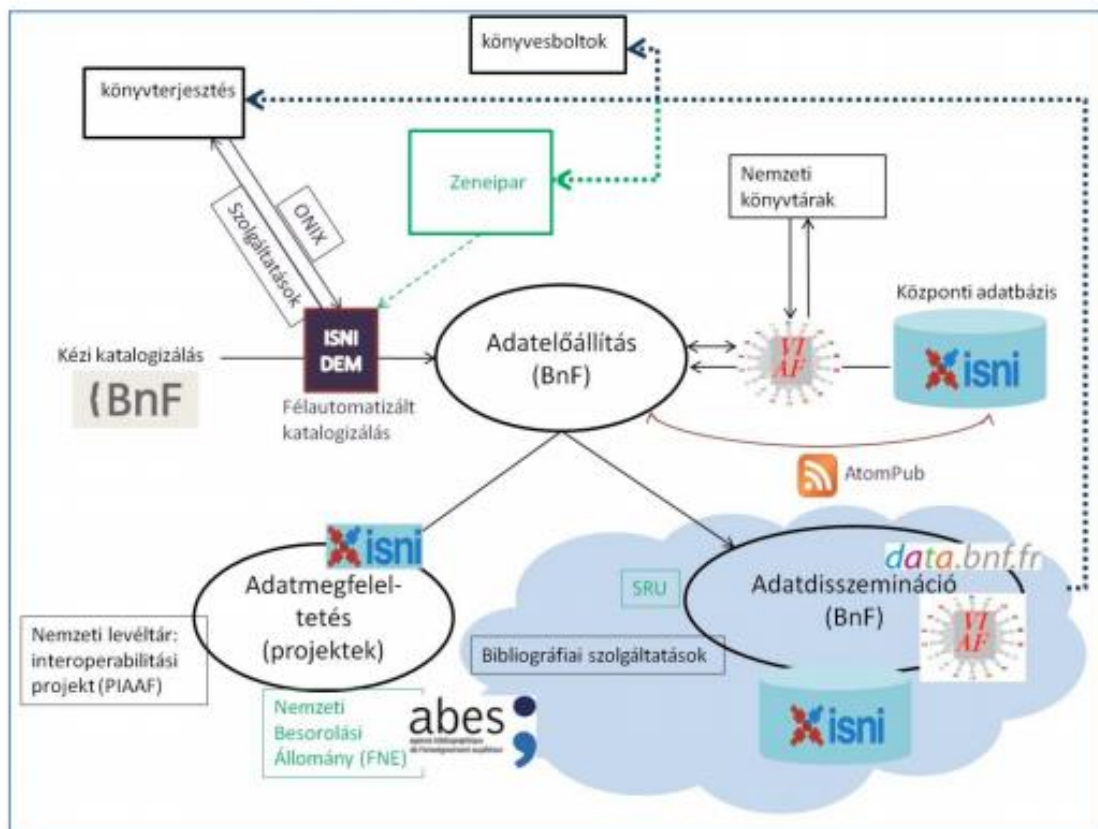


BnF-ISNI közvetlen (automatizált) adatkapcsolat

Aktuálisan a kötelezpéldányként beérkezett mintegy – 65 ezer szerzőtől származó – 75 ezer könyv 90 százalékának a feldolgozása ezen munkafolyamat szerint zajlik. A kiadónak bejelentést kell tennie egy elektronikus űrlap kitöltésével a beküldött kiadványról. A kinyomtatott és a kiadvánnyal együtt kézbesített űrlap segíti a dokumentumazonosítást. A piac 10 százalékát lefedő két kiadói csoport viszont már közvetlenül küldi be a kiadványok adatait ONIX-formátumban az ún. „kötelezpéldány extranet” applikáción keresztül. Az ONIX-rekordokból szoftveres úton kinyert közreműködői adatokat a BnF munkatársai besorolási rekordokká alakítják, majd az AtomPublishing Protocol (AtomPub)

segítségével feltöltik az ISNI adatbázisába. A kiutalt új névazonosítók 8–10 napon belül kerülnek vissza a kiadókhöz.

Az ISNI-igénylések előállítására tehát a kiadóktól származó ONIX-adatok alapján történik egy új fejlesztésű szoftver, az ISNIDEM segítségével. Az alkalmazás mindenképp ellenőrzi, hogy létezik-e már a szerző a katalógusban. Ha a válasz igen, az igénylést összekapcsolják a vonatkozó besorolási rekorddal (egyelőre manuálisan, de a tervek szerint a jövőben automatizált módon). Ha a szerző még nem szerepel a BnF katalógusában, létrehoznak egy új besorolási rekordot az ISNIDEM rendszerben. Egyelőre ez is manuálisan történik, de ugyancsak tervezik automatizálni a folyamatot. Az igényléseket elküldik az ISNI központi adatbázisába, a kiadókat pedig rögtön értesítik, ahogy az azonosítókat megkapták. A folyamat, amely a kiadóval folytatott párbeszédévé teljesedik ki, legfeljebb egy hetet vesz igénybe.



adatfolyamok az optimalizálás után

2016 végén a BnF katalógusa 1,7 millió személynév és 380 ezer testületnév-besorolási rekordot tartalmazott. Összesen 1,4 millió ISNI-t vezettek vissza a rekordokba, ezek eloszlása a személyi nevek és testületi nevek rekordjai között 80:20%. A cél a 100 százalékos lefedettség elérése a személyneveket illetően.

A BnF-nél kiemelt cél az adatok újrahasonosításának előmozdítása, ami egyben hatékonyan segíti elő az ISNI elterjedését. A katalógusadatok nemcsak bibliográfiai és szemantikus szolgáltatásokban (data.bnf.fr), illetve a VIAF-ban hasznosulnak, hanem különböző kiadói és könyvterjesztői adatbázisok is építenek rájuk (pl. Dilicom).

A VIAF és az ISNI közötti eltérések

Tudvalevő, hogy a VIAF elsődleges felhasználási területe a könyvtárak számára az adataik láthatóságának növelése, míg az ISNI a kiadókkal, könyvterjesztőkkel és jogkezelő szervezetekkel való együttműködés szempontjából jelent elsősorban perspektívát. A két azonosító közöttük lévő átfedések ellenére mind műszaki értelemben, mind a finanszírozást illetően különbözik. Az eltéréseket az alábbi táblázat foglalja össze.

VIAF

ISNI

| | | |
|---------------------------------|---|---|
| Fenntartás/ Működtetés | az OCLC szolgáltatása, amelynek fejlesztését, működését egy a könyvtárak képviselőiből álló tanácsadó testület (VIAF Council) koordinálja | nemzetközi szabvány, fenntartása a Nemzetközi ISNI-Ügynökség hatáskörébe tartozik, technikai háttéréről az OCLC gondoskodik |
| Finanszírozás/ Üzleti modell | ingyenes szolgáltatás | tagsági díj a nemzeti ügynökségek részéről, valamint az egyes azonosítók kiosztása után járó díjak |
| Lefedettségi | személyek, testületek, művek, kifejezési formák és földrajzi nevek | személyek és testületek |
| Adattartalom | csak nyilvános adatok | személyes (titkosított) adatok is |
| Az azonosító szerepe | (klaszterezés) a nagyobb nemzeti könyvtárak, tudományos könyvtárak és egyéb közgyűjtemények besorolási rekordállományára építve, az adatok közzététele és terjesztése | szabványos nemzetközi (ISO) azonosítók létrehozása építve a VIAF-ból származó releváns adatokra; adatkapcsolatokat generál más (nem-VIAF) forrásokkal – platformként szolgál a cross-domain adatkapcsolódásokhoz; kizárólag a kiutalt ISNI-azonosítókat teszi közzé és terjeszti (tehát azokat, amelyek esetében az egyes rekordokról bizonyossággal elmondható, hogy egymástól különböznek, illetve hiteles információk alapján deduplikálták őket) |
| Technikai megoldások | aratja az adatokat, csoportosítási (klaszterezési) eljárást alkalmaz, a rekordok megjelenítése és letöltése többféle formátumban lehetséges | köteget feltöltést alkalmaz, illetve csoportosítási eljárást, végzi a rekordcsoportok (klaszterek) karbantartását; az online számigénylés API-n keresztül megvalósítható; lehetséges végfelhasználói közreműködés (crowdsourcing) (kontrollált módon), valamint az online szolgáltatások testre szabása a tagok és nyilvántartó ügynökségek számára; szakemberekből álló csapat (Quality Team) biztosítja az adatbázis minőségének ellenőrzését és a hibák javítását; a rekordok megjelenítése többféle formátumban (XML, RDF) lehetséges |
| Adatforrás | könyvtárak besorolási állománya, a kulturális örökség terület egyéb adatszolgáltatói (múzeumok, levéltárak) | különböző adatszolgáltatók (könyvtárak, jogkezelő szervezetek, kutatói adatbázisok, a zeneipar és könyvpiar szereplői); az ISNI-ügynökségek által feltöltött vagy az online számkiutalás során szolgáltatott adatok |
| Adatmodell | a VIAF semlegesnek tekinthető az egyes szabványok és formátumok tekintetében | az ISNI ISO szabványként egyedi adatmodell kialakítását igényli, ugyanakkor adatkezelési politikája alkalmassá teszi az eltérő forrásokból származó adatok befogadására |

| | | |
|-----------------------------|--|--|
| Adatperzisztencia | a klaszterek tartalma a feltöltésekkel gyakran változik | a a cél a lehetőségekhez mérten stabil, azaz perzisztens és megbízható azonosítók létrehozása; a rendszer olyan algoritmussal rendelkezik, amely csak azoknak a klasztereknek utal ki ISNI-számot, amelyek megbízhatósága elér egy meghatározott szintet |
| Frissítések | a VIAF havonta frissül | folyamatos frissülés az ügynökségek igényei alapján, illetve az online számkialások tükrében, beleértve a Minőségcsoport rekordszerkesztési műveleteit, illetve a végfelhasználói igénylések kezelését |
| Adatdisszemináció és licenc | a VIAF nyílt kapcsolt adatként terjeszthető ODC-By licenc10 alatt | az ISNI egy hivatkozható nyilvántartás, amely egy API-n keresztül nyilvánosan lekérdezhető; az adatok nyílt licenc alatt felhasználhatók, jelenleg tervezik nyílt kapcsolt adattá fejlesztését |
| Célközönség | könyvtárak, a kulturális örökség intézményei, oktatás és kutatás | globális és több szektort érint (könyvtárak, levéltárak, kutatás, zene- és filmipar, könyv- és folyóirat-kiadás, jogkezelők stb.) |
| Példa felhasználásra | a a katalogizálás forrása, a könyvtári entitások összekapcsolása repozitóriumokkal (HUB-funkció), mint például a Wikipedia és a Wikidata | linkek automatizált létrehozása a besorolási rekordokhoz a kiadói és könyvtári adatok összekapcsolásával |

A VIAF közgyűjteményi (elsősorban könyvtári) rekordok alapján és közgyűjtemények (elsősorban könyvtárak) által épül, valamint a nevek csoportosítása során nem alkalmaz ellenőrzési mechanizmusokat. Ezzel szemben az ISNI felhasználási köre kiterjedtebb, a teljes kreatív ágazatot magában foglalja, a zeneipart is beleértve, így a benne rejlő adatkapcsolási (linked data) potenciál is jelentősebb. Az ISNI igénylésekor egy szakemberekből álló csapat (ISNI Quality Team) ellenőrzi, hogy az egy azonosítóhoz rendelendő névváltozatok valóban egy személyt takarnak-e, valamint javítják az ISNI referencia-adatbázisban (ISNI Reference Database) előbukkanó hibákat. Az ISNI nyilvántartása tehát megbízhatóbb, hitelesebb adatokra épül. Továbbá az ISNI egy nemzetközi (ISO) szabvány, a VIAF az OCLC tulajdona, jóllehet a kétféle azonosító között van kapcsolat, és az OCLC műszaki partnerként van jelen az ISNI-adatbázis működtetésében.

Az ISNI különös adottsága, ami ugyancsak megkülönbözteti a könyvtári besorolási rekordok nemzetközi szintű azonosítására használt VIAF-tól, hogy különbséget tesz személy és személyiség között, amit a személyes adatok védelmének (privacy) kezelésével valósít meg. Ez közelebbről azt jelenti, hogy lehetőség van arra, hogy egy személy, ha érdeke úgy kívánja, több személyiséget (álnevet) használjon anélkül, hogy a személyiségek, illetve a hozzájuk kapcsolódó nevek közötti kapcsolat nyilvános volna. Az ISNI névterét használó könyvtári rendszerekbe tehát a kapcsolati információ alaphoz nem kerül át. A nyilvánosság ugyanakkor – természetesen – a későbbiekben módosítható. Az ISNI-kód egy négy blokkból álló számsorozat. Lentebb egy magyar politikus ISNI-száma látható.

ISNI: 0000 0001 2141 7284

Az ISNI oldalára ugorva bibliográfiai rekord leírásához hasonló oldalt kapunk, ahol a – különböző forrásokból származó – névváltozatokon túl elérhető még a személyhez kapcsolható más személyek linkjei, a személyhez köthető művek listája és egyéb megjegyzések.

| | | | |
|---|--------------|----------------|---|
| shortlist | title data | search history | 1 |
| results search [or] ISN:0000000121417284 1 hits | | | |
| labels | sources data | marc21 | |

Please help us improve this record

If you have any supplemental information about the identity listed here, please click in this box to go to the contribution form.

Thank you in advance!

ISNI: 0000 0001 2141 7284

Name: Orbán, Viktor
 Orbán Viktor (magyar politikus, a Fidesz elnöke)
 Victor Orbán
 Viktor Orbán
 Viktor Orbán (actual primeiro-ministro da Hungria)
 Viktor Orbán (Hongaars politicus)
 Viktor Orbán (Hungarian politician)
 Viktor Orbán (Hungarian politician, chairman of Fidesz)
 Viktor Orbán (personnalité politique hongroise)
 Viktor Orbán (politico húngaro)
 Viktor Orbán (politico ungherese)
 Viktor Orbán (ungarischer Politiker)
 Viktor Orbán (ungarsk politiker)
 Viktor Orbán (węgierski polityk)
 Viktors Orbāns
 Виктор Орбан
 Виктор Орбан (унгарски политичар)
 Віктор Орбан
 Orbán, Viktor
 Orbán, Viktor
 اوربان، ویکتور
 ویکتور اوربان
 (ویکتور اوربان (سیاستمدار اهل مجارستان))
 विक्रम अदरान
 오르반 빅토르
 オルバーン・ヴィクトル
 奧班 維克多

Dates: 1963-

Creation class: Language material
 Im
 Projected medium
 Text

Creation role: author

☐ Related names

☐ Titles

☐ Notes

az ISNI oldala Orbán Viktorról

2.11.2.3 PIM személynévtér

Magyarországon a Petőfi Irodalmi Múzeumban kezdtek el elsőként építeni személynévteret, amelynek azonosítóit ma már a személyek magyar Wikipédia-oldalaira is beillesztik a többi globális személynév-azonosító mellé. A PIM személynévtér számmal azonosítja a személyeket.

<https://opac-nevter.pim.hu/record/-/record/PIM123214>

.A névtér oldalain elérhető a PIM adatbázisaiból a kiválasztott személyhez kapcsolható adatok teljes köre úgy, a személy életrajzi adatai, személyes bibliográfiai adatai, esetleges kitüntetései stb.

Tarlós István

Database: INT; IKN
Unified name form: Tarlós István
Date of birth/death: 1948-
Profession: gazdasági szakmérnök; mélyépítő mérnök; polgármester
Birthplace: Budapest
Date of birth: 1948. V. 26.
Schools/Career: III. ker. Önkormányzat

- ▼ Award:
- ▼ Data sources:
- ▼ Wiki:
- Works

Link to this record: <http://resolver.pim.hu/auth/PIM123214>

a PIM személynévtér oldala Tarlós Istvánról

2.11.2.4 ULAN

A Getty Intézet 1984-ben kezdett bele a művészek alapadatainak gyűjtésébe, rendszerezésébe, és ennek eredményeként született meg az ULAN, a művésznevek egyesített listája (Union List of Artist Names), amely egy rögzített (korlátozott) szköppal rendelkező személynévtérként is értelmezhető. A szolgáltatás oldalán az olvasható, hogy magukat strukturált szótárként definiálják.⁸ Sokféle adat felvehető a rendszerbe, az alábbi adattípusokat kötelező kitölteni a rendszerben egy új elem felvételekor.

- record type
- name
- name source
- display biography
- nationality
- role
- birth date and death date

Az ULAN gyűjti és egymástól elkülöníti a művészek neveit, de csak a személyeket látja el egyedi azonosítóval, a neveiket már nem.

⁸ „Az ULAN egy strukturált szótár, amely művészek, mecénások, cégek, múzeumok és más kapcsolódó személyek vagy szervezetek neveit tartalmazza a műalkotások, építészeti alkotások és gyűjtemények területén” (<http://www.getty.edu/research/tools/vocabularies/ulan/about.html>).



Click the icon to view the hierarchy.

[Semantic View](#) (JSON, JSONLD, RDF, N3/Turtle, N-Triples)

ID: 500339585

Record Type: Person

Mozart, Wolfgang Amadeus (Austrian composer, 1756-1791)

Names:

- Mozart, Wolfgang Amadeus** ([preferred](#),U,index,LC)
- Wolfgang Amadeus Mozart** (U,display)
- Mo□□□art, Vol'fgang Amadei** (U)
- Mötsaruto** (U)
- Mo□□□art, Iogann-Krizost Vol'fgang Gotlib** (U)
- Mozart, Johann Chrysostom Wolfgang Amadeus** (U)
- Mozart, W. A.** (U)
- Mozart, Wolfgango Amadeo** (U)
- Mo□□□art, V. A.** (U)
- Mocartas, V. A.** (U)
- Motsart, Volphnkank Amedaios** (U)
- Mo□□□art, Volfang Amadeus** (U)
- Mozzart, Apollo** (U)
- Mozart, Joannes Chrysostomus Wolfgangus Amadeus** (U)
- Mozart, Johannes Chrisostomus Wolfgangus Theophilus** (U)
- Mozhate** (U)
- Моцарт, Вольфганг Амадей** (U)
- מוצרט** (U)
- מוצרט, וולפגנג אמדאוס** (U)
- 莫札特** (U)

Nationalities:

Austrian ([preferred](#))

Roles:

composer (music) ([preferred](#))

Gender: male

Birth and Death Places:

Born: [Salzburg](#) (Salzburg, Austria) (inhabited place)
Died: [Vienna](#) (Vienna state, Austria) (inhabited place)

List/Hierarchical Position:

- ... Non-Artists
- [Mozart, Wolfgang Amadeus](#) (I,U)

Biographies:

(Austrian composer, 1756-1791) [VP Preferred]
(artist, 1756-1791) [GRL]

Additional Names:

Sources and Contributors:

- Моцарт, Вольфганг Амадей [GRL]
..... [Library of Congress Authorities database](#) (n.d.) n 80022788
- מוצרט [GRL]
..... [Library of Congress Authorities database](#) (n.d.) n 80022788
- מוצרט, וולפגנג אמדאוס [GRL]
..... [Library of Congress Authorities database](#) (n.d.) n 80022788
- 莫札特 [GRL]
..... [Library of Congress Authorities database](#) (n.d.) n 80022788
- Mocartas, V. A. [GRL]
..... [Library of Congress Authorities database](#) (n.d.) n 80022788
- Mo□□□art, Iogann-Krizost Vol'fgang Gotlib [GRL]
..... [Library of Congress Authorities database](#) (n.d.) n 80022788
- Mo□□□art, V. A. [GRL]
..... [Library of Congress Authorities database](#) (n.d.) n 80022788
- Mo□□□art, Volfang Amadeus [GRL]
..... [Library of Congress Authorities database](#) (n.d.) n

W.A. Mozart adatai a Getty ULAN oldalán

Az ULAN azonosítóit gyakran feltüntetik a Wikipédia művészekről szóló oldalain is, és ebben a funkciójában már – korlátozott skóppal bíró – globális névtérként is tekinthetünk rá.

2.11.2.5 GDN

A Német Nemzeti Könyvtár besorolási adatállománya (Gemeinsame Normdatei) a német nyelvterületen használt rendszer, amely az RDA szabályzat alapján rendel egyedi azonosítókat többféle entitáshoz, így például személyekhez, testületekhez, földrajzi helyekhez, tárgyszavakhoz.

| | |
|---------------------------------|--|
| GND | |
| Link zu diesem Datensatz | http://d-nb.info/gnd/4008684-7 |
| Geografikum | Budapest |
| Andere Namen | Haupt- und Residenzstadt Budapest Budapest / Tanács (Spitzenorgan) Székesfőváros Budapest Boedapest Ofenpest Ofen-Pest Aquincum (röm.) Budimpešta Budapesta Budapešta Budapeszt Budapešt Budapestum Buda Pest Ofen Pest-Buda |
| Quelle | B 2006 M Druckorte des 16. bis 19. Jhs., S. 36 Homepage (Stand: 10.01.2017): http://budapest.hu/ |
| Erläuterungen | Definition: Hauptstadt Ungarns, 1.1.1873 durch Zusammenschluß von Buda, Obuda u. Pest gegründet |
| Zeit | 1873- |
| Land | Ungarn (XA-HU) |
| Koordinaten | E 019°02'23" / N 047°29'52" E019.039909 / N047.498010 |
| Vorgänger | Buda Pest (Stadt) |

Budapest GDN-rekordjának részlete

Ezeket az adatokat szabadon hozzáférhetővé téve más szolgáltatások számára a GDN-azonosítók – valamely entitástípusra szűkítve a fókuszot – működhetnek globális névtérként is. A Wikipédia-oldalakon gyakran beágyazzák a tárgyalt dolog GDN-ID-jét, ami önmagában mutatja, megerősíti azt a tényt, hogy a GDN valóban globális névtérként funkcionál.

GDN: 4008684-7

2.11.2.6 BNF

A Francia Nemzeti Könyvtár (Bibliothèque nationale de France, BnF) névazonosító rendszere ugyanúgy és ugyanazért szolgálhat globális névtérként, mint a német nemzeti könyvtár (GDN) besorolási állománya, a GDN. A Wikipédia a BNF-azonosítókat is gyakran elhelyezi a megfelelő oldalakon.

Budapest (Hongrie)

Vedette matière nom géographique. S'emploie en tête de vedette ou en subdivision.

La capitale hongroise Budapest est la réunion en 1872 de trois communes : Buda (Ofen en allemand), Óbuda (Aquincum=Altofén, la ville romaine) et Pest (Contra Aquincum ou Transaquincum)

<Employé pour :

Altofén (Hongrie)
Aquincum (ville ancienne)
Buda (Hongrie)
Bude (Hongrie)
Contra Aquincum (ville ancienne)
Óbuda (Hongrie)
Ofen (Hongrie)
Ofenpesth (Hongrie)
Pest (Hongrie)
Pesth (Hongrie)
Transaquincum (ville ancienne)

<<Terme(s) générique(s) :

[Hongrie -- Divisions politiques et administratives](#)

Source(s) :

Pesth und Ofen mit ihren Einwohnern, besonders in medicinischer und anthropologischer Hinsicht, 1838 . - L'Hunyadi Janos, source purgative à Bude (Ofen en Hongrie), 1871 . - Az aquincumi amfiteátrumok, 1956
Révai új lexikona, 1998 . - Magyar helységnév-azonosító szótár, 1992 . - GDEL

Domaine(s) : 914

Correspondance(s) exacte(s) :

- LCA (Library of Congress Authorities) : Budapest (Hungary) <http://id.loc.gov/authorities/names/n79091691>

Notice n° : FRBNF11933405

Création : 81/02/12 Mise à jour : 13/03/01

Budapest BNF-rekordja

A BNF -azonosító egy karaktersorozat, amelyben lehetnek betűk is a számok mellett.

BNF: cb119334051


2.11.2.7 A könyvtárak besorolási állománya

Amikor személynevet, testületi vagy földrajzi nevet kell felvenni a bibliográfiai rekordokba, a könyvtárak a neveket írják be, nem pedig névazonosítókat a megfelelő mezőkbe. Ez a megoldás nem teszi lehetővé a nevek egyértelmű kezelését. Lényegében ez az oka annak, hogy kidolgozták és használatba vették a besorolási állományok (authority records) rendszerét. A besorolási állományok bevezetésének és fenntartásának célja egyfelől a nevek egyértelmű kezelése, a szinonimák és homonimák feloldása, tehát az egyértelműsítés, másfelől a kitüntetett névalak kiválasztása abban az esetben, ha egy névhordozóhoz (személy, testület, földrajzi hely) több név is tartozik. A kitüntetett névalakra sokáig azért volt szükség, mert csak ezáltal lehetett egyértelmű (egy az egyhez) kapcsolatot kialakítani a nevek és a névhordozók között. Ezt a – jogos – igényt a digitális világban már másként lehet kielégíteni, így a kitüntetett névalak jelentősége csökken. Természetesen addig, ameddig a különböző szakmai (például könyvtárosi) közösségek igényt tartanak a preferált névalak minősítés kiosztására, addig ezt a lehetőséget fent kell tartani, de a nemzeti névterek építése és használata során, amikor többféle szakma képviselőinek kell egymással együttműködve dolgoznia, figyelembe kell venni azt is, hogy a preferált névalak besorolása szakmánként, kontextusonként változhat.


Adding authority Personal Name


Save

0 1 2 3 4 5 6 7 8


000 - LEADER 

▲ 00 fixed length control field *


001 - CONTROL NUMBER 

003 - CONTROL NUMBER IDENTIFIER 


▲ 00 control field *


005 - DATE AND TIME OF LATEST TRANSACTION 

▲ 00 control field *

008 - FIXED-LENGTH DATA ELEMENTS 

▲ 00 fixed length control field *

010 - LIBRARY OF CONGRESS CONTROL NUMBER 

014 - LINK TO BIBLIOGRAPHIC RECORD FOR SERIAL OR MULTIPART ITEM 

részlet egy besorolási rekord adatbeviteli űrlapjáról

A könyvtári besorolási állomány névtérnek tekinthető, hiszen ugyanazokat a funkciókat valósítja meg, mint amiket a névtértől elvárunk, még akkor is, ha gyakran más eszközöket, más megoldásokat használ a célok eléréséhez.

2.11.2.8 ISADN

Az 1989-ben elindult projektben az ISADN (International Standard Authority Data Number) céljaként a személyek egyértelmű azonosítását lehetővé tevő –számszerű – azonosító rendszer felállítását és működtetését tűzték ki. A projekt végül nem valósult meg, de mintául szolgált a később sikeressé váló globális személyazonosító rendszerek, mint a VIAF vagy az ISNI számára (Habibzadeh & Yadollahie 2009).


2.11.2.9 Wikidata

A Wikipédia elvileg felkínálhatná az egyértelmű azonosítás lehetőségét kiválasztott entitástípus (mondjuk a személyek) számára, hiszen a webcímekek egyedisége garantálni tudná a személyek egyedi azonosítását. A Wikipédia gyakorlati működés módja azonban e lehetőség kihasználását nem teszi lehetővé. A wiki-bejegyzések címének szabad változtatgatása ugyanis teljesen ellentétes a névterekkel szemben támasztott elvárásokkal, amelyek szerint az azonosítónak (a címnek) minden pillanatban, az időben folytonosan egyedinek és perzisztensnek kell lennie.

A Wikipédia tartalmait „adatszerűsíteni” kívánó Wikidata projekt már alkalmasabbnak tűnik az egyedi azonosítói szerep betöltésére, hiszen a kiválasztott entitás (típus, individuum) számára perzisztens azonosítót adnak a projekt keretein belül, de egyelőre kérdéses, hogy az alkalmazott, választott informatikai megoldás alkalmas^{sá} teszi-e a Wikidata egész rendszerét arra, hogy névtérként is működjön.

A Wikidata célja elsősorban az, hogy RDF-tripletek formájában – géppel is értelmezhető – állításokat lehessen tenni adott entitásról, individuumról, de lehet azonosítórendszerként is értelmezni, hiszen minden wikidata-oldalon leírt dolognak egyedi azonosítója van.

English Not logged in Talk Contributions Create account Log in



WIKIDATA

Item [Discussion](#) [Read](#) [View history](#)

Budapest (Q1781)

capital city of Hungary [edit](#)
 Buda Pest | Buda-Pest

[In more languages](#) Configure

| Language | Label | Description | Also known as |
|-----------|----------|-------------------------|------------------------------|
| English | Budapest | capital city of Hungary | Buda Pest Buda-Pest |
| Hungarian | Budapest | Magyarország fővárosa | Bp. Buda Pest Óbuda |
| French | Budapest | capitale de la Hongrie | |
| German | Budapest | Hauptstadt von Ungarn | Ofen Alt-Ofen |

[All entered languages](#)

Budapest Wikidata-bejegyzésének része (binnen a Wikidata-azonosítóval: Q1781)

A Wikidata oldalaira RDF-állításokat lehet feltenni, amelyeket aztán – egy másik felületen – SparQL-lekérdezések segítségével értelmezni, hasznosítani lehet további célokra. Az állítások szintaxisára, szemantikájára vannak belső szakmai útmutatók, de a rendszer nem ellenőrzi ezek betartását, ami bizonyos mértékig kétségeket ébreszt a rendszer konzisztenciáját illetően.

Statements

| | | |
|------------------------|-------------------|---------------------------------|
| instance of | city | edit |
| | ▼ 0 references | + add reference |
| | capital | edit |
| | ▼ 0 references | + add reference |
| town status in Hungary | ▶ 1 reference | edit |
| | county of Hungary | edit |
| | ▼ 0 references | + add reference |
| | | + add value |
| part of | Central Hungary | edit |
| | ▶ 1 reference | + add value |

Budapest – RDF-állításokat tartalmazó – Wikidata-oldalának része

2.12 Névstruktúra

A névtér fogalmának alaposabb elemzésére – sok egyéb további szemponton túl – azért van szükség, hogy

- láthassuk azt, hogy a névterek miben térnek el más rendszerektől;
- meg tudjuk mondani, mi a névtér, és mi nem az, és mi alapján tudjuk ezt eldönteni;
- minél pontosabban el tudjuk különíteni egymástól a különböző névtereket, névtértípusokat, tudjuk, hogy mi az azonos, és mi az eltérő bennük;
- tisztázhatjuk, hogy mi kapcsolat a köznévtér, a tulajdonnévtér, valamint a tudásszervezési rendszer fogalmi között.

Ezen kérdések megválaszolásához első lépésben arra van szükségünk, hogy röviden felvázoljunk egy relációelméletet, majd az itt definiált relációk segítségével pontosan meghatározzuk a névterekkel, tudásszervezési rendszerek, ezek típusaival kapcsolatos fogalmakat.

2.13 Relációelmélet

A névterek, a tudásszervezési rendszerek (KOS-ok) matematikai értelemben olyan struktúráknak tekinthetők, amelyek a névtérbe sorolt elemek közt adott relációkat definiálnak. A névterek egyfelől a névhalmaz elemeinek típusában, másfelől az elemek közti relációk típusaiban térnek el egymástól. A köznévterek vizsgálatakor, jellemzésekor ezért fontos lehet, hogy milyen relációkról beszélhetünk. A tudásszervezési rendszerek (főleg az osztályozási rendszerek és teauruszok) építésének több évszázados történetében nem túl nagy mennyiségű relációtípust kezeltek, ami nyilván nem véletlen, hiszen az emberi tudás terminológiai, fogalmi rendszerbe foglalása nem elsősorban a feldolgozni, leírni kívánt tudásterület sajátosságaitól, hanem az emberi tudástermelés, az emberi gondolkodás általános szabályszerűségeitől függ.

A következőkben egy olyan rövid és formális relációelméletet vázolunk fel, amelyben a névterek építkezésekor szükséges relációk mind benne lesznek, és amely – szükség esetén – szabadon bővíthető. A fogalmak formális meghatározása során használjuk a logikában kiterjedt módon alkalmazott szimbólumokat a logikai alapfogalmakra, műveletekre.

2.13.1 Relációelméleti alapfogalmak

Az itt bemutatott relációelméletet a legalapvetőbb relációs fogalmak (reláció, relációtulajdonságok, elemi relációk, függvény, művelet, nevezetes matematikai relációk, relációművelet) definiálásával kezdjük, majd összegyűjtjük az ontológiai alapfogalmakat. Ezek után és ezek alapján tudjuk meghatározni azokat a KOS-relációkat, amelyeket az egyes tudásszervezési rendszerek, névterek jellemzéséhez, leírásához lehet felhasználni.

2.13.1.1 A reláció fogalma

Mind a reláció fogalmát, mind a hozzá kapcsolódó két fontos fogalom, az értelmezési tartomány és értékkészlet fogalmát halmazelméleti alapon határozzák meg. Ezeket a következő módon definiálhatjuk.

reláció

| | |
|--|-----------------------------|
| $\text{relation}(x, y) \equiv R(x, y)$ | $\{\{x\}, \{x, y\}\}$ |
| $\text{domain}(R)$ | $\{y: \exists x(R(y, x))\}$ |
| $\text{image}(R)$ | $\{y: \exists x(R(x, y))\}$ |

A relációfogalomra támaszkodva meghatározhatjuk a legelemibb relációtípusokat is. Ezek az univerzális vagy teljes reláció, az azonosság, identitás vagy diagonális reláció, az üres vagy null reláció, valamint a különbözőségi, egyenlőtlenségi, antidiagonális relációk.

elemi relációk

| | |
|------------------------------------|---|
| $\text{universal_relation}(x, y)$ | $\forall m \forall n (R(x, y))$ |
| $\text{identity_relation}(x, y)$ | $\forall m \forall n (R(x, y) \leftrightarrow x = y)$ |

| | |
|-------------------------|--|
| empty_realtion(x,y) | $\forall m \forall n (\neg R(x, y))$ |
| diversity_relation(x,y) | $\forall m \forall n (R(x, y) \leftrightarrow x \neq y)$ |

2.13.1.2 Relációtulajdonságok

A relációk szemantikájának kibontása előtt először azt kell felidézni, hogy a különböző relációfogalmak milyen formális, szintaktikai tulajdonságokkal rendelkezik(nek). A névterek számára hasznos relációk jellemzéséhez a matematikában használt relációtulajdonságra tudunk támaszkodni.

relációtulajdonságok

| | |
|---------------------------------|---|
| reflexív | $\forall x (R(x, x))$ |
| irreflexív | $\forall x (\neg R(x, x))$ |
| nem-reflexív | $\exists x (\neg R(x, x))$ |
| balról totális | $\forall x \exists y (R(x, y))$ |
| jobbról totális | $\forall y \exists x (R(x, y))$ |
| másodlagosan reflexív | $\forall x \forall y (R(x, y) \rightarrow R(y, y))$ |
| fordított másodlagosan reflexív | $\forall x \forall y (R(x, y) \rightarrow R(x, x))$ |
| koreflexív | $\forall x \forall y (R(x, y) \rightarrow x = y)$ |
| szimmetrikus | $\forall x \forall y (R(x, y) \rightarrow R(y, x))$ |
| aszimmetrikus | $\forall x \forall y (R(x, y) \rightarrow \neg R(y, x))$ |
| antiszimmetrikus | $\forall x \forall y ((R(x, y) \wedge R(y, x)) \rightarrow x = y)$ |
| nem-szimmetrikus | $\exists x \exists y (R(x, y) \wedge \neg R(y, x))$ |
| összefüggő | $\forall x \forall y (x \neq y \rightarrow (R(x, y) \vee R(y, x)))$ |
| erősen összefüggő | $\forall x \forall y (R(x, y) \vee R(y, x))$ |
| gyengén összefüggő | $\forall x \forall y \forall z (R(x, y) \wedge R(x, z) \rightarrow (y = z \vee R(y, z) \vee R(z, y)))$ |
| dichotóm | $\forall x \forall y (x \neq y \rightarrow (R(x, y) \wedge \neg R(y, x)) \vee (\neg R(x, y) \wedge R(y, x)))$ |
| trichotóm | $\forall x \forall y (R(x, y) \wedge \neg R(y, x) \wedge x \neq y) \vee (\neg R(x, y) \wedge R(y, x) \wedge x \neq y) \vee x = y$ |
| balról egyértelmű | $\forall x \forall y \forall z ((R(x, z) \wedge R(y, z)) \rightarrow x = y)$ |
| jobbról egyértelmű | $\forall x \forall y \forall z ((R(x, y) \wedge R(x, z)) \rightarrow y = z)$ |
| tranzitív | $\forall x \forall y \forall z ((R(x, y) \wedge R(y, z)) \rightarrow R(x, z))$ |
| intranszítív | $\forall x \forall y \forall z ((R(x, y) \wedge R(y, z)) \rightarrow \neg R(x, z))$ |
| kvázitranszítív | $\forall x \forall y \forall z ((R(x, y) \wedge R(y, z) \wedge x \neq z) \rightarrow R(x, z))$ |
| aliostranzitív | $\forall x \forall y \forall z ((R(x, y) \wedge R(y, z)) \rightarrow (R(x, z) \vee x = z))$ |
| nem-tranzitív | $\exists x \exists y \exists z ((R(x, y) \wedge R(y, z)) \wedge \neg R(x, z))$ |
| cirkuláris | $\forall x \forall y \forall z ((R(x, y) \wedge R(y, z)) \rightarrow R(z, x))$ |
| acirkuláris | $\forall x \forall y \forall z ((R(x, y) \wedge R(y, z)) \rightarrow \neg R(z, x))$ |

| | |
|--------------|--|
| euklideszi | $\forall x \forall y \forall z ((R(x, y) \wedge R(x, z)) \rightarrow R(y, z))$ |
| aneuklideszi | $\forall x \forall y \forall z ((R(x, y) \wedge R(x, z)) \rightarrow (\neg R(y, z) \wedge \neg R(z, y)))$ |
| katénáris | $\forall x \forall y \exists z (R(x, y) \rightarrow ((R(y, z) \vee R(z, x)))$ |
| akatenáris | $\forall x \forall y \forall z (R(x, y) \rightarrow (\neg R(y, z) \wedge \neg R(z, x)))$ |
| konfluens | $\forall x \forall y \forall w \exists z ((R(x, y) \wedge R(x, w)) \rightarrow (R(y, z) \wedge R(w, z)))$ |
| akonfluens | $\forall x \forall y \forall w \exists z ((R(x, y) \wedge R(x, w)) \rightarrow \neg (R(y, z) \wedge R(w, z)))$ |

2.13.1.3 Függvény

A reláció kategóriája segítségével lehet definiálni a függvény vagy leképezés fogalmát, majd erre támaszkodva lehet meghatározni további kategóriákat, mint az inverz függvény vagy inverz leképezés, az egy-egyértelmű, a bitotális függvény vagy 1:1-es reláció, az injektív vagy egyrétű függvény, a szürjektív függvény vagy ráképzés vagy szuperjektív vagy lefedő, a bijektív vagy egyrétűen lefedő függvény, a sorozat függvény, a monoton vagy monoton növekvő függvény, illetve az antiton vagy monoton csökkenő függvény.

függvénytulajdonságok

| | |
|--|---|
| function $\equiv F$ | left-total(F) \wedge right-unique(F) |
| inverse_function(m, n) $\equiv F^{-1}$ | right-total(F) \wedge left-unique(F) |
| biunique(F) | left-unique(F) \wedge right-unique(F) |
| bitotal(F) | left-total(F) \wedge right-total(F) |
| injective(F) | function(F) \wedge left-unique(F) |
| surjective(F) | function(F) \wedge right-total(F) |
| bijjective(F) | injective(F) \wedge surjective(F) |
| series(F) | function(F) \wedge domain(F) = N |
| monotone(F(x, y)) | $\forall x \forall y (x \leq y \rightarrow F(x) \leq F(y))$ |
| antitone(F(x, y)) | $\forall x \forall y (x \leq y \rightarrow F(y) \leq F(x))$ |

2.13.1.4 Művelet

A művelet (operáció vagy belső összekapcsolás) fogalma a függvényből származtatható. A művelet meghatározása után sorban lehet definálni az egyes művelettulajdonságokat, mint az asszociatív vagy társítható vagy csoportosítható művelet, a kommutatív vagy felcserélhető művelet, O disztributív Q-ra vagy szétosztható vagy széttagolható művelet, Q disztributív O-ra vagy szétosztható vagy széttagolható művelet, idempotens vagy önhatványozó vagy azonos hatványú művelet, O abszorbtív Q-ra vagy elnyelő vagy adjunktív művelet, Q abszorbtív O-ra vagy elnyelő vagy adjunktív művelet.

művelet és művelettulajdonságok

| | |
|----------------------------------|---|
| operation(x, y) $\equiv O(x, y)$ | function(x, y) \wedge domain = cod |
| associative(O(x, y)) | $\forall x \forall y \forall z (xO(yOz) = (xOy)Oz)$ |
| commutative(O(x, y)) | $\forall x \forall y (xOy = yOx)$ |

| | |
|-----------------------------|---|
| distributive($O(x, y) Q$) | $\forall x \forall y \forall z ((x \circ (y \circ z)) = (x \circ y) \circ (x \circ z))$ |
| distributive($Q(x, y) O$) | $\forall x \forall y \forall z ((y \circ z) \circ x = (y \circ x) \circ (z \circ x))$ |
| idempotent($O(x, y)$) | $\forall x (x \circ x = x)$ |
| absorptive($O(x, y) Q$) | $\forall x \forall y (x \circ (x \circ y) = x)$ |
| absorptive($Q(x, y) O$) | $\forall x \forall y (y \circ (x \circ y) = x)$ |

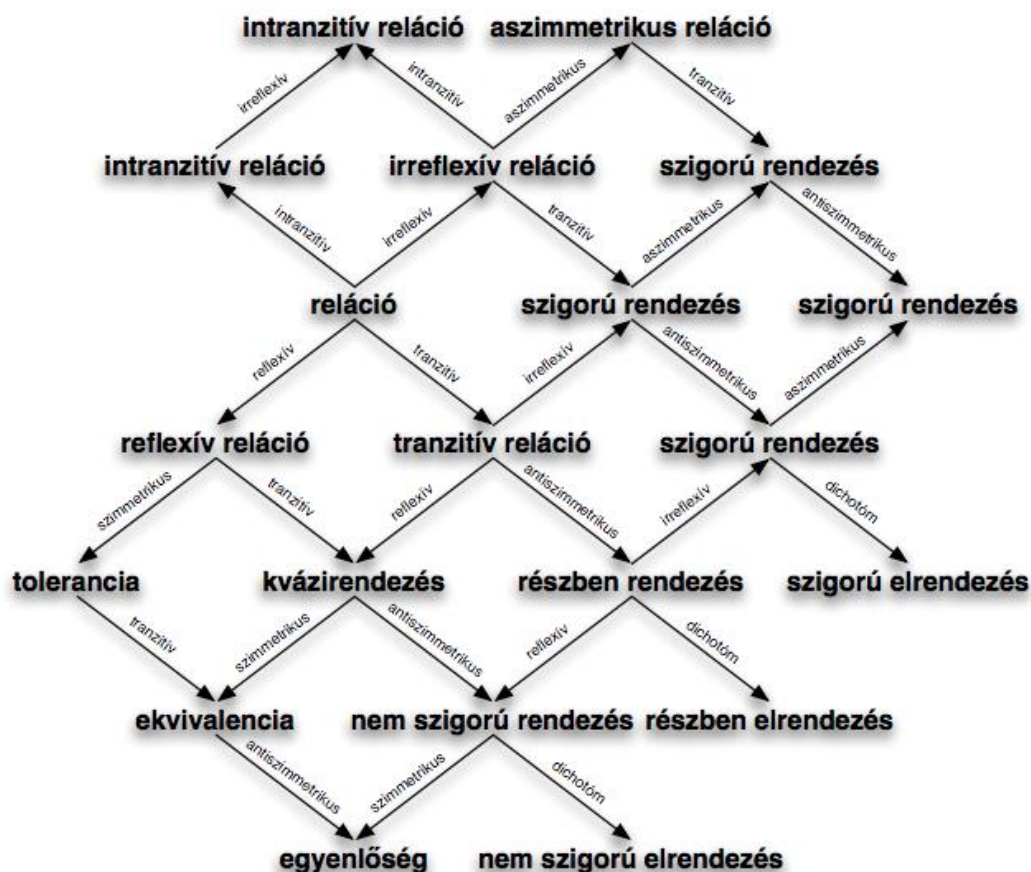
2.13.1.5 Nevezetes matematikai relációk

A matematikában gyakran használt relációknak saját nevük van, és ezeket a relációtulajdonságok alapján szokás meghatározni. A leggyakrabban használt, nevezetes relációk között van a tolerancia vagy kompatibilitási vagy hasonlósági reláció, a részleges ekvivalencia reláció, az ekvivalencia reláció, a részben vagy parciális vagy részleges vagy gyenge vagy gyenge részben rendezés vagy nem szigorú rendezés, a gyenge rendezés, a lineáris rendezés, a szigorú vagy erős részben vagy irreflexív rendezés, a nem szigorú elrendezés reláció vagy teljes parciális vagy lineáris rendezés, a parciális vagy totális parciális rendezés, a szigorú elrendezés vagy teljes szigorú vagy lineáris szigorú vagy totális szigorú rendezés, az előrendezés vagy kvázirendezés, a gyenge rendezés vagy lineáris előrendezés, a gyenge részben rendezés, a részben elrendezés vagy teljes gyenge részben rendezés, valamint a tournament reláció.

nevezetes matematikai relációk

| | |
|---------------------------|---|
| tolerance | reflexive \wedge symmetric |
| partial_equivalence | transitive \wedge symmetric |
| equivalence | reflexive \wedge symmetric \wedge transitive |
| partial_order | reflexive \wedge antisymmetric \wedge transitive |
| weal_order | complete \wedge transitive |
| linear_order | complete \wedge antisymmetric \wedge transitive |
| strict_order | irreflexive \wedge transitive |
| linear_partial_order | reflexive \wedge antisymmetric \wedge transitive \wedge dichotomous |
| linear_strict_order | irreflexive \wedge transitive \wedge dichotomous |
| preorder | reflexive \wedge transitive |
| weak_order | reflexive \wedge transitive \wedge dichotomous |
| weak_partial_order | antisymmetric \wedge transitive |
| linear_weak_partial_order | antisymmetric \wedge transitive \wedge dichotomous |
| tournament | transitive \wedge dichotomous |

Az eddig bemutatott relációk közül sokat (bár nem az összeset) kiválasztva a relációtulajdonságok alapján felrajzolhatunk egy ábrát, amiben látszanak a relációk közti kapcsolatok is.



nevezetes relációk közti kapcsolatok a relációtulajdonságok alapján

2.13.1.6 Relációműveletek

A relációkkal is végezhetünk műveleteket, és ezek közül a legfontosabbak saját megnevezéseket kaptak. Itt bemutatjuk a relációkonvertálás vagy konverzképzés vagy inverzképzés vagy megfordítás, a relációinvertálás vagy inverzképzés, a kontrapozíció vagy transzpozíció, a relációkomplementálás, a duálreláció vagy duális reláció, a relációmetszés, a relációegyesítés vagy relációúnió, a relációszorzás vagy relatív szorzat vagy kompozíció vagy multiplikáció vagy szuperpozíció vagy relációkompozíció, valamint a relációösszeadás vagy összeadás vagy relatív összeadás műveletét.

relációműveletek

| | |
|--|---|
| R-conversion $\equiv R^{-1}(x, y)$ | $\forall x \forall y (R^{-1}(x, y) \leftrightarrow R(y, x))$ |
| R-inversion $\equiv R^{-i}(x, y)$ | $\forall x \forall y (R^{-i}(x, y) \leftrightarrow R(\neg x, \neg y))$ |
| R-contraposition $\equiv R^{-p}(x, y)$ | $\forall x \forall y (R^{-p}(x, y) \leftrightarrow R(\neg y, \neg x))$ |
| R-complementation $\equiv R^{-t}(x, y)$ | $\forall x \forall y (R(x, y) \leftrightarrow \neg R(x, y))$ |
| R-dual $\equiv R^{-d}(x, y)$ | $\forall x \forall y (R^{-d}(x, y) \leftrightarrow \neg R(y, x) \equiv R^{-c}(x, y))$ |
| R-intersection $\equiv \times(R(x, y), Q(x, y))$ | $\forall x \forall y (R \times Q(x, y) \leftrightarrow (R(x, y) \wedge Q(x, y)))$ |
| R-union $\equiv +(R(x, y), Q(x, y))$ | $\forall x \forall y (R + Q(x, y) \leftrightarrow (R(x, y) \vee Q(x, y)))$ |

| | |
|--|--|
| R-composition $\equiv *(R(x,y), Q(x,y))$ | $\forall x \forall y \forall z (R * Q(x, y) \leftrightarrow (R(x, z) \wedge Q(z, y)))$ |
| R-addition $\equiv \dagger(R(x, y))$ | $\forall x \forall y \forall z (R \dagger Q(x, y) \leftrightarrow (R(x, z) \vee Q(z, y)))$ |

2.13.1.7 Reifikáció

A relációk használata során gyakran előfordul, hogy azokat módosított értelemben használjuk, mégpedig úgy, hogy nem vesszük figyelembe a reláció teljes jelentését, csak annak egy részét. Példaként az 'x apja y-nak' kétargumentumú relációra hivatkozhatunk, amelynek reifikált alakjából elhagyjuk az egyik argumentumot és az eredeti relációs minőséget belevetítjük a reláció meghagyott argumentumát adó entitásba: 'x apa'. Ezt műveletet nevezik reifikációnak, aminek van két típusa, a bal-, illetve jobb-reifikáció.

a reifikáció művelete

| | |
|---|---|
| left-reification $(x,y) \equiv 'R(x, y)$ | $\forall x \forall y (R(x, y) \leftrightarrow R(x))$ |
| right-reification $(x,y) \equiv R'(x, y)$ | $\forall x \forall y (R'(x, y) \leftrightarrow R(y))$ |

2.13.2 Ontológiai alaprelációk

Ebben a fejezetben azokat az ontológiai alapfogalmakat vesszük számba, amelyeket minden adatmodell felépítése során használni kell, és amelyek között többféle kapcsolatot is értelmezhetünk. Azért fontos ez a kitekintés, mert ezeket a kapcsolatokat nem nehéz összekeverni, ami aztán könnyen pongyolasághoz, többértelműséghez vezet.

A filozófiai, ontológiai szakirodalomban régóta használt, széles körben elterjedt a partikularé-univerzálé fogalomkettős, bár a két terminus helyett – különösen a filozófián túli tudásterületeken – gyakran használnak más kifejezéseket. Amíg partikularé, illetve az individuum, az instanciacia, az előfordulás, az elem, a példány terminusaival mindig a világ valamilyen egyedi, minden mástól elkülöníthető dologára hivatkozunk, addig a univerzálé, illetve a concept, a property, az osztály, a típus terminusaival mindig valamely jellemző mentén közös dolgok csoportját, gyűjteményét jelöljük. Nem szokás azonban használni a gyűjtemény vagy a halmaz terminusát az univerzálé helyett, és ez a mozzanat felveti annak szükségességét, hogy a partikularé-univerzálé kettős mellé felvegyük a gyűjtemény fogalmát is, és persze tisztázzuk a három ontológiai alapfogalom egymáshoz való viszonyát (Bittner et al. 2004).

Bár az elem és az individuum halmazelmélethez közelebb álló terminusok, belül nem teszünk különbséget a partikularé szinonimái között (bár azért leggyakrabban a partikularé és az individuum terminusokat fogjuk használni). Helytelennek tartjuk viszont a concept, az osztály vagy a property terminusok használatát az univerzálé fogalma helyett, ezért az univerzálé szinonimájaként csak a típus terminusát tartjuk elfogadhatónak. A kérdés tehát az, hogy mi a különbség az individuum, az univerzálé és a gyűjtemény fogalmi között.

Mivel a világban létező dolgok leírására alkalmas fogalmainkat akarjuk valahogyan tipizálni, ezért legelőször a partikularé (individuum) és univerzálé fogalmait kell meghatározunk. Ezt egy speciális metarelációval, az instanciálás (példányosítás) relációval tehetjük meg. Ezt a definíciót használják a jelenkori filozófusok (Lowe 1998), de ezt az értelmezést veszi át a Dolce ontológia is (Gangemi et al. 2001). Az instanciálás metarelációval azt fejezhetjük ki, hogy egy fogalommal leírt „dolognak” mennyi példánya lehetséges a világban. Ha a fogalommal leírt dolog csak „egymagában” létezik, tehát nincsenek példányai, legfeljebb csak annyit mondhatunk, hogy „önmaga példánya”, akkor beszélünk a partikularé fogalmáról.

Ezzel szemben az univerzálé fogalmával olyan dolgot illetünk, amelynek lehetséges több példánya is. Mondhatjuk azt is, hogy a partikularé az instanciálás metareláció jobbról egyértelmű szűkítése, míg az univerzálé e szűkítés komplementere. A három eddig tárgyalt fogalmat Gangemiék alapján a következő formulákkal írhatjuk le (Gangemi et al. 2001).

Először az instanciálás vagy instanciája, előfordulása, példánya relációt definiáljuk.

instanciálás reláció

| | |
|--------------------------|---------------------------------|
| instance = ε | asymmetric \wedge aneuclidean |
|--------------------------|---------------------------------|

Az instanciálás reláció segítségével már meghatározhatjuk a partikularé vagy individuum, illetve az univerzálé fogalmait is.

partikularé és univerzálé

| | |
|------------|---|
| particular | $\forall x(\text{particular}(x) \leftrightarrow \neg \exists y(\text{instance}(y, x)))$ |
| universal | $\forall x(\text{universal}(x) \leftrightarrow \neg \text{particular}(x))$ |

Van tehát három alapvető metafogalmunk, köztük az instanciálás reláció, de amikor továbblépünk, figyelni kell arra, hogy nem szabad összekevernünk az instanciálás metarelációt azzal a relációval, amit a partikularé és az univerzálé fogalmi között értelmezhetünk és aminek az instanciája nevet adhatjuk. Bár az instanciája reláció nagyon hasonlít az eleme relációhoz, mégsem lehet azonos vele, mivel másfajta fogalmak között értelmezzük. Amíg az eleme reláció egy individuum (partikularé) és egy gyűjtemény (halmaz) közötti reláció, addig az instanciája reláció egy partikularé és egy univerzálé között áll fent (Bittner et al. 2004, Gangemi et al. 2001). A meglévő különbségek miatt tehát egymástól különböző relációnak kell tartanunk őket, de a meglévő hasonlóságok miatt az eleme reláció jeléhez (\in) hasonló szimbólummal jelöljük az instanciája metarelációt (ε). Az instanciája relációban van valamilyen intenzionális kapcsolat, ami teljesen hiányzik a halmaz és elemei között fenntartott eleme relációból. Ezért van szükség az univerzálé fogalma mellé felvenni a gyűjtemény kategóriáját, hogy az individuumok és a gyűjtemények közti extenzionális kapcsolatot kifejezhessük az eleme reláció alkalmazásával.

A három fogalomtípus eltérő minőségét szemléltethetjük az időbeli „viselkedésükkel”. Az individuumok elveszthetik részeit vagy épp új részeket „szerezhetnek” (az ember sejtei folyamatosan elhalnak, illetve megszületnek). Az univerzálék is hasonlóképpen viselkednek, csak nem a részeit, hanem a példányaikat veszíthetik el vagy nyerhetnek újakat belőlük. A gyűjtemény (a halmaz) viszont mindig az elemei, összetevői (vagyis az individuumok) által azonosítható, ezért nem lehetséges, hogy különböző időpontokban különböző elemei legyenek.

Van tehát három ontológiai metaegységünk, melyek között különböző relációkat értelmezhetünk. A következő táblázatban először felsoroljuk azokat a kapcsolatokat (1-8. sor), amelyek a három entitástípus között léteznek a partitivitás atommentes (részhalmaz-szerű) és atomos (eleme-szerű) relációtípusa, illetve a genericitás szerint. Ezt kiegészítjük a feltüntetett relációk fontosabb inverzeivel (1-10, 4-9, 7-11, 5-12).

metaszintű keresztrelációk

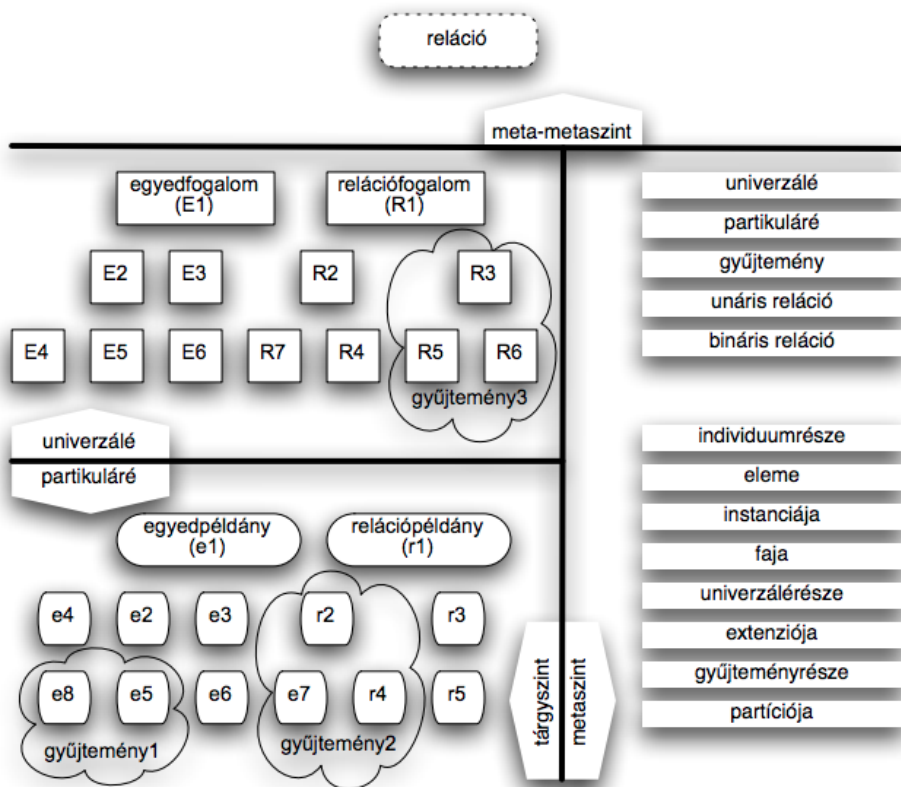
| | reláció jele | bal argumentum | jobb argumentum |
|---|-----------------|----------------|-----------------|
| 1 | individuumrésze | individuum | individuum |
| 2 | eleme | individuum | gyűjtemény |
| 3 | instanciája | individuum | univerzálé |
| 4 | faja | univerzálé | univerzálé |
| 5 | univerzálérésze | univerzálé | univerzálé |
| 6 | extenziója | gyűjtemény | univerzálé |
| 7 | gyűjteményrésze | gyűjtemény | gyűjtemény |

| | | | |
|----|------------------|------------|------------|
| 8 | partíciója | gyűjtemény | individuum |
| 9 | neme | univerzálé | univerzálé |
| 10 | individuumegésze | individuum | individuum |
| 11 | gyűjteményegésze | gyűjtemény | gyűjtemény |
| 12 | univerzáléegésze | univerzálé | univerzálé |

A metasztintú keresztrelációkat önálló szimbólumokkal jelöljük, és a relációk argumentumait infix-módban adjuk meg (tehát a két argumentum közé tesszük a relációjelet). A fenti relációk – legalább az időhöz való viszonyukban – különböznek egymástól, és különböző erősségű formulákkal írhatók le (ezeket itt nem mutatjuk be, részletesen kibontva lásd: Bittner et al. 2004).

A táblázat az ontológiák nagyon fontos metareláció-típusait mutatja be, de jeleznünk kell azt is, hogy egy nagyon fontos reláció hiányzik a felsorolásból. Az intenzió kategóriája – az extenzió mellett – a fogalmak nélkülözhetetlen „tartozéka”, amit – ennek ellenére – a kortárs ontológiák és ontológiaelméletek egyike sem tud (vagy akar) megragadni. Ez persze nem véletlen, hiszen az intenzió, vagyis a fogalmak jelentését, tartalmát kifejező bármilyen állítás, tehát „tetszőleges” logikai formula lehet, amit eddig még nem sikerült egyetlen metafogalomtípushoz rendelni. Mi sem tudjuk formalizálni és a modellbe építeni ezt a fogalmat, de fontosnak tartjuk, hogy egyrészt azért jelezzük a hiányát, másrészt rögzítsük a lehetséges argumentumait: az ‘intenziója’ reláció egy formula és egy univerzálé között teremt kapcsolatot.

Az univerzálé és a partikularé (illetve a gyűjtemény) kategóriái mellett – még mindig a metasztinten maradván – szükségünk van egy másik fontos fogalmi megkülönböztetésre is. A matematikai szinten definiált legáltalánosabb metafogalmunk, a reláció fogalmához a meta-metasztinten hozzákapcsoltuk az extenzionális relációelméletet, melyből két relációtípust „lehozhatunk” a metasztintre és deklarálnak az unáris reláció és a bináris reláció metafogalmait (meg kell jegyezzük, hogy ez a dichotómia még mindig szintaktikai meghatározáson alapul, tehát ezek miatt sem kell semmiféle plusz ontológiai elkötelezettséget vállalnunk). Az unáris és a bináris reláció két fogalomtípusát messzemenően hasznosíthatjuk a tárgyszintű építkezésben, hiszen a bináris reláció metafogalma adja a tárgyszint relációfogalmának típusát, az unáris reláció metafogalmat pedig a tárgyszinten az egyedfogalomnak feleltethetjük meg (mint a tárgyontológia két legfelső kategóriáját). Ha az eddig bemutatott dimenziókat egymásra vetítjük, akkor meghatározhatjuk az ontológia tárgyszintjének legfontosabb metafogalmait. Az univerzálé-partikularé kettősség mentén választhatjuk el egymástól a modell generikus, illetve instanciális szintjét, míg az egyed- és relációfogalmak (vagyis az unáris és a bináris reláció metafogalmak) segítségével megadhatjuk a tárgyszint generikus és instanciális hierarchiájának két-két csúcselemét. A generikus szinten tehát lehetnek egyedfogalmaink és relációfogalmaink, az instancia szinten pedig előfordulhatnak egyedpéldányok és relációpéldányok, és természetesen a gyűjtemények mindkét szinten összefoghatnak univerzálékat és partikularékat egyaránt. A négy ontológiai réteg, a metafogalmak és a köztük levő metarelációk, valamint a tárgyszint legfontosabb összetevőit mutatjuk be a következő ábrán.



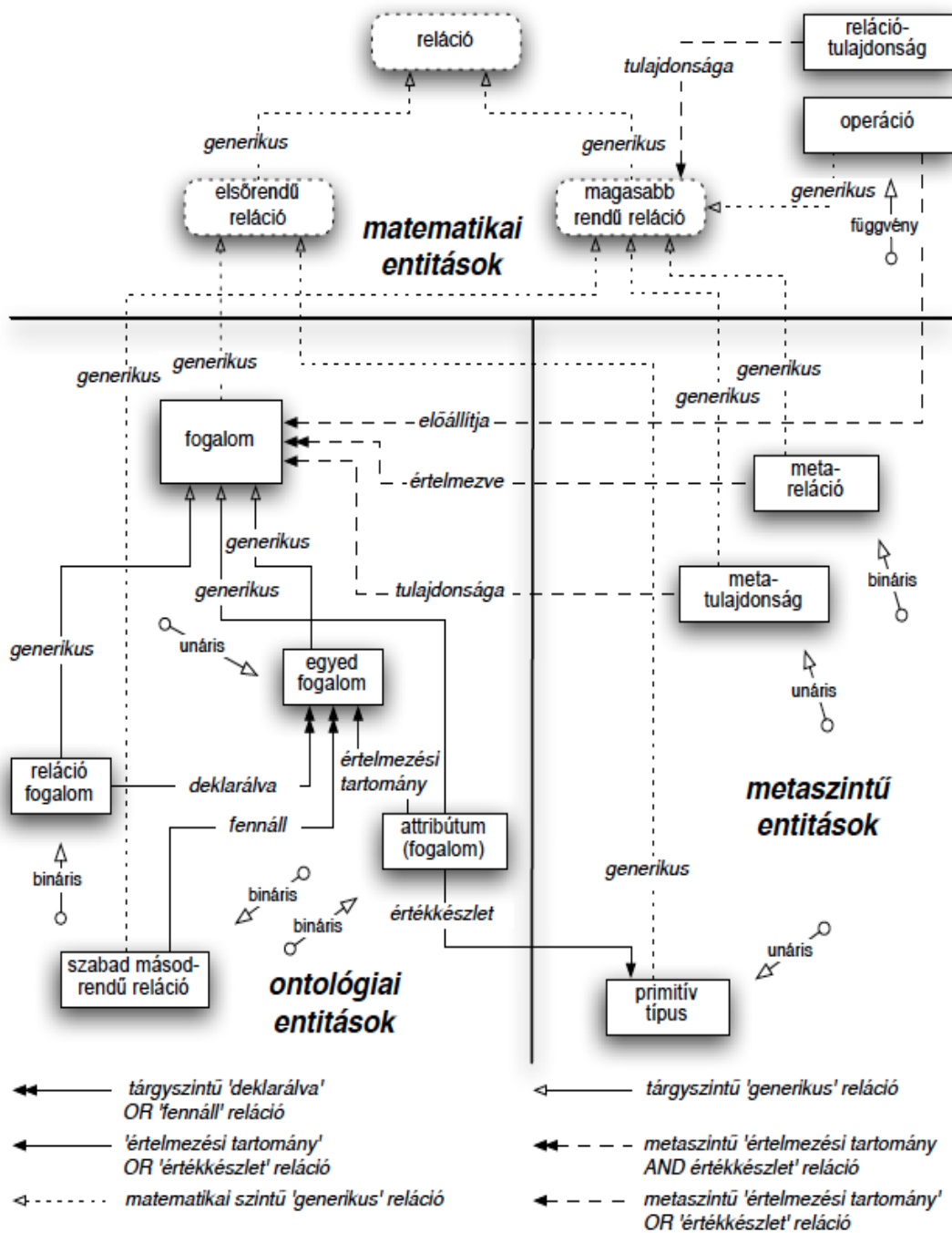
alapfogalmak egymáshoz való viszonya

A keresztrelációk elméletében kellett még definiálni az olyan metafogalmakat, mint az összekapcsolódás (connection), elhelyezkedés (location) vagy a függőség (dependence), melyek közül az utóbbi azért különösen fontos, mert a fogalmak tartalmát jelentő ismertetőjegyek, tulajdonságok formalizálhatóságának ígéretét jelenti.

Az eddig említett elméletek az ontológiatervezési munkával kapcsolatosak. A tényleges ontológiaépítő munka a tárgyszintű építkezést jelenti, amikor az ontológia szerkesztői – saját ontológiai elfogultságaik, elkötelezettségeik alapján – új és új fogalmakat vesznek fel, melyeket ismertetőjegyekkel jellemeznek, illetve meghatározzák az új fogalom helyét a fogalmi struktúrában.

A tárgyszinten a generikus modellnek megfelelően vehetjük fel az univerzálékat, melyekre a metaszinteken előzetesen meghatározott metafogalmak (metatulajdonságok, relációs tulajdonságok, műveletek, generikus alárendeltje reláció stb.), kategóriák (absztrakt, konkrét, egyedfogalom, relációfogalom stb.).

Elsősorban a tárgyszintű ontológiaépítés lehetőségeit szem előtt tartva az alábbi ábrán mutatjuk be a generikus modell legfontosabb fogalmait.



ontológiai szintek és alapfogalmak

2.13.3 KOS-relációk

A tudásszervezési rendszerek, a névterek építése során sokfajta relációra van szükség – igaz, a különböző tudásszervezési rendszerek, valamint a kétféle névtér, a tulajdonnévtér és a köznévtér számára más relációk a fontosak. Vannak relációk, amelyek mindegyik névtér esetében kellenek, de vannak olyanok, amelyeket csak a köznévtér vagy tudásszervezési rendszerek építése során vehetünk használatba. Ebben a fejezetben ezeket vesszük sorba. Elemzésünket az alábbi relációkra terjesztjük ki.

- faja/neme vagy generikus (főlé- és alárendeltje) reláció
- része/egésze vagy partitív (főlé- és alárendeltje) reláció
- homonima reláció
- szinonima reláció
- meghatározója (előzménye, következménye) reláció

- hierarchikus alá-fölérendeltje (tartalmazás) reláció
- lexikografikus reláció
- rokona (egyéb) reláció

2.13.3.1 Generikus reláció

A tudásszervezési rendszerek építésében (vagy általánosabban: az emberi tudás fogalmakkal való reprezentációjában) a legfontosabb reláció a generikus reláció. Az egyértelműség kedvéért érdemes rögzíteni, hogy itt nem egyetlen relációról, hanem inkább egy relációkettősről kell beszélnünk. A generikus reláció fogalmát általában a generikus alárendeltje reláció értelemben szokták használni, de nem kevés alkalommal szükség lehet a reláció inverzére, a generikus fölrendeltje relációra is. Ha nem zavaró, akkor a generikus reláció kifejezéssel hivatkozhatunk a generikus alá- és fölrendelt relációra együttesen.

A generikus reláció a relációk (predikátumok, univerzálék) közt teremt kapcsolatot, ezért másodrendű relációnak kell tekintenünk. A relációt definiáló formulával azt fejezzük ki, hogy ha veszünk egy általánosabb (Q) és egy speciálisabb fogalmat (P), akkor az utóbbi bármely példányára igaz lesz, hogy az előbbi példány is, miközben ez fordítva nem igaz.

generikus reláció

| | |
|----------------------------|--|
| generic_subordinate(P,Q) | $\forall x \forall P \forall Q ((P(x) \rightarrow Q(x)) \wedge \neg(Q(x) \rightarrow P(x)))$ |
| generic_superordinate(P,Q) | $\forall x \forall P \forall Q ((Q(x) \rightarrow P(x)) \wedge \neg(P(x) \rightarrow Q(x)))$ |

Más kontextusokban ugyanerre a fogalomra használják az 'osztályozási reláció' kifejezést is, sőt, a 'típusa' relációt is gyakran ebben az értelemben alkalmazzák. A keresztrelációk elméletében korábban a 'neme' és 'faja' kifejezéseket alkalmaztuk erre a két fogalomra. Előfordul még az inklúzió mint magába foglalás, a szubszumpció mint tartalmazás terminusa is, amelyek inkább a generikus fölrendelés értelmét fejezik ki. Olykor felbukkan a hipo- és hipernima kifejezéspár is, aminek jelentése a generikussággal azonos. Az 'osztály-alosztály', illetve a 'típus-altípus' besorolások a generikusság itt bemutatott értelmét fejezik ki, ezért ezek a terminusok tekinthetők egymás szinonimáinak, de vigyázni kell akkor, amikor a – szintén széles körben elterjedt – IS-A relációt akarnánk használni a generikus reláció szinonimájaként, mert ezzel a relációval lehetnek problémák. A generikus reláció nagyon fontos tulajdonsága a tranzitívitas, és ez a minőség nem minden esetben áll fenn – a maga teljességében – az IS-A reláció esetében. Ennek az az oka, hogy az IS-A reláció használata során előfordul, hogy azt kiterjesztik egyfelől az 'IS-A-TYPE-OF' (vagy az 'IS-A-KIND-OF', illetve 'IS-A-SORT-OF') relációra, másfelől az 'IS-AN-INSTANCE-OF' relációra egyaránt, és ez – olykor – nem várt következményekhez vezethet. Vegyünk egy példát! A következő három mondat mindegyike igaz lehet.

állat IS-A élőlény

puli IS-A állat

Bodri IS-A puli

Az első két esetben a generikus relációval kötünk össze kategóriákat, ám a harmadikban már a példány relációt használjuk, ami furcsa változást hoz az 'is-a' relációval összekapcsolt entitások esetében. Nem pontos azt mondani, hogy a tranzitívitas minősége sérül, inkább egy furcsa változás áll be a tranzitívitas sorban. A generikus ('IS-A-TYPE-OF') reláció esetében elméletileg mindig van lehetőség arra, hogy a generikus relációval összekötött kategóriák közé be lehessen tenni egy újat úgy, hogy a tranzitívitas megmarad. A fenti példát kiegészítve mondhatjuk még a következőt.

puli IS-A kutya

kutya IS-A állat

Ekkor a puli és az állat fogalmi közé betettünk egy újat (a kutya fogalmát), ami által létrehoztunk egy négy lépcsőből álló tranzitív sort.

puli IS-A kutya IS-A állat IS-A élőlény

Amikor a generikusként értelmezhető IS-A relációval összekapcsolt elemek közé beteszünk egy instanciájaként felfogható IS-A relációt, akkor a fent jelzett alábonthatóság lehetősége már nem áll fent. Ha azt mondjuk, hogy:

Bodri IS-A puli

akkor a 'Bodri' alá már nem vehetünk fel semmi mást az IS-A reláció segítségével. Ez nem borítja fel a tranzitivitást, de balról lezárja új elemek felvételének lehetőségét a generikus reláció segítségével. Az IS-A reláció jobb oldalán sosem szerepelhet a 'Bodri' elem. Ezzel szemben a valódi generikus struktúra esetén bármelyik elem szerepelhet a kapcsolat mindkét oldalán. Leszámítva persze a generikus hierarchia csúcspontját álló 'valami' fogalmat.

A névtérépítés során az egyik legfontosabb feladat annak biztosítása, hogy a generikus relációt és a példány relációt mindig elkülönítsük egymástól, mert az összevont kezelésük könnyen zavart okozhat. A példány és a generikus relációk különbsége segít majd nekünk később, a névterek definiálásakor abban, a névterek különböző típusait egyértelműen és pontosan elhatárolhassuk egymástól.

A generikus reláció kapcsán fontos figyelembe venni azt a megkülönböztetést, miszerint a hierarchikus relációnak két fajtája is van: a **polihierarchikus** és a **monohierarchikus** reláció. A monohierarchikus struktúrán belül minden elemnek egyetlen közvetlen felettese van, míg a polihierarchiára ez nem igaz, vagyis itt megengedett a több szülő létezése. A tranzitivitást nem változtatja meg ez a fogalmi felbontás, de a struktúra elemei közti kapcsolatok minősége már más a két rendszeren belül. A monohierarchia fagráfokkal reprezentálható (aminek legfőbb jellemzője „körmentesség”, vagyis az, hogy két pont között mindig csak egyetlen útvonal létezik), míg a polihierarchiára csak annyit mondhatunk, hogy azok tranzitív, irányított gráfok, amiben két elem között többféle útvonal is létezhet.

2.13.3.2 Partitív reláció

A rész/egész relációval vagy másként a partitív alárendeltje/partitív fölérendeltje relációval a mereológia elmélete foglalkozik. A része (partitív alárendeltje) reláció az egésze (partitív fölérendeltje) inverze, és ez relációkettős is tranzitív, hierarchikus – akárcsak generikus reláció. A tranzitivitás azonban csak a pontosan definiált – és egymástól is jól elkülönített – altípusokra érvényes, nem általában a partitív relációra. Sok esetben érvényesül a tranzitivitás. Vegyünk két példát!

Ha Érd Pest megye része, és Pest megye Magyarország része, akkor Érd is Magyarország része.

Ha az ujjam a kezem része, és a kezem a karom része, akkor az ujjam a karom része is.

Könnyen lehet azonban olyan példát találni, amikor a partitív reláció intranzitivitást tapasztalhatunk.

A zongorista keze része a zongoristának, a zongorista része a zenekarnak, de a zongorista keze nem része a zenekarnak.

Az ellentmondásosnak tűnő helyzet magyarázata egyszerű: többféle partitív reláció létezik, és ezek között nem érvényes a tranzitivitás. Olyan mereológiai elméletre van ezért szükség, amely kezelni, magyarázni tudja ezt a jelenséget. A versengő mereológiai elmélet közül itt a Peter Simons és Archello Varzi nevéhez köthető elméletet fogadjuk el (Simons 2000, Varzi 1996).

A partitív relációra igaz az alábbi három tulajdonság.

mereológiai axiómák

| | |
|------|---|
| (a1) | $\forall x(P(x, x))$ |
| (a2) | $\forall x\forall y(P(x, y) \wedge P(y, x) \rightarrow x = y)$ |
| (a3) | $\forall x\forall y\forall z(P(x, y) \wedge P(y, z) \rightarrow P(x, z))$ |

Az első három axióma rögzíti, hogy a P partitív reláció reflexív (a1), antiszimmetrikus (a2) és tranzitív (a3). A következő lépésben a P relációval néhány új mereológiai relációt definiálunk, majd az újakkal még újabbakat hozunk létre:

| | | |
|----------------|---|---------------------------|
| (1) $PP(x, y)$ | $\forall x\forall y(P(x, y) \wedge \neg P(y, x))$ | valódi része |
| (2) $O(x, y)$ | $\forall x\forall y\exists z(P(z, x) \wedge P(z, y))$ | átfedése |
| (3) $U(x, y)$ | $\forall x\forall y\exists z(P(x, z) \wedge P(y, z))$ | kívüllevősége |
| (4) $OX(x, y)$ | $\forall x\forall y(O(x, y) \wedge \neg P(x, y))$ | részhatáros átfedése |
| (5) $UX(x, y)$ | $\forall x\forall y(U(x, y) \wedge \neg P(x, y))$ | részhatáros kívüllevősége |
| (6) $PO(x, y)$ | $\forall x\forall y(OX(x, y) \wedge OX(y, x))$ | valódi lefedése |
| (7) $PU(x, y)$ | $\forall x\forall y(UX(x, y) \wedge UX(y, x))$ | valódi kívüllevősége |

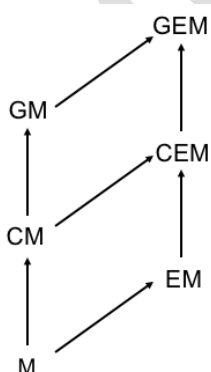
Sorban definiáljuk a következő relációkat: valódi része (1), átfedése (2), kívüllevősége (3), részhatáros átfedése (4), részhatáros kívüllevősége (5), valódi lefedése (6), valódi kívüllevősége (7). Ezek után a P része, O átfedése és U kívüllevősége relációk segítségével először meghatározhatjuk a kiterjeszthetőség (erős) elvét (a4), majd az (a5)-(a7) mereológiai axiómák elfogadásával definiálhatjuk a zárt (extenzionális) mereológia elméletét.

| | |
|------|--|
| (a4) | $\forall x\forall y(\neg P(x, y) \exists z(P(z, x) \wedge \neg O(z, y)))$ |
| (a5) | $\forall x\forall y(U(x, y) \rightarrow \exists z\forall w(O(w, z) \leftrightarrow (O(w, x) \vee O(w, y))))$ |
| (a6) | $\forall x\forall y(O(x, y) \rightarrow \exists z\forall w(P(w, z) \leftrightarrow P(w, x) \wedge P(w, y)))$ |
| (a7) | $\forall x\forall y\exists z((P(z, x) \wedge \neg O(z, y)) \rightarrow \exists z\forall w(P(w, z) \leftrightarrow (P(w, x) \wedge \neg O(w, y))))$ |

További axiómák felvételével egyrészt definiálhatjuk az általános (extenzionális) mereológia elméletét (a8) másrészt meghatározhatjuk az atomos (a9), illetve atom-nélküli mereológiákat (a10).

| | |
|-------|---|
| (a8) | $\exists x\phi \rightarrow \exists z\forall y(O(y, z) \leftrightarrow (\exists x(\phi \wedge O(y, x)))$ |
| (a9) | $\forall x\exists y(P(y, x) \wedge \neg \exists zPP(z, y))$ |
| (a10) | $\forall x\exists y(PP(y, x))$ |

Nem szükséges a fenti axiómák mindegyike ahhoz, hogy egy mereológiai elméletet definiálni lehessen. Varzi többfajta mereológiai elméletet különített el más és más axiómák együttes elfogadásával.



| | |
|-----|---|
| M | mereológia: (a1)-(a3) |
| EM | extenzionális mereológia: (a1)-(a4) |
| CM | zárt mereológia: (a1)-(a3), (a5)-(a7) |
| CEM | zárt extenzionális mereológia: (a1)-(a7) |
| GM | általános mereológia: (a1)-(a3), (a8) |
| GEM | általános extenzionális mereológia: (a1)-(a4), (a8) |

Ezek a mereológiai elméletek különböző erősségű és tartalmú ontológiai elkötelezettségek mentén érvényesek. A fenti hat elmélet egyikében sem szerepel azonban az az utolsó két mereológiai axióma, amelyek alapján elkülöníthetjük az atomos és nem-atomos partitív relációkat egymástól. Az (a9) és

(a10) jelű tételek alkalmazásával a fenti elméletek tovább bonthatók atomos vagy atom-nélküli alelméletekre.

A partitív reláció is tranzitív. Jeleztük azonban, hogy a tranzitivitás nem általában a partitív reláció egészére, hanem annak valamelyik típusán belüli kapcsolatokra igaz. A fenti formulák alapján most már mondhatjuk azt, hogy a partitív reláció tranzitivitása mindig csak valamelyik – fent definiált – mereológiai elméleten belül érvényesül. A partitív reláció tranzitivitása akkor válik érvénytelenné, amikor egy atomos és atom-nélküli partitív relációra akarjuk érvényesíteni. Az emberi test és a testrészek közötti viszonyt atom-nélküli partitív relációval írhatjuk le, ezzel szemben a testületek és személyek közti szervezeti tagságot atomos partitív reláció alkalmazásával tudjuk megragadni. A kétféle rész-egész relációt azonban nem szabad összekeverni, különben ahhoz a tranzitivitási deficithez jutunk, amit korábban már bemutattunk a zenekari tag zongorista keze kapcsán. Természetesen lehet többféle partitív relációt használni egy rendszeren belül, de ilyen esetekben a tranzitivitás létezését csak adott partitív reláción belül várhatjuk el, a – részben – eltérő minőségű partitív relációk közt nem számíthatunk rá. Mindebből az következik a névtérépítésre vonatkozóan, hogy – szükség esetén – többféle partitív relációt kell definiálni és használni, hogy az ilyen keveredést elkerüljük.

2.13.3.3 Szinonima

A szinonima reláció a nyelv formai többértelműségét mutató nyelvi reláció. Nyelvi konstrukciók (tehát nyelvi formák) közötti kapcsolatot fejez ki azáltal, hogy azok jelentései (tartalmi) közti azonosságot fejezi ki adott nyelven belül. A magyar nyelvben a 'kutya' és az 'eb' egymás szinonimái. Bármelyiket használjuk, ugyanarra az állatra utalunk vele. A következő két mondatnak – durván – ugyanezt a jelentést tulajdonítjuk.

A postást megharapta a kutya.

A postást megharapta az eb.

A szinonima relációban álló terminusok egymással felcserélhetők egy nyelvi konstrukción belül anélkül, hogy a megnyilatkozás jelentése ezáltal módosulna. Ebből az is következik, hogy a szinonima reláció szimmetrikus, meg az is, hogy valamifajta egyenlőséget fejezhetünk ki vele. A relációnak az a lényege, hogy a szinonim egységek ugyanarra a jelentésre mutatnak. A nyelvhasználatunk lényege az, hogy szintaktikailag jól-formált nyelvi konstrukciókat (szavakat, kifejezéseket, mondatokat) – konvencionális módon, tehát teljesen önkényesen – összekapcsolunk jelentésekkel. Amikor ezt tesszük, akkor a jelentések tartalmát nyelvi formák reprezentálják. A szinonimitás azt jelenti, hogy két – egymástól különböző – nyelvi formátum ugyanazt a tartalmat hordozza.

A szinonima reláció esetében fontos megkötés az egynyelvűség, hiszen két különböző nyelv egységei is mutathatnak ugyanarra a jelentéstartalomra, és ezeket nem szokás egymás szinonimáinak tekinteni. Az angol 'dog' ugyanazzal a jelentéssel bír, mint a magyar 'kutya' szó, de ezek inkább egymás nyelvi fordításai, mint egymás szinonimái.

Fontos megjegyezni, hogy a szinonima reláció által összekapcsolt nyelvi egységeknek a valóságban sosincs teljesen azonos jelentése, bizonyos nyelvhasználati kontextusokban nem lehet feltétlen kicserélni egymással a – máshol, máskor – szinonim egységeket. A „kutya hideg van odakint” mondat helyett nem mondhatjuk azt, hogy 'eb hideg van odakint'.

2.13.3.4 Homonima

A homonima reláció a nyelv tartalmi többértelműségét mutató nyelvi reláció. Jelentések (tehát nyelvi tartalmak) közötti kapcsolatot fejez ki azáltal, hogy azok nyelvi megjelenései (formái) közti azonosságot fejezi ki adott nyelven belül. A homonimának két típusát el lehet különíteni aszerint, hogy írásos vagy szóbeli nyelvi egységekről van-e szó. Előbbi esetben homográfáról, utóbbi esetben homofóniáról beszélnek. Mivel itt rögzített, tehát írás-alapú kommunikációról van szó, a homonima relációt mindig írásos nyelvi egységek között értelmezzük. A homonima esetén arról van szó, hogy ugyanazzal a nyelvi alakkal más jelentés fejezünk ki, tehát különbözik a tartalom. A homonima leggyakrabban idézett,

klasszikus, magyar példája az 'ár' szóval kapcsolatos. Az alábbi négy nyelvi egység alakilag megegyezik, de jelentését tekintve különböző.

ár mint árhullám

ár mint értékmérő a piacon

ár mint szűrőszerszám

ár mint területi mértékegység

A homonima reláció meghatározásakor ugyanazokat a megoldásokat kell alkalmaznunk, mint a szinonima reláció esetében, csak itt az azonosságot az eltérő jelentésű megnyilatkozások nyelvi alakjaira kell vonatkoztatnunk.

2.13.3.5 Meghatározója (előzménye, következménye) reláció

A tudásszervezési rendszerek közül a tezauruszokban használják azt a relációpárt, amelyet nem igazán lehet egyértelmű és egzakt jelentéssel ellátni, inkább többféle jelenséget egybefogó ernyőfogalomként lehet értelmezni. Ebben a – sokszor az okság fogalmához kapcsolódó – szövevényes jelentéshálóban is lehet inverzrelációt definiálni, annak érdekében, hogy a különböző alkalmazási helyzetekben a kényelmesebb megoldást (a kényelmesebb irányban kifejezett relációt) lehessen választani.

A meghatározója relációpár nagy általánosságban az ok-okozat, cél-eszköz-következmény, tárgy-folyamat-funkció fogalomcsoporttal jellemezhető kapcsolatokat fejezi ki a dolgok között. Valamiféle meghatározottságot jelölhetünk vele. Amikor ennek a meghatározottságnak az eredmény oldalára fókuszálunk, akkor beszélhetünk a **rezultánsa** relációról. Ezzel az jelöljük, hogy a bal-oldali relátumnak (dolognak, tárgynak, folyamatnak) mi a meghatározottja (rendeltetése, okozata, eredménye, terméke, célja, tárgya, következménye), amit a jobb-oldali relátummal fejezünk ki. Az inverzreláció, amit szoktak **előfeltétele** vagy kiindulása relációnak nevezni, azt fejezi ki, hogy a bal-oldali relátum (dolog, tárgy, folyamat) mi a meghatározója (létének, létrehozásának, működésének, meghatározásának oka, előidézője, eszköze, alapja), amire a jobb-oldali relátummal utalunk.

Ha a relációkettőt szemantikailag kellő mértékben szűk jelentéstartományon belül tartjuk, akkor elképzelhető, hogy hierarchikus struktúrákat lehet építeni ebben a dimenzióban is. Például bizonyos típusú események esetében lehet beszélni az egymást követő események hierarchikus láncolatáról.

2.13.3.6 Hierarchikus alá-fölrendeltje (tartalmazás) reláció

A generikus és a partitív reláció mentén a tudásszervezési rendszerek elemei hierarchikus struktúrába, alá-, fölrendeltségi kapcsolatrendszerbe szervezhetők. Mivel ez a két reláció szabatosan definiált, mindkettő egyértelmű, konzisztens struktúrát eredményez. Vannak azonban olyan tudásszervezési rendszerek is, amelyben csak annyit követelnek meg, hogy a rendszer egységei között létezen valamilyen tartalmazási, hierarchikus alá- és fölrendelési kapcsolat úgy, hogy a relációk szemantikai egyértelműségét nem követelik meg, csak az elemek között értelmezhető hierarchikus jelleg létét. Angolul pontosan fejezi ki ezt a minőséget a 'broader term', illetve 'narrower term' kifejezőpár, ami csak arra utal, hogy valamilyen szempont szerint az így összekapcsolt kategóriák között valamilyen mennyiségi, tartalmazási kapcsolat létezik, semmi több.

Amikor egy hierarchikus rendszert úgy építünk fel, hogy közben a tartalmazásnak többféle értelmét engedjük meg egyszerre, akkor egy pongyola, mert jelentését tekintve többértelmű, de ettől még jól rendezett struktúrához jutunk. Ezt a pongyolaságot és a belőle fakadó inkonzisztenciát ezért szokás bevállalni, mert ilyen esetekben a rendszer építése, használata egyszerűbb, hiszen kevesebb – olykor szemantikai értelemben nehéz – intellektuális döntést kell hozni.

2.13.3.7 Lexikografikus rendezési reláció

A lexikografikus (ábécé szerinti) rendezési reláció nyelvi egységek lineáris sorba rendezését jelenti, matematikai értelemben **gyenge teljes rendezésnek** felel meg a **nyelvi egységek** halmazán. A könyvek, tanulmányok végén található hivatkozáslisták bibliográfiai tételekből állnak (általában a szerzők neveivel kezdődve), és ezeket a tételeket lexikografikus rendezés szerint sorolják be egymás mögé. A tudásszervezési rendszerek egységeinek, terminusainak megjelenítését, felsorolását általában a legfontosabb struktúraképző reláció mentén szokták megoldani, de sokszor adódhat olyan helyzet, amikor az egységeket vagy azok egy részét (például egy hierarchián belül az azonos szinten levő, azonos szülő alá tartozó egységeket) a lexikografikus rendezéssel teszik ki egymás után.

2.13.3.8 Rokona (egyéb) reláció

A tudásszervezési rendszerek világában a rokona reláció nem a biológiai értelemben vett leszármazási rokonságot vagy a társadalmi értelemben társadalmi viszonyt jelenti, hanem pusztán csak arra szokás utalni vele, hogy az általa összekötött két relátum között van-e valamilyen – kellő pontossággal nem meghatározható – hasonlóság. Ezt a relációt akkor veszik fel két elem között, ha az adott tudásszervezési rendszeren belül a jelezni kívánt kapcsolat nem feleltethető meg egyik rendszeren belüli, jól-definiált relációnak sem (például teauruszokon belül a generikus vagy partitív relációnak), de egy gyengébb hasonlósági, rokonsági relációval érdemes rögzíteni ezt a gyenge, csak egy pongyola, nem jól-definiált relációval kifejezhető kapcsolatot.

2.14A használatba vehető relációk listája

Az eddig elemzett relációk közül az alábbiakat fogjuk a tudásszervezési rendszerek leírásakor használni. Az egyszerűség kedvéért itt „egyként” hivatkozunk azokra a relációkra is, amelyeket az inverz párjukkal együtt szoktak alkalmazni. Az egyértelműség kedvéért jelezzük azt is, ha az elméleti bemutatás során alkalmazott névhez képest másfajta megnevezés mentén kerülhet használatba egy reláció.

KOS-relációk listája

| | | |
|-----------------|--|--|
| R ₁ | lexikografikus rendezési reláció | |
| R ₂ | ekvivalenciareláció | |
| R ₃ | hierarchikus alárendeltje (tartalmazás) alárendeltje reláció | |
| R ₄ | különbözőségi reláció | |
| R ₅ | generikus alá- és fölérendeltje relációpár | fölérendeltje-alárendeltje, faja-neme relációpár |
| R ₆ | partitív alá- és fölérendeltje relációpár | része-egésze, pars-totum relációpár |
| R ₇ | következménye-előzménye relációpár | meghatározója (rezultáns)-meghatározója (előfeltétel) relációpár |
| R ₈ | rokona (egyéb) reláció | |
| R ₉ | lásd és helyette szinonimareláció-pár | |
| R ₁₀ | címkegyakorisági függvény | |
| R ₁₁ | címke-címke együttjárás reláció | |
| R ₁₂ | felhasználó-címke együttjárás reláció | |

| | |
|-----------------|---|
| R ₁₃ | felhasználó-dokumentum együttjárési reláció |
| R ₁₄ | felhasználói aktivitás függvény |
| R ₁₅ | dokumentumcímkézési gyakorisági függvény |
| R ₁₆ | tetszőleges relációfogalom |
| R ₁₇ | tetszőleges osztályfogalom |
| R ₁₈ | tetszőleges attribútumfogalom |

2.15 Tudásszervezési rendszerek

A nevek rendezett gyűjteményét, vagyis nevek adott halmazát és a rajtuk értelmezett relációk együttesét matematikai értelemben **struktúrának** nevezzük. Más ismeretterületeken ugyanerre a képződményre a **tudásszervezési rendszerek** (Knowledge Organization System, KOS) fogalmát alkalmazzák. Többféle tudásszervezési rendszert szokás elkülöníteni az egyszerűbb struktúráktól az egyre összetettebbek felé haladva, úgymint a **terminuslista**, az **osztályozási rendszer**, a **tezaurusz** és az **ontológia**, illetve az építkezési logikát tekintve el lehet választani az előző kontrollált rendszereket a kontroll-nélkül szervezett **folkszonómiától**.

A tudásszervezési rendszerek a világ leírásához szükséges terminusokat gyűjtik egybe, és a terminusok közötti relációkon keresztül a megkönnyítik a hozzáférést magukhoz a terminusokhoz. A különböző típusú tudásszervezési rendszerek igazából csak a relációikban különböznek egymástól, a felvehető terminusokra nincs semmilyen ontológiai megkötés. Elvileg mindegyik rendszer alkalmas arra, hogy a terminusaival tetszőleges tudásterületet lefedhessünk.

A tudásszervezési rendszerek bizonyos értelemben érzéketlenek arra, hogy tulajdonneveket vagy közneveket kell-e kezelniük, gyakran elő is fordul, hogy egy konkrét tudásszervezési rendszerben keverten szerepelnek tulajdon- és köznevek.

A tudásszervezési rendszerekre vagy legalábbis egy részükre lehet (szokták) alkalmazni a névtér fogalmát, ami a fenti értelmezések alapján érthető, kezelhető. Szerencsésebb lenne a tulajdonnév és köznévi rendszereket jobban elkülöníteni egymástól, de inkább az a fontos, hogy adott rendszerről tudjuk, milyen névállományból épül fel. Ha fontos, a tulajdonnévtereket megfelelő jelző felvételével lehet jelezni.

2.15.1 Formális keretrendszer

A tudásszervezési rendszerek abban a különböznek egymástól, hogy a rendszer elemei között milyen relációkat engedélyeznek felvenni. Éppen ezért matematikai értelemben a tudásszervezési rendszerek matematikai struktúráknak tekinthetők.

Matematikailag a KOS a tárgyszavak tartóhalmazán értelmezett struktúra, amelyből természetesen többféle is lehetséges annak megfelelően, hogy milyen relációkat engedünk meg a tárgyszavak között. A struktúra az alábbi formában írható fel.

$$\text{KOS} = \langle D, R_1, R_2, \dots, R_i, R_n \rangle, \text{ ahol}$$

D a tudásszervezési rendszer tárgyszavaiból álló tartóhalmaz,
 R_i az elemeken (tárgyszavakon) értelmezhető reláció.

A matematikai leírást azonban ki kell egészítenünk, ugyanis később látni fogjuk, hogy a tudásszervezési rendszerek különböző típusai nemcsak a tárgyszavaikban és a relációikban, hanem más jellemzőikben is eltérhetnek egymástól. Mielőtt azonban bővítenénk a tudásszervezési rendszer fogalmának összetevőit, a tudásszervezési rendszerek típusairól kell pár szót szólnunk. A különböző

archívumépítési gyakorlatokban, a hálózati kultúra időszakát megelőző évszázadban, érdemben és kiterjedt módon háromféle tudásszervezési rendszert vettek használatba, majd a hálózati kommunikáció további típusok megjelenését tette lehetővé. Az alábbi öt KOS-típusról van szó.

- hagyományos archívumokban (is)
 - terminuslista
 - taxonómia
 - tezaurusz
- csak digitális archívumokban
 - ontológia
 - folkszonómia

Ezek mindegyike a rá jellemző matematika struktúrával, valamint a rendszerek gyakorlati működtetésére vonatkozó szabályrendszerrel írható le. A tudásszervezési rendszerek alaphalmaza (D) azokat a szavakat, kifejezéseket (terminusokat) tartalmazza, amelyeket a dokumentumokhoz lehet rendelni. A tudásszervezési rendszerek különbségeit az a tény határozza meg, hogy milyen – szintaktikai, szemantikai vagy pragmatikai – relációkat (R_i) engedünk meg felvenni a rendszer elemei között. Nem elég azonban csak a matematikai struktúrára figyelni, ha igazán meg akarjuk érteni a tudásszervezési rendszerek jelenségét. Azt is fel kell vennünk e rendszerek jellemzői közé, hogy van-e, és ha igen, akkor milyen felügyelet, milyen kontroll van a metaadat-hozzárendelési munka folyamatában. Előbb persze meg kell mondanunk, miért is van szükség ennek fi gyelmebevételére. Nos, ha a tartalmi metaadatok dokumentumokhoz rendelésének az a fő funkciója, hogy egyértelműen jellemezzük általuk a dokumentumok tartalmát, akkor az egyértelműséget (vagyis a többértelműségek elkerülését) biztosítanunk kell valahogy. A többértelműségek elkerülése pedig megfelelő szakértelmet, fegyelmezett munkarendet, szakmai kontrollt, kontrollált szótárakat és tudásszervezési rendszereket kíván. Ezt persze nem olyan könnyű formalizálni, hiszen olyan kérdésekre kell választ találnunk e szempont alapján, mint:

- $Q_1 = \{ \text{támasztanak-e bármilyen szakmai feltételt, szaktudással kapcsolatos elvárást a munkát végzők személyével szemben?} \}$
- $Q_2 = \{ \text{van-e bármilyen munkaszervezési szabályrendszer, ellenőrzési mechanizmus a munka menetére vonatkozóan, azaz kik, milyen jogosultságokkal vehetnek részt a munka egyes részfolyamataiban?} \}$
- $Q_3 = \{ \text{kik rendelhetik a KOS-rendszer elemeit a dokumentumokhoz?} \}$
- $Q_4 = \{ \text{kik szerkeszthetik, módosíthatják, bővíthetik a tudásszervezési rendszer elemeit, relációit?} \}$

A tudásszervezési rendszerek építésének és alkalmazásának kontrolljára vonatkozó fenti Q_i kérdésekre különböző válaszokat adhatunk, és ezt a feltételegyüttest vagyis az S_j társadalmi normák összefüggő rendszerét érdemes felvenni a tudásszervezési rendszer jellemzői közé. A dolgokat kissé leegyszerűsítve a következő tevékenységekre vonatkozó normákat kell rögzítenünk:

- S_1 – kinek szabad új tárgyszót létrehozni a tudásszervezési rendszerben
- S_2 – kinek szabad új relációt létrehozni a tudásszervezési rendszerben
- S_3 – kinek szabad két tárgyszót relációba állítani a tudásszervezési rendszerben
- S_4 – kinek szabad tárgyszót dokumentumhoz rendelni a katalógusban
- S_5 – csak a tudásszervezési rendszer elemeit szabad-e a dokumentumokhoz rendelni

Az öt norma közül az első három a tudásszervezési rendszerek építésével, az utolsó kettő a katalógusok bővítésével kapcsolatos. A számítógépek világában a fenti normák mind kezelhetők azáltal, hogy a digitálisan szabályozzuk, kinek van írási joga a tudásszervezési rendszer és/vagy a katalógus elemeire, illetve milyen adatokat lehet egymással összekapcsolni. Az írási, szerkesztési jogosultságokat is figyelembe véve már felírhatjuk a tudásszervezési rendszerek teljesebb formuláját, amelynek segítségével aztán majd megmutathatjuk az egyes típusok közti különbségeket is:

$$KOS = \langle D, R_1, R_2, \dots, R_n, S_1, S_2, S_3, S_4, S_5 \rangle, \text{ ahol}$$

D a tudásszervezési rendszer tárgyszavaiból álló tartóhalmaz,

R_i az elemeken (tárgyszavakon) értelmezhető reláció ($i=1, \dots, n$)
 S_j a tudásszervezési rendszer

A fent bemutatott összetevőkkel már adott az az általános keret, amelyre támaszkodva elég pontosan megragadhatjuk a történelmileg létező, szélesebb körben elterjedt tudásszervezési rendszerek legfontosabb jellemzőit. Minden tudásszervezési rendszerben van egy közös reláció, a lexikografikus rendezés, amely a tárgyszóhalmaz elemeinek ábécé szerint való sorba állítását jelenti. A kérdés az, hogy melyek azok a relációk (és normák), amelyek alapján különbséget tehetünk az egyes típusok között. A KOS-ok leírásához pontosan definiált relációkat kell használnunk, amelyeket korábban már bemutatunk, és most felsoroluk az alábbi táblában.

KOS-relációk listája, 2.

| | |
|---|--|
| R_1 lexikografikus rendezési reláció | R_{10} címkegyakorisági függvény |
| R_2 ekvivalenciareláció | R_{11} címke-címke együttjárési reláció |
| R_3 hierarchikus alárendeltje (tartalmazás) reláció | R_{12} felhasználó-címke együttjárési reláció |
| R_4 különbözőségi reláció | R_{13} felhasználó-dokumentum együttjárési reláció |
| R_5 generikus alá- és fölérendeltje relációpár | R_{14} felhasználói aktivitás függvény |
| R_6 partitív alá- és fölérendeltje relációpár | R_{15} dokumentumcímkézési gyakorisági függvény |
| R_7 következménye-előzménye relációpár | R_{16} tetszőleges relációfogalom |
| R_8 rokona (egyéb) reláció | R_{17} tetszőleges osztályfogalom |
| R_9 lásd és helyette szinonimareláció-pár | R_{18} tetszőleges attribútumfogalom |

A bemutatott keretrendszer és a fenti relációk alapján az egyes tudásszervezési rendszereket a következő relációk segítségével definiálhatjuk.

2.15.2 Terminuslista

A *terminuslistáknak* van a legegyszerűbb szerkezetük, mert ott a tartóhalmaz elemein csak egy *lexikografikus rendezést* engedünk meg. Ilyenek a könyvek végén található indexek (név- és tárgymutatók), amelyek a könyv legfontosabb kulcsszavait sorolják fel ábécé szerint, minden kulcsszóhoz hozzárendelve az oldalszámokat, ahol a kulcsszavak a szövegben előfordulnak. De ide tartoznak a különféle egységesített névlisták, sőt a keresőmotorokban használt 'invertált index' technológiája is.



L

Licensed Practical Nurse, 46
 Living trusts, 72–73, 80
 Living will, 87–89, 230
 Long-term care insurance policies,
 59–62

M

Massage, 206–207
 Meals and feeding, 192–93—
 Alzheimer's patients, 194
 appetite, 194–95, 238, 240
 eating aids, 127–28
 solutions to problems, 195–96
 see also Nutrition and diet
 Medicaid, 12, 13, 15, 30, 56–58—
 eligibility, 57, 79–80
 services not covered, 57
 waiver programs, 56
 Medicare, 13, 16, 22, 53–55—
 claims disputes, 55
 eligibility, 53
 hospice coverage, 311
 what is covered, 20–21, 22, 30, 34,
 41–42, 54, 119
 what is not covered, 53, 54
 Medication, 26–29, 30–31—
 cautions, 144–45
 directions for taking, 28
 drug interaction, 28
 generic drugs, 27, 28
 label abbreviations, 144
 over-the-counter drugs, 28
 overdoses, 29
 record keeping, 143–44

Medigap (Medicare supplemental) insur-
 ance, 55
 Mental health considerations, 31—
 Alzheimer's, 288–305
 boredom, 219
 dementia, 288
 depression, 217–18, 233–34
 fear of aging, 218
 Moving patients—
 helping to stand, 279
 in bed, 270–77
 transfer to and from wheelchair,
 280–85
 using mechanical lift, 278–79

N

Nail care, 164, 170
 Nursing homes, 4, 11–14
 Nutrition and diet—
 fluid intake, 242
 food safety, 239
 increasing calories and protein
 intake, 240–41
 nutritional guidelines for elderly, 240
 prevention of bone loss, 242–43
 sources of nutrients, 243–44

O

Occupational therapy, 204
 Odor prevention, 187–88
 Older Americans Act, 59

P

Pain management, 234—
 reduction techniques, 224–26
 types of pain, 224

tárgyszómutató mint terminuslista egy könyvből

Mivel a terminuslisták elemei között csak kevés relációt vesznek fel, igazából csak arra lehet használni őket, hogy egyfajta leltárként funkcionáljanak, mutassák, hogy az adott rendszeren belül milyen terminusok, nyelvi egységek fordulnak elő. Ehhez persze lehet további információkat is hozzárendelni (mint a fenti ábrán látható tárgyszómutató esetében), de ez a hozzárendelés már olyan hozzáadott értéként tekinthető, amely a terminuslista használatából fakad, nem pedig magából az alaprendszerből. Ettől még érdemes megbecsülni ezt az egyszerűséget, mert sok esetben komoly előny származhat abból, ha ilyen terminuslistákat „nyilvántartási” célokra használunk. **A szűkebben értelmezett tulajdonnévterek is terminuslistának minősíthetők.**

2.15.2.1 Formális jellemzés

A terminuslista a legegyszerűbb struktúrával rendelkező tudásszervezési rendszer. Ez a lapos struktúra nem kínál túl sok funkciót, viszont a rendszer nagyon egyszerűen kezelhető. A terminuslista formulája a következőpp írható fel.

$$KOS_{tlist} = \langle D, R_1, R_2, S_1, S_4, S_5 \rangle, \text{ ahol}$$

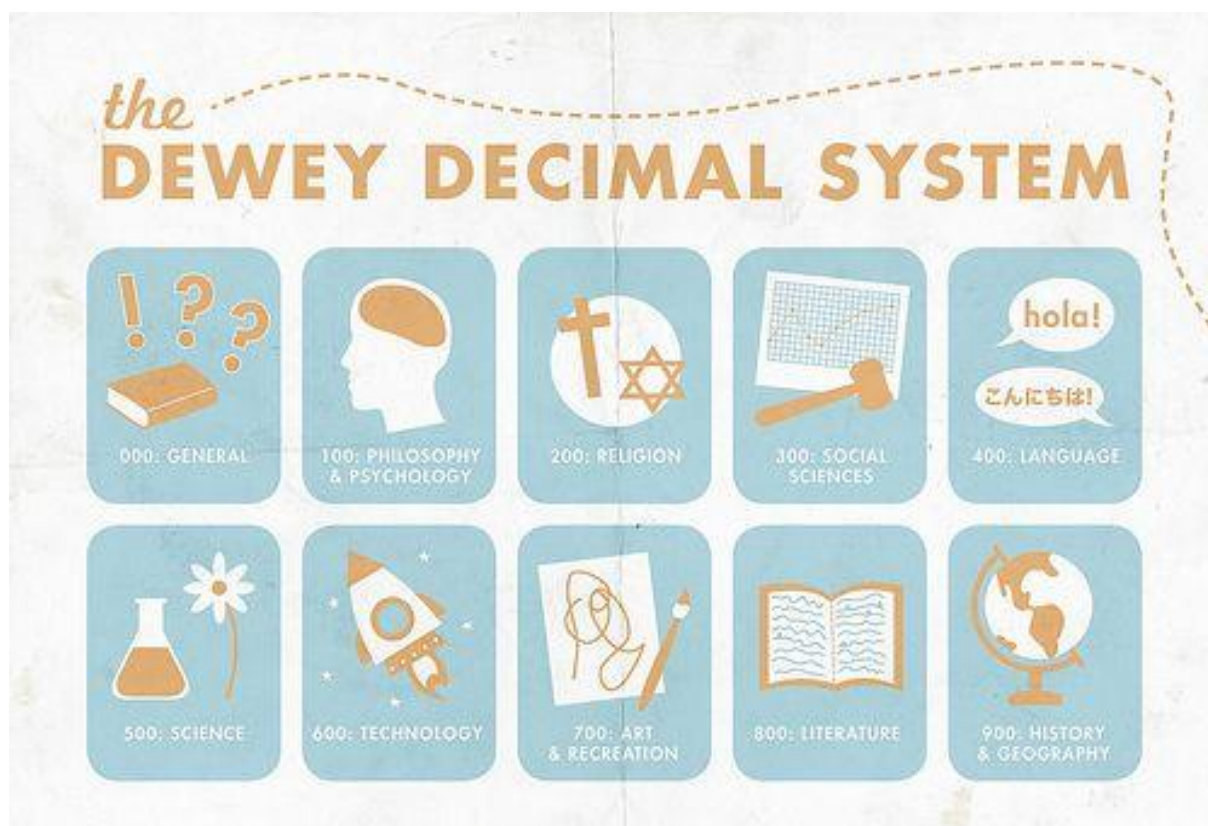
R_1 lexikografikus rendezési reláció
 R_2 ekvivalenciareláció

A terminuslista esetében az S_2 és az S_3 szabály nem érvényesíthető, hiszen ebben a rendszerben új relációt nem lehet definiálni. Az ilyen rendszernek a lexikografikus rendezés mellett van még egy másik relációja: a lista elemeit az R_2 ekvivalenciareláció kapcsolja össze. A földrajzi nevek egységesített adatbázisát például az kapcsolja össze egyetlen rendszerré, hogy minden tételéről azt állítjuk, hogy ekvivalensek egymással abban a tulajdonságukban, hogy valamennyien földrajzi entitások tulajdonnevei. Az ekvivalenciareláció fenntartásával azt kell „garantálnunk”, hogy a terminuslista elemei – az alkalmazott szempont szerint – azonos minőségűek lesznek (tehát nem keverednek különböző típusú elemek, mondjuk személynevek a földrajzi nevekkal). Más relációt nem lehet a terminuslista elemei közé felvenni. Attól függően, hogy milyen típusú terminusokról van szó, változhat az a gyakorlat, hogy fenntartják-e a szavak, kifejezések bekerülését szabályozó S_1 normát. Az igazán komolyan vett két előírás a katalógusépítésre vonatkozik, ezek a tárgyszó és a dokumentum összekapcsolását szabályozó S_4 és S_5 normák.

2.15.3 Taxonómia

Elmélyült elemzések segítségével különbséget lehetne tenni a taxonómia, az osztályozási rendszer, a klasszifikáció fogalmai között, de itt nem szükség figyelembe vennünk ezeket a finom eltéréseket, így a továbbiakban azonos jelentéssel használjuk ezeket a kategóriákat.

A klasszifikációs rendszereknek a klasszikus példája a könyvtárak világában használt ETO, az Egységes Tizedes Osztályozási rendszer (UDC, *Universal Decimal Classification*), vagy annak őse a DDC, a Dewey Decimal Classification system. A könyvtárakban a kezdetektől fogva elemi igény volt arra, hogy a könyveket könnyen és gyorsan meg lehessen találni tartalmi leírások, címkék alapján. Ennek az igénynek a kielégítésére dolgozták ki és tartották fent a decimális osztályozási rendszereket, amelyek a könyvek tartalmának jellemzéséhez, leírásához adtak segítséget azzal, hogy egy kontrollált és hierarchikusan összeszervezett osztályozó rendszert kínáltak fel. Az ETO mindent átfogó információs (indexelő és kereső) nyelv, decimális jelzetelésű osztályozási rendszer, amely egyrészt a 10 főosztályból álló táblázatból (amelyek 10-10 osztályra, azok mindegyike 10-10 alosztályra, azok pedig 10-10 szakcsoportra oszthatók, fogalmaikhoz alosztások fűzhetők), másrészt az osztályozó fogalmak összekapcsolására szolgáló szabályokból áll.



a DDC legfelső szintű kategóriái

Elvileg tetszőleges mélységig le lehet menni a hierarchiában, ami nagyon pontos leírás lehetőségét kínálja, bár ez a másik oldalról komoly szerkesztési és fenntarthatósági kérdéseket is felvet.

Dewey has a rich notational structure

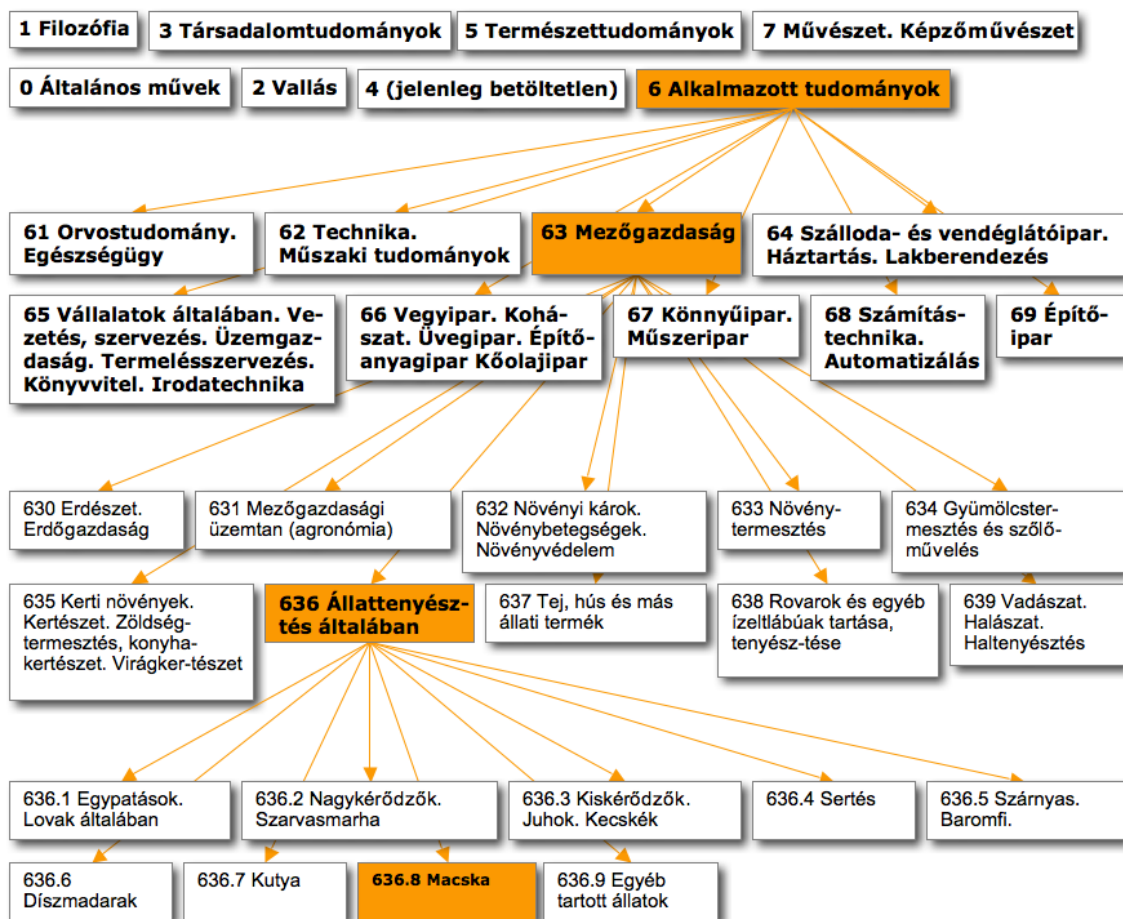
The Lancashire cotton industry : a study in economic development

Assigned DDC Code: 338.4767721094276



a DDC hierarchiájának alábontása több, mint húsz mélységű tételig

A hierarchikus szerveződés miatt nagyon sok osztályozási jelzetet lehet a rendszerbe felvenni, és a tizes osztályozási logika miatt gyorsan lehet benne keresni. Lehet olyan nézetben is megmutatni a rendszer egységeit, amelyek szemléletesek lehetnek a használóik számára, bár ez az elvi lehetőség a gyakorlatban sokszor nem áll rendelkezésre.



a 'macska' lexikai egység helye az ETO-ban

A taxonómiák népszerűsége az egyszerű kezelhetőségükben rejlik. Ez az egyszerűség persze viszonylagos. A terminuslistákhoz képest ugyanis itt már szemantikai elvárásokat kell figyelembe vennünk, hiszen a hierarchikus alárendelési reláció alkalmazása (akármi legyen is az értelme egy konkrét taxonómia esetében) szükségszerűen mindig szemantikai kényszerek betartásával jár együtt. Ezért van az, hogy ezen rendszerek használata esetében már megkövetelnek valamilyen szaktudást, és szabályozzák a munkafolyamat menetét is. Utóbbi mozzanat több részre osztható. A taxonómia elemeinek halmazát felfoghatjuk olyan *kontrollált szótárként* is, amelynek elemeit nem tetszőleges módon, hanem csak adott szabályokhoz igazodva, tehát csak kontrollált módon lehet bővíteni. Ez egyfelől korlátot jelent a tárgyszó-hozzárendelési munka során, mert előírásokhoz igazodó, tehát *fegyelmezett* munkavégzést követel meg az erre a feladatra előzetesen felkészített, *képzett archivátoroktól*, másrészt az ilyen rendszernek szüksége van egy olyan folyamatra, amely során a folyamatos változtatási igényeket ki lehet elégíteni, vagyis bővíteni, módosítani kell a rendszer valamely részét. Ez azt is jelenti, hogy a rendszer fenntarthatóságának érdekében szükség van *taxonómiaépítő szaktudásra*. Ezen feltételek, pontosabban az ezek teljesülésére vonatkozó kérdés azonban felvet két újabb, nagyon fontos tudásszociológiai, tudományfilozófiai kérdést:

- $Q_5 =$ {mennyire egységesen értelmezi az osztályozó közösség a taxonómia elemeit?}
- $Q_6 =$ {lehet egyetlen egységes rendszerbe rendezni valamely dokumentumgyűjteményt jellemző tudásterület fogalomkészletét?}

A kérdések megválaszolása szétfeszítené jelen tanulmány kereteit, ezért a részletes kifejtéstől el kell tekintenünk. A hierarchikus jelleg elvileg a gyors kereshetőséget támogatja, de azt meg kell itt jegyezni, hogy az ilyen struktúrák sok esetben merevekké válhatnak, és olykor a létrehozóik elfogultságai is visszatükröződnek bennük. Kiváló példát hozott erre Clay Shirky, aki a DDC elemzése közben vette

észre, hogy a történelmi régiók tárgyiszó alá sorolt címkék között „egy szinten” szerepelt Ázsia, Afrika és a Balkán félsziget (Shirky 2005).

D: History (general)

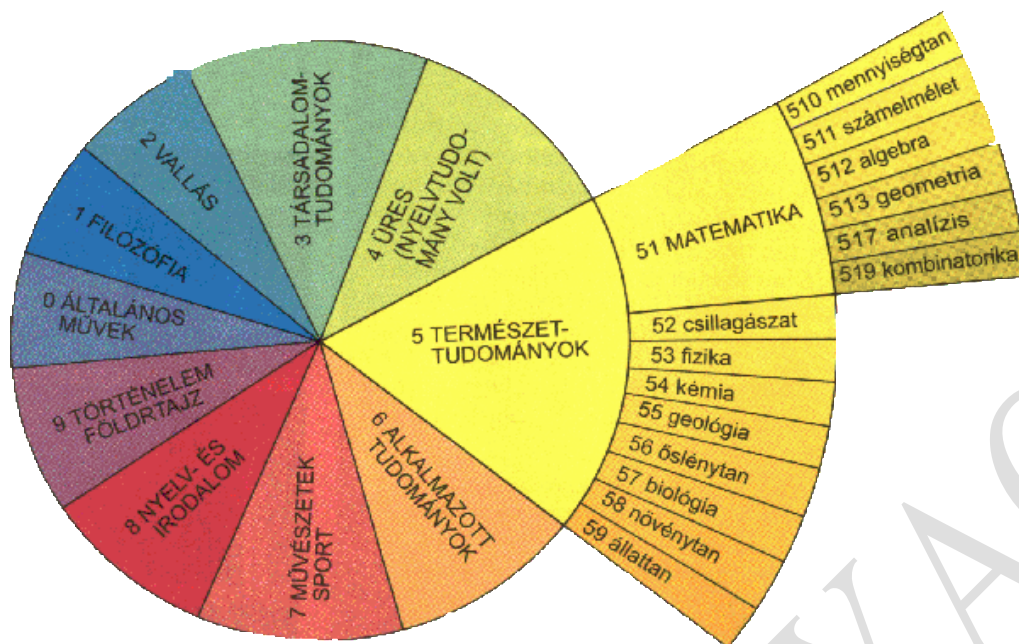
| | |
|-------------------|-------------------------|
| DA: Great Britain | DK: Former Soviet Union |
| DB: Austria | DL: Scandinavia |
| DC: France | DP: Iberian Peninsula |
| DD: Germany | DQ: Switzerland |
| DE: Mediterranean | DR: Balkan Peninsula |
| DF: Greece | DS: Asia |
| DG: Italy | DT: Africa |
| DH: Low Countries | DU: Oceania |
| DJ: Netherlands | DX: Gypsies |

Ez azért volt feltűnő, mert a méretét (és ezáltal kicsit a jelentőségét) tekintve a Balkán félsziget jóval kisebb a másik két kontinenshez képest. Ezt az aránytalanságot jelezte azzal Shirky, hogy bekarikázta a három területet egy földgömbön.



Clay Shirky ábrája a DDC kapcsán

Shirky végül rájött arra, mi lehet az oka ennek a furcsaságnak. Azért volt azonos szinten a három régióra mutató egység, mert nagyjából azonos számban adtak ki könyvet Amerikában a három földrajzi régióról, tehát a könyvtári polcokon elfoglalt helyigényük alapján voltak egyenrangú kategóriák a könyvtárosok számára. Akármennyire is jogos és elfogadható szempont ez a könyvtáros világ számára, ez a szempont nyilván elfogult. Ezek után pedig fel lehet tenni a kérdést, hogy vajon lehet-e elfogultság nélkül tudásszervezési rendszert építeni? A valószínűsíthető válasz pedig az, hogy nem nagyon. Minden tudásszervezési rendszernek lehet létjogosultsága egy adott tudásterületen, az arra jellemző elfogultságokat figyelembe véve, de bajos olyan rendszert építeni és feltételezni, amely univerzális ígérennyel léphetne fel, vagyis azzal a céllal, hogy minden tudásterületen, minden alkalmazási célra egyaránt felhasználható legyen.



az ETO rendszerének főbb kategóriái a természettudományokon belül

Az ETO folyamatosan karbantartott, klasszifikációs rendszer, amely alapvetően köznévtérnek minősíthető, még ha vannak benne tulajdonnevek is.

2.15.3.1 Formális jellemzés

A *taxonómiák* (más néven *osztályozási* vagy *klasszifikációs rendszerek*) már két fontos szemantikai relációt tartalmaznak (a lexikografikus rendezés – „kötelező” – szintaktikai relációján túl). Ez a tudásszervezési rendszer úgy van felépítve, hogy az elemei hierarchikusan egymás alá vannak rendelve – valamilyen *tartalmazási reláció* alapján. Ezt az alárendelési relációt lehet tiszta és pongyola értelmezésben is használni (a tiszta értelmezés esetben az alárendelési reláció a *generikus alárendeltje* relációval egyezik meg, a pongyola megközelítés keverten alkalmazza a generikus és a *partitív*, vagy más hierachikus relációt, például az *előzménye* viszonyt). A *hierarchia* leírásához azonban nem elégséges egyetlen relációt értelmeznünk a rendszeren, noha a közvélekedés gyakran megelégszik ezzel a megoldással. Arra is szükség van, hogy egy második relációval biztosítani lehessen azt, hogy az azonos felettes elem alá rendelt elemek különbözzenek egymástól, vagyis definiálni kell egy *különbözőségi relációt*.⁹ A taxonómia rendszerét a következőképpen írhatjuk le:

$$KOS_{tax} = \langle D, R_1, R_3, R_4, S_1, S_3, S_4, S_5 \rangle, \text{ ahol}$$

| | | | |
|-------|-----------------------|--------------|------------------------|
| R_1 | lexikografikus | rendezési | reláció |
| R_3 | hierarchikus | alárendeltje | (tartalmazási) reláció |
| R_4 | különbözőségi reláció | | |

Ebben a szisztémában az S_i szabályok közül már négyet érvényesítenek, csak az S_2 norma hiányzik, hiszen a két szemantikai kapcsolaton túl nincs mód más reláció alkalmazására a rendszerben, ebből következően pedig nincs is szükség az új relációk felvételét szabályozó normára.

⁹ Lehetne még erősebb feltételt is előírni és a JEPD-elv teljesülését elvárni. Ez annyival több a közvetlenül függő elemek különbözőségének elvárásától, hogy azt is megköveteli, hogy az azonos szinten levő fogalmak „együttes terjedelme” megegyezzen a fölöttes elem terjedelmével. A JEPD-elv (Jointly Exhaustive and Pairwise Disjoint) magyar fordítása 'együttesen kimerítő és kölcsönösen kizáró' lehetne. Bővebben lásd: [Bittner et al. 2004]

2.15.4 Tezaurusz

Ahogy a közgyűjtemények világán belül a taxonómiák a könyvtári világban jöttek létre először, ott terjedtek el széles körben, úgy a tezauruszok használata is ehhez az intézménytípushoz köthető. A tezauruszok iránti igényt a taxonómiák gyenge és könnyen inkonzisztenssé váló struktúrájával való elégedetlenség teremtette meg. Az osztályozási rendszerben csak egy lényegi relációtípust, a hierarchikus alá-fölrendeltséget lehet alkalmazni, és ez rákényszerítette, rákényszeríti a mindenkori felhasználókat arra, hogy minden hierarchikus relációt egyként kezeljenek e rendszeren belül. A rendszer elemei között pedig csak a hierarchikus láncok mentén lehetett kapcsolatot teremteni és keresgélni, ami könnyen merevvé tette az ilyen rendszereket. A tezauruszok megteremtésekor éppen e hátrányok leküzdése érdekében előre definiáltak több, pontosan meghatározott relációt, amelyek egyfelől a kapcsolatok sokkal pontosabb kifejezésére voltak alkalmasak, másfelől a rendszer elemei között több relációtípust is alkalmazva, többféle kapcsolati, bejárési útvonal alakulhatott ki, és ez jóval nagyobb rugalmasságot biztosított a tezauruszok számára.

A pontosabb szabályok szerinti működésmód igénye megkövetelte, hogy a tezauruszokra pontos és számon kérhető elvárásokat fogalmazzanak meg. Idővel a tezauruszok építésére vonatkozó szabályokat nemzetközi (és magyar) szabványokban rögzítették. A legfrissebb nemzetközi szabvány ISO 25964, amit két részletben publikáltak, 2011-ben és 2013-ban. A magyar szabványt 1987-ben adták ki (MSZ 3418-87 Magyar nyelvű információkereső tezauruszok szerkezete, részei és formái).

A tezauruszok tartalmas tételeit (vezér)deszkriptornak nevezik, amelyekhez a tezaurusz relációinak segítségével további tételeket lehet kapcsolni. A rendszerbe fel lehet venni nem-deszkriptorokat is, amelyek a szinonimitás kezelésére alkalmasak. A nem-deszkriptorok arra használhatók, hogy segítségükkel jelezzék, hogy a rendszeren belül van már olyan deszkriptor, amely be van kötve a tezaurusz struktúrájába, a nem-deszkriptornak így csak annyi a funkciója, hogy elvezessen a hozzá rendelt deszkriptorhoz, ahol további tartalmas továbblépési lehetőséget lehet találni.

A magyar tezaurusz szabványban az alábbi relációkat definiálták.

a magyar tezauruszszabvány által hdefiniált relációk

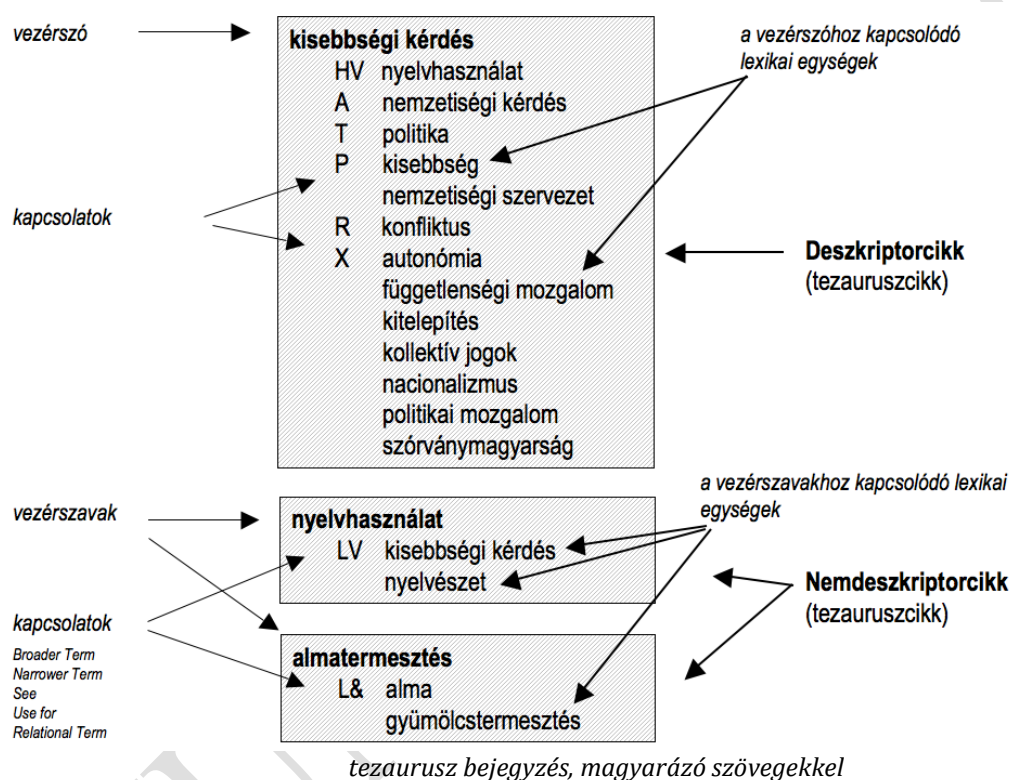
| | | |
|---|------------------------------|----------------|
| H | szinonimája (lásd, helyette) | R ₉ |
| F | fölérendeltje | R ₅ |
| A | alárendeltje | R ₅ |
| T | egésze (totum) | R ₆ |
| P | része (pars) | R ₆ |
| R | meghatározója (rezultáns) | R ₇ |
| E | meghatározója (előfeltétel) | R ₇ |
| X | rokonsági kapcsolat | R ₈ |

Ezeket a relációkat a következő módon határozzák meg a szabványban.

- H – szinonima reláció: a jelet követő nemdeszkriptor a deszkriptorcikk élén álló vezérdeszkriptort helyettesíti
- F/A – fölé- és alárendeltségi (generikus, nem-faj) reláció: a két fogalom egymásból lineárisan származtatható (valódi részhalmoz)
- F – fölérendeltje reláció: a mögötte levő fogalom fölérendeltje a vezérdeszkriptornak
- A – alárendeltje reláció: a mögötte levő fogalom alárendeltje a vezérdeszkriptornak
- T/P – egész-rész (totum-pars; szerkezeti) reláció: a fizikai, szerkezeti kapcsolat meghatározása
- T – egésze reláció: a mögötte levő fogalom egésze a vezérdeszkriptornak
- P – része reláció : a mögötte levő fogalom része a vezérdeszkriptornak

- R/E – rezultánsa és előfeltétele reláció:
- R – rezultánsa reláció: „a vezérdeszkriptor által jelölt tárgy, folyamat, stb. rendeltetése, okozata, eredménye, terméke, célja, tárgya, következménye (együttvéve: meghatározottja) a jelet követő deszkriptor által jelölt folyamat, tárgy, stb. (együttvéve: meghatározó).”
- E – előfeltétele, kiindulása reláció: a vezérdeszkriptor által jelölt folyamat, tárgy, stb. létének, létrehozásának, működésének, meghatározásának oka, előidézője, eszköze, alapja, kialakulása, azaz meghatározója.
- X – rokonsági reláció: más módon ki nem fejezhető kapcsolat, ellentét, hasonlat vagy személynév.

A szabványban rögzített elvárások alapján a teauruszok teauruszcikkekből, deszkriptorokból állnak, és ezek a tételek sokféle módon kapcsolódhatnak egymáshoz. Egy adott tétel több komponense a vezérdeszkriptor, amihez a relációk mentén vannak hozzárendelve a további tételek (amelyek lehetnek deszkriptorok vagy nem-deszkriptorok).



Ez az előny a kezdetektől arra a szerepre predesztinálta a teauruszt, hogy a könyvtári, archívumi világ legjobb, és ezért a legelterjedtebb tudásszervezési rendszerévé váljék, ez azonban nem valósult meg. A teaurusznak az osztályozási rendszerekkel kellett viaskodnia, de a taxonómia-teaurusz csatát már szinte az első pillanatban az előbbi nyerte meg. A csata kimenetelét a könnyebb kezelhetőség döntötte el.¹⁰ Mindez persze felveti azt, hogy egy alaposabb tárgyalás során figyelniük kellene arra a szempontra is, hogy milyen kötelezettségei vannak a katalogizálást, metaadat-hozzárendelést végző embernek.

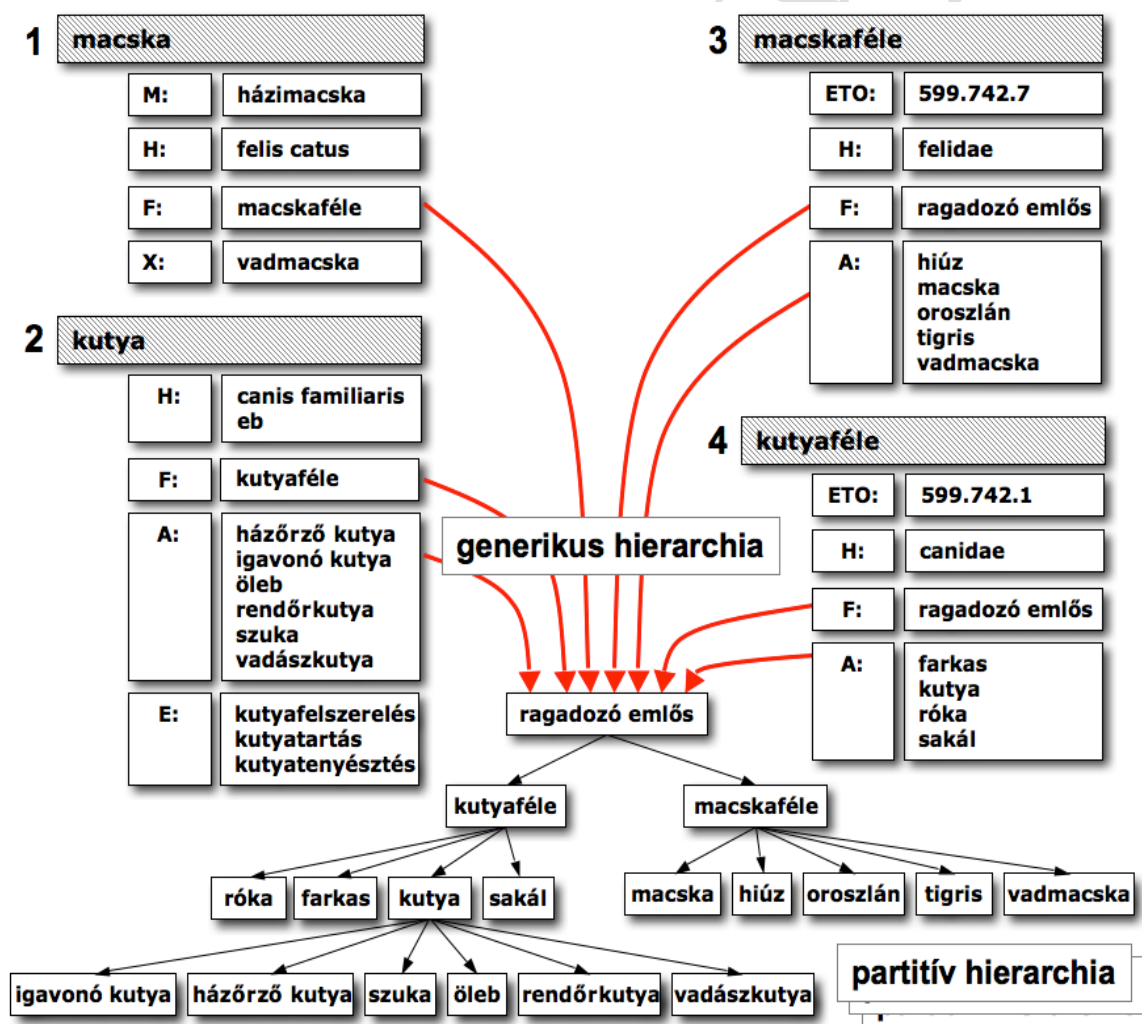
Ha a taxonómiák építéséhez kontrollált szótárra, kompetenciafeltételek fenntartására, fegyelemre, a munkafolyamatok ellenőrzésére van szükség, akkor ez még inkább így van a teauruszok esetében, hiszen ott jóval bonyolultabb struktúrát kell fenntartani, több szempontra kell figyelni, és nagyobb szaktudást kell elvárni a rendszert építőktől. Két – együttesen nehezen vagy egyáltalán nem teljesíthető – elvárás áll itt szemben egymással. Minél inkább szakterületi kérdésről van szó, annál megbízhatóbbnak lehet tartani a szakteauruszok (szaktaxonómiák) használatát az adott tudásterület leírásában, de annál inkább szükség van szakképzett, fegyelmezett, és ezért drága munkaerő

¹⁰ ETO-jelzet alapján nem csak a művet, hanem a példányt is meglehetősen találni: a szakkatalógusban kinézett 943.9"1848" : 323.1 jelzetű könyvért csak oda kellett menni a 9-essel kezdődő polcokhoz, az ETO jelzet magára a könyvre is rákerülhetett. Az ETO-nak az analóg világban ilyen gyakorlati haszna is volt.

alkalmazására is. Ha pedig valami sokba kerül, akkor mindig felmerül a kérdés, hogy ki fogja hosszútávon megfizetni. A kontrollált rendszerek hanyatlásának magyarázatában ez a döntő mozzanat: egyre inkább az látszik, hogy a web kontextusában egyre kevésbé hajlandók pénzt áldozni erre. Annál is inkább nehéz megfizettetni a kontroll árát, mert az utóbbi években megjelent új jelenség sokak számára egy alternatív megoldás lehetőségét sejteti.

Bár a tezauszok elbuktak mind a hagyományos világban a taxonómiákkal folytatott versenyben, mind a hálózati világ újonnan megjelent KOS-rendszereivel szemben, a tezauszok szemantikailag pontos és gazdag belső struktúrája megmaradt vonzó és követendő példát jelentő alternatívának, és ezzel részben előképét jelentette a számítógéppel kezelt formális ontológiáknak, és kezdetektől fogva táplálta a hozzájuk fűzött reményeket.

A tezausz pontos relációi mentén ugyanis elméletileg mindig adva volt a lehetőség arra, hogy logikai következtetéseket lehessen végezni e relációk segítségével. Az alábbi ábrán azt mutatjuk be, hogy az egyes tezauszcikkekben rögzített generikus relációs kapcsolatok arra voltak alkalmasak, hogy ezek mentén következtetéseket lehessen levonni a különböző tárgyszavakkal leírt dokumentumokra vonatkozóan (például ha egy könyvhöz a kutya terminus volt hozzárendelve, akkor a generikus reláció mentén lehetett ajánlani olyan könyveket is, amelyek a kutyafélékről, a ragadozó emlősökről szóltak – ha esetleg a kutyákról nem találtak megfelelő könyveket – azon az alapon, hogy minden kutya kutyaféle, valamint minden kutyaféle ragadozó emlős).




tezauszcikkek közti kapcsolatok a generikus reláció mentén

Az ilyen következtetéseket a hagyományos világban az ember mindig képes volt elvégezni, de a humán erőforrással kapcsolatos kapacitási gondok pont ezen a területen jelentkeztek a legerősebben. A géppel

végzett következtetéseket pedig már a formális ontológiák segítségével próbálták, próbálják megoldani. A teaurusz túl bonyolult volt ahhoz, hogy emberekkel hatékonyan kezelni lehessen, viszont még nem volt elég bonyolult ahhoz, hogy a gépek érdemben használni tudják.

ID: 300007842

Record Type: [concept](#)

 **cable-stayed bridges** (<bridges by construction>, bridges (built works), ... Built Environment (hierarchy name))












Note: Bridges having a deck that is directly supported from towers or pylons by straight cables without vertical suspenders. Construction of cable-stayed bridges usually follows the cantilever method. After the tower is built, one cable and a section of the deck are constructed in each direction. Each section of the deck is pre-stressed before continuing. The process is repeated until the deck sections meet in the middle, where they are connected. The ends are anchored at the abutments.

Terms:












- cable-stayed bridges** (**preferred**,C,U,English-P,D,U,PN)
- cable-stayed bridge** (C,U,English,AD,U,SN)
- bridges, cable-stayed** (C,U,LC,English,UF,U,N)
- cable braced bridges** (C,U,English,UF,U,N)
- stayed bridges** (C,U,English,UF,U,N)
- cable-stayed girder bridges** (C,U,English,UF,U,U)
- stayed girder bridges** (C,U,English,UF,U,N)
- tuibruggen** (C,U,Dutch-P,D,U,U)
- tuibrug** (C,U,Dutch,AD,U,U)
- kabelbrug** (C,U,Dutch,UF,U,U)
- kabelbruggen** (C,U,Dutch,UF,U,U)
- Schrägseilbrücken** (C,U,German,D,PN)
- Schrägseilbrücke** (C,U,German-P,AD,SN)
- puentes de cables oblicuos** (C,U,Spanish-P,D,U,PN)
- punte de cables oblicuos** (C,U,Spanish,AD,U,SN)

Facet/Hierarchy Code: [V.RK](#)

Hierarchical Position:

-  Objects Facet
-  Built Environment (hierarchy name) (G)
-  Single Built Works (hierarchy name) (G)
-  single built works (built environment) (G)
-  <single built works by specific type> (G)
-  <single built works by function> (G)
-  transportation structures (G)
-  <transportation structures by form> (G)
-  bridges (built works) (G)
-  <bridges by construction> (G)
-  cable-stayed bridges (G)

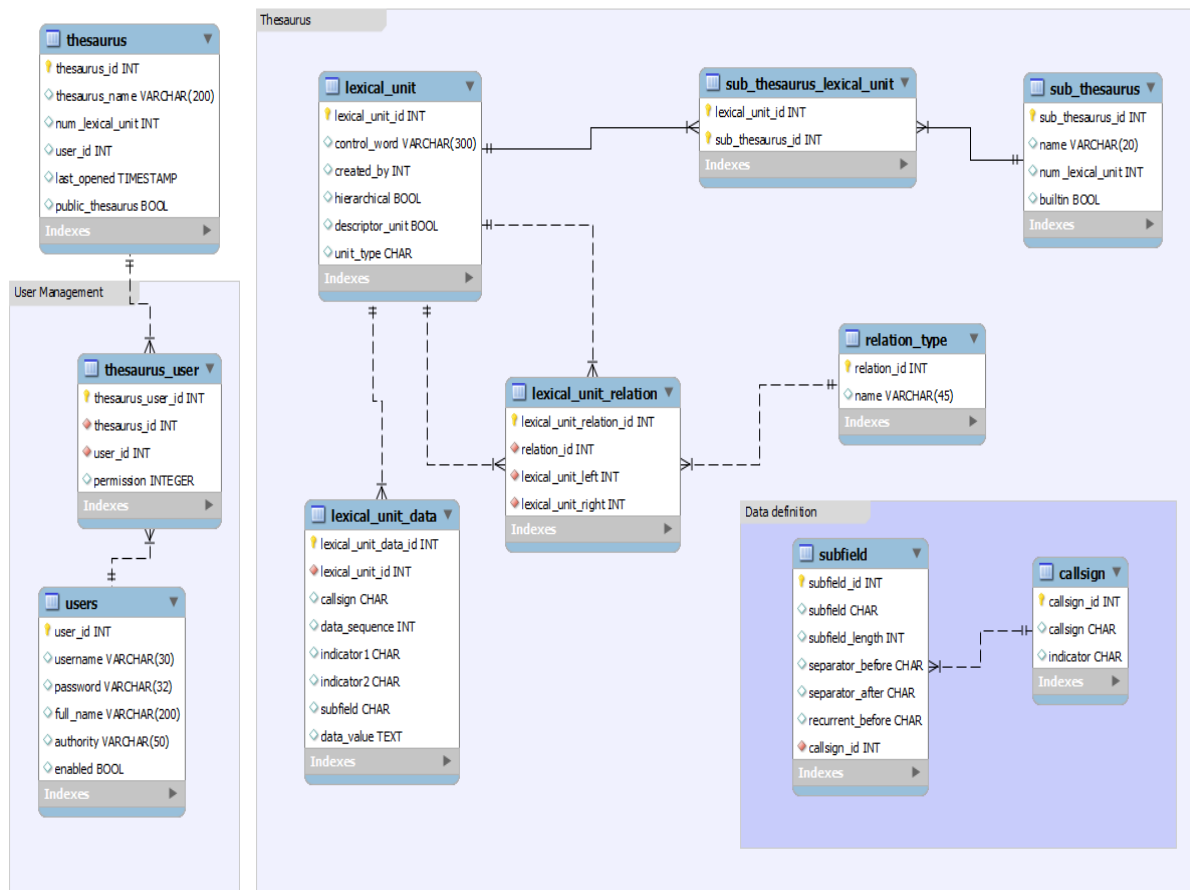
Additional Parents:

-  Objects Facet
-  Components (hierarchy name) (G)
-  components (objects parts) (G)
-  <components by specific context> (G)
-  architectural elements (G)
-  <structural elements and components for structural elements> (G)
-  structural elements (G)
-  structural systems (G)
-  cable structures (G)
-  cable-stayed structures (G)
-  cable-stayed bridges (G)

Additional Notes:

German Brücke mit einer Tragfläche, die mittels Kabel ohne vertikale Unterstützung direkt von Türmen oder Pylonen getragen wird. Der Bau einer Schrägseilbrücke erfolgt meist im Freivorbauverfahren. Nach Errichtung des Turms werden in jede Richtung jeweils ein Kabel und ein Segment der Tragfläche gebaut. Jeder Teil der Tragfläche wird vorgespannt, bevor mit dem weiteren Bau fortgefahren wird. Der Prozess wird wiederholt, bis sich die Tragflächenteile in der Mitte treffen, wo sie verbunden werden. Die Enden werden in den Stützpfählern verankert.

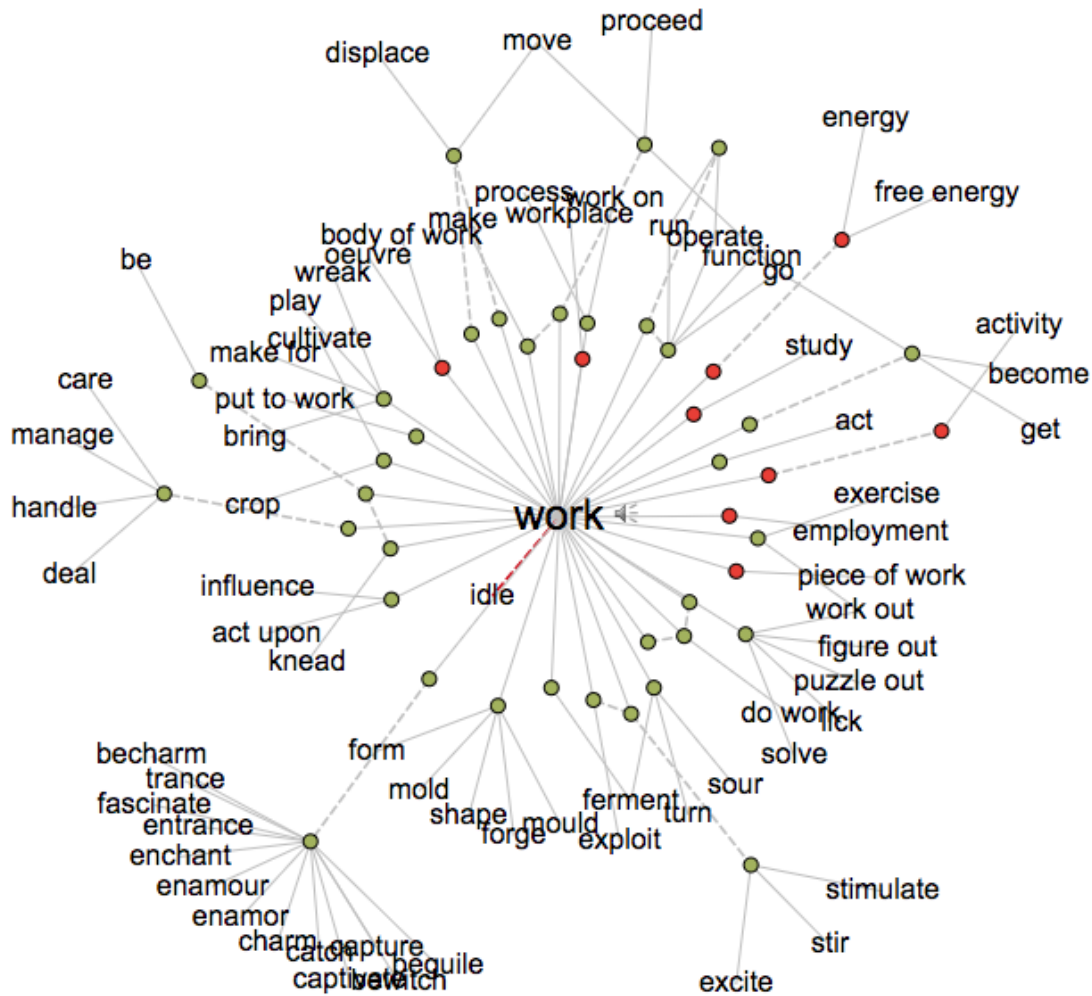
Getty Art & Architecture Thesaurus Online: 'cable-stayed bridges' tétel



a Rellex adatmodellje

A sémák áttekintése alapján látható, hogy az adatmodell központjában a tezaszuscikkek magját jelentő relációs tábla áll, amelyben a vezérszkriptor és a különböző relációk mentén hozzá kapcsolható lexikai egységek kapcsolatát rögzítik.

Mindig is adta magát, hogy a tezaszuszban rejlő gazdagabb struktúra valahogyan hasznosítani kellene az egységek és kapcsolataik megjelenítése során. Ezt az elvi lehetőséget, és régi kíváncsalmat akkor lehetett megvalósítani, amikor a tezaszuszokat is számítógéppel lehet kezelni, és a grafikus felületeken már el lehetett szakadni a hagyományos íráskultúra vizuális kereteitől. Részben ilyen „elszakadási” kísérletként is értékelhetjük az olyan kezdeményezéseket, amelyek a tezaszuszok egységeit, struktúráját a térben szabadabb módon, a lineáris írástechnikák korlátaitól elszakadva akarták megjeleníteni. Az egyik legrégebbi és legsikeresebb ilyen kísérlet a vizuális tezaszusz, amelynek felületén a kétdimenziós felületet szabadon és teljesebb módon kihasználva lehet feltüntetni a kiválasztott tételeket és a hozzájuk tartozó további egységeket. Ez a megoldás természetesen csak a felhasználói élményben okoz változást, a rendszer belső felépítése, működése ugyanolyan marad.



a visualthesaurus.com képernyője a 'work' terminusra

Mindezek után nézzük meg, formálisan hogyan jellemezhetjük a tezauszrt mint tudásszervezési rendszert.

2.15.4.1 Formális jellemzés

A széles körben használt tudásszervezési rendszerek közül a tezauszban van a legtöbb – többségében jól-definiált – reláció, így elméletileg ez tekinthető a legnagyobb és legpontosabb leíróerővel rendelkező megoldásnak.

$KOS_{tez} = \langle D, R_1, R_5, R_6, R_7, R_8, R_9, S_1, S_2, S_3, S_4, S_5 \rangle$, ahol

| | | | |
|-------|------------------------------------|-----------|--------------------------|
| R_1 | lexikografikus | rendezési | reláció |
| R_5 | generikus | alá- és | fölérendeltje relációpár |
| R_6 | partitív | alá- és | fölérendeltje relációpár |
| R_7 | következménye-előzménye | | relációpár |
| R_8 | rokona | (egyéb) | reláció |
| R_9 | lásd/helyette szinonimareláció-pár | | |

A teaurusz esetében az S_i szabályok mindegyikét be kell tartani (igaz, hogy az S_2 -es szabályt nem mindig, sőt gyakrabban nem érvényesítik, vagyis nem engedik, hogy új relációt lehessen felvenni a rendszerbe, de a formális modellbe mégis be kell emelnünk ezt a szabályt, mert előfordulhatnak olyan teauruszok is, amelyekben a szabványokban rögzített relációkhoz képest további relációt is definiálnak). Vannak azonban a formulában jelzett relációk, relációpárok, amelyek más minőséget adnak a teauruszoknak. Mivel a teaurusz több, pontosan definiált relációt tartalmaz a taxonómiához képest, ezért összetettebb struktúrát képezhetünk le vele, és a gazdagabb szemantika, a nagyobb kifejezőerő miatt sokkal pontosabban, rugalmasabban és megbízhatóbban lehet vele a dokumentumok tartalmát leírni.¹¹

2.15.5 Folkszonómia

Alvin Toffler 1980-ban írt könyvében már előrevetítette, hogy a termelés és fogyasztás hosszú időszakon keresztül egymástól elváló társadalmi szerepei újra közelebb kerülnek egymáshoz. A jelenség fontosságát azzal is hangsúlyozni akarta Toffler, hogy új terminust alkotott: a *prosumer* kifejezést [Toffler 2001]. Toffler nem látta, nem láthatta előre a hálózati kommunikáció web2-es trendjeit, de ettől még tény, hogy ezek a fejlemények messzemenően igazolták a toffleri jóslatot. Persze, ezt a jelenséget – a web2.0 kifejezésen túl – nagyon sokféle terminussal próbálták meg leírni, Toffler *prosumer* kategóriája csak egy a sok közül. A felhasználók által létrehozott tartalom (user generated content), *produser*, *social media*, *social web*, *peer production*, *crowdsourcing*, *collective intelligence*, *wikinómia*, *a tömegek bölcsessége* – sokáig lehetne sorolni a versengő kifejezéseket. Ha a navigáció támogatására szolgáló tudásszervezési rendszereket vizsgáljuk, akkor a megfelelő web2-es kategória a *folkszonómia* (folksonomy) fogalma.

Maga a 'folkszonómia' terminus Thomas Vander Wal egyik blogbejegyzésében jelent meg először 2004-ben (azóta kétszer, két helyen is megpróbálkozott a fogalom definiálásával (Vander Wal 2007, 2005b), de érdemes megemlíteni azt a tényt, hogy az *etnoklasszifikáció* kifejezéssel Susan Leigh Star már 1996-ban nagyon hasonló értelmű fogalmat hozott létre (Star 1996). Ezeket a rendszereket gyakran *címkézési rendszereként* (tagging system, tagsonomy), olykor *közösségi címkézési rendszerként* is emlegetik, magát a metaadat-kezelési tevékenységet pedig címkézésnek hívják, de használnak még további neveket is, mint például *grassroots classification*, *cooperative classification* vagy *distributed classification system*, *folk classification*, *folk taxonomy*, *social classification*, *open tagging*, *free tagging*, *faceted hierarchy*.

Az új terminust Vander Wal a 'folk' és 'taxonomy' szavak összevonásából gyártotta. Azt akarta kifejezni vele, hogy a tárgyszavazási tevékenységet nem csak az erre felhatalmazott, kontrollált tárgyszórendszert kezelő szakember, hanem bárki (a „nép” bármely tagja) végezheti. Tartalmilag nem volt szerencsés erre a taxonómia kifejezést ráhúzni, mert a folkszonómiák nem hogy egy osztályozási rendszert, de semmilyen más struktúrát nem képesek felépíteni, fenntartani. Ennek pedig pont az az oka, amit az új kifejezés előtagja fejez ki, hogy ti. ha bárki felvehet új címkét, bárki bármilyen címkét a dokumentumokhoz rendelhet, akkor semmi sem garantálja az , hogy a címkék között konzisztens kapcsolat alakulhasson ki (és maradjon fenn).

A folkszonómia nagyon lapos struktúrának tekinthető, csak abban különbözik az egyszerű terminuslistáktól, hogy a címkék gyakorisági értékeire lehet relevanciakezelést építeni. A folkszonómiák alapján épített címkefelhők a címkék méreteivel, forgalmukkal kapcsolatos számosságával jelzik azt, hogy a rendszeren belül melyikek a közösség által fontosabbnak ítélt (mert gyakrabban használt) címkék. Ez a gyakoriságjelzés mint relevanciakezelő kritérium sok esetben valóban használható, de olyan is előfordul, nem is kevésszer, hogy ez a megoldás gyakorlatilag használhatatlan. Erre később mutatunk példákat.

¹¹ A teaurusz R_5 , R_6 , R_7 relációi egyenként" mind hierarchikus struktúrát biztosítanak, ezért ezek úgy tekinthetők, mint amelyek egy-egy alosztályát adják a taxonómiák R_3 'hierarchikus alárendeltje' relációjának. Ez azt (is) jelenti, hogy ha e három teauruszreláció mentén összekapcsolt elemeket egyetlen – általánosabb – relációval kötnénk össze, akkor az R_3 relációval képezhető struktúrához nagyon hasonló képződményt kapnánk. A teauruszban tehát egy „rejtett” taxonómia is található.

A folkszonómiák hatalmas előnye a rendszer könnyű kezelhetősége. Nem követel meg semmilyen feltételt a címkét kiosztó felhasználótól a rendszer a címkék használatakor, szabadon lehet új címkét felvenni, régit használni. Ebből persze többféle probléma is előállhat. Ha nincs semmi szabály, akkor ugyanazt a tárgyszót fel lehet venni egyes számban, de többes számban (lány, lányok, vagy ugyanazt a jelentésű tartalmat fel lehet venni más szófajú (főnévi, melléknévi vagy igei) alakban (futok, futás, futó), vagy alkalmazni lehet szinonim jelentésű tárgyszavakat (kutya, eb), és ezek mind külön címkéknek fognak megfelelni. A folkszonómiák védelmében szokták felhozni, hogy a hibás, „nem helyes” alakú címkék kisebb számban fordulnak majd elő, és a gyakoriságuk alapján majd nem lesznek láthatók, így nem is fognak zavarni. Ez azonban nem feltétlen igaz. Az Instagram hashtag-jei kapcsán a későbbiekben bemutatjuk azt, hogy ez effajta szűrés nem feltétlen működik minden esetben eredményesen.

A folkszonómiák esetében további komoly problémáját jelenti a *kontextuskötöttség*, ami alatt azt a jellegzetességet értjük, hogy a felhasználók gyakran olyan címkéket rendelnek a dokumentumokhoz, amelyeknek csak a saját kontextusokban van igazi referenciája, és annak ismerete nélkül semmilyen értelemben nem ad tartalmi eligazítást a címke. A klasszikus példája ennek az, amikor a felhasználók a 'me' címkét ragasztják a – mondjuk – az általuk feltöltött képekhez, de ugyanilyen semmitmondó a 'selfie' (és még sok minden más, hasonló) címke is.

A jelenség első híres példája a del.icio.us webcímen működő közösségi címkézési rendszer volt. Azóta a kép- és videómegosztók, a mikroblogok és a közösségi oldalak is ezt a technikát használják: a Twitter vagy a Facebook hashtagjei, a Flickr, Instagram vagy a YouTube címkézési gyakorlata is folkszonómiaként minősíthetők.

advertising ajax apple architecture art article audio blog blogging blogs book books business community computer cool css culture daily design development diy download education email english entertainment fashion fic finance firefox flash food for forum framework free freeware fun funny furniture gallery game games google graphics green gtd hardware health history home hosting house howto html humor icons imported inspiration interesting internet iphone java javascript jobs language learning library lifehacks linux list mac magazine management maps marketing media mobile movies mp3 music news online opensource osx photo photography photos photoshop php plugin politics productivity programming psychology python radio rails recipe recipes reference research resources rss ruby rubyonrails science search security seo sga shop shopping social software tech technology tips tool tools toread travel tutorial tutorials tv ubuntu video videos web web2.0 webdesign webdev wiki windows wordpress work writing xml youtube

a del.icio.us címkefelhője

A del.icio.us címkefelhője műfajt teremtett, az „utánzó” többféle módon is igyekeztek kihasználni a műfajban benne rejlő adottságokat. Voltak, akik a látvány fele mozdultak el, és a címkefelhőt „önálló” képként kezelték.

Trends for you · [Change](#)

Budapest
4,818 Tweets

#mustread
8,544 Tweets

#feedly
3,145 Tweets

HRK Game

#TheVoiceKids
108K Tweets

Gabriel
478K Tweets

#ColombiaDecide 🇨🇴
30.3K Tweets

#CambioDeMando
67.8K Tweets

#اسوا_بنك_واجته
22.3K Tweets

#SelfiesForAriana
27.5K Tweets

a Twitter személyes hashtag-ajánlata

Persze, a hashtagelés nem mindig, nem feltétlenül nyújt segítséget az eligazodásban, kapcsolódásban. A bal oldalon látható az a címkeajánlat, amit a Twitter kínált fel egy olyan felhasználónak, aki hosszú évek inaktivitása után belépett újra a szolgáltatás felületére. A 'Budapest' címkén kívül semmilyen értékes, értékelhető címkét nem láthatunk, a többségüknél talán még a véletlenszerű kiosztás is több relevanciát nyújthat.

Az ajánlott címkék listájában jól látszik a folkszonómiák kontextuskötöttsége, felszínes tartalmassága. A '#mustread' címke csak annyit képes üzeni nekünk, hogy fontos, de persze ez is csak akkor lenne számunkra, ha a címkét kiosztó személy és köztünk valamilyen értékközösség lenne, aminek jelzésére ez a hashtag nem alkalmas. A kontextuskötöttség mint hátrány a személyes névmások ('me', 'my mom'), a személynevek ('Gabriel') gyakori alkalmazásában érhető tetten, amikor tudjuk, hogy a címke referenciája csak a címkéző kontextusában határozható meg pontosan, annak ismerete nélkül a címke semmire sem mutat.

A hashtagek tartalmatlanságát szemléltethetjük még az Instagram képmegosztó száz legnépszerűbb címkéjével. A listában vannak olyan címkék, amelyek bírnak némi tartalmi eligazító erővel (dog, food, makeup, sky, flower, cat, gym), de ezek vannak kevesebben. A kontroll hiánya megmutatkozik abban is, hogy nem lehetséges a szintaktikai egységesítés sem. A címkék felbukkanhatnak egyes és többes számban vagy szinonim alakban, és ilyen esetekben különböző címkéknek számítanak, példa lehet erre a girl (17.) és a girls (59.) vagy a picoftheday (13.) és a photoftheday (3.) címke is. Sok olyan címke van a listában, amely teljes mértékben kontextusfüggő, érdemi tartalommal nem rendelkezik, például a my (85.), a me (7.), a family (26.), a selfie (12.), followme (10.) vagy a baby (63.). A népszerű címkék nagy része pedig valamilyen állapotot vagy értékelést jelöl, ami megint csak kevés eligazító erővel bír, például a fun (18), a funny(66.), beautiful (4.), happy (8.) vagy a love (1.). Vannak olyan címkék is, amelyek valamilyen divatjelenség miatt kerülhettek be a topszázba, vagy egy közösség szlengjeként funkcionálhatnak, és a kontextus ismerete nélkül nem is lehet tudni a jelentésüket: f4f (46.), lfl (48.), tbt (4.), vsco (50.), ootd (62.).

| | | | |
|--------------------|-----------------|-------------|------------------|
| 100. #birthday | 75. #black | 50. #vsco | 25. #art |
| 99. #red | 74. #gym | 49. #dog | 24. #food |
| 98. #workout | 73. #motivation | 48. #lfl | 23. #instalike |
| 97. #sweet | 72. #yummy | 47. #pretty | 22. #igers |
| 96. #wedding | 71. #healthy | 46. #f4f | 21. #repost |
| 95. #blackandwhite | 70. #instacool | 45. #travel | 20. #smile |
| 94. #instalove | 69. #hot | 44. #sky | 19. #tagforlikes |
| 93. #fit | 68. #iphoneonly | 43. #music | 18. #fun |

| | | | |
|------------------|----------------|----------------------|-------------------|
| 92. #instafood | 67. #instapic | 42. #swag | 17. #girl |
| 91. #instasize | 66. #funny | 41. #followforfollow | 16. #instadaily |
| 90. #flowers | 65. #instagram | 40. #beach | 15. #friends |
| 89. #iphonesia | 64. #night | 39. #beauty | 14. #summer |
| 88. #tweetgram | 63. #baby | 38. #sun | 13. #picoftheday |
| 87. #design | 62. #ootd | 37. #vscocam | 12. #selfie |
| 86. #instafollow | 61. #makeup | 36. #bestoftheday | 11. #follow |
| 85. #my | 60. #cat | 35. #fitness | 10. #followme |
| 84. #webstagram | 59. #girls | 34. #life | 9. #fashion |
| 83. #followback | 58. #cool | 33. #amazing | 8. #happy |
| 82. #christmas | 57. #lol | 32. #follow4follow | 7. #me |
| 81. #work | 56. #party | 31. #nofilter | 6. #cute |
| 80. #blue | 55. #foodporn | 30. #style | 5. #beautiful |
| 79. #instacool | 54. #tflers | 29. #instagood | 4. #tbt |
| 78. #work | 53. #hair | 28. #nature | 3. #photooftheday |
| 77. #blue | 52. #photo | 27. #likeforlike | 2. #instagood |
| 76. #pink | 51. #sunset | 26. #family | 1. #love |

az Instagram 100 legnépszerűbb hashtagje (2018)

Bár azt egyáltalán nem mondhatjuk, hogy a folkszonómiák tudatos tervezés eredményeként jöttek volna létre, kialakulásukat és elterjedésüket mégis a taxonómiákkal való összehasonlítás révén érthetjük meg. Ahogy azt az osztályozási rendszerek leírásakor jeleztük, hamar kiderült, hogy hálózati környezetben ezek a tudásszervezési rendszerek sikertelennek bizonyultak, mert rugalmatlanok voltak, kiépítésük és használatuk nehéznek és költségesnek bizonyult. Jóval rugalmasabb és egyszerűbb, a felhasználók számára elfogadhatóbb, használhatóbb megoldásra volt szükség. Ezt hozták el a folkszonómiák, amelyeket épp ezért minősíthetünk felhasználóbarát tudásszervezési rendszereknek is.

A folkszonómiák használatakor ahelyett, hogy a dokumentumokat egy hierarchikus struktúrájú osztályozási rendszer elemeivel próbálnánk meg leírni, egyszerűen címkéket ragasztunk rájuk úgy, hogy a címkék formátumára, jelentésére semmilyen megkötés nem teszünk. A címke kifejezheti a dokumentum jellemző tulajdonságát ('film', 'magyar', 'dokumentumfilm'), lehet emlékeztetőként használni ('megnézni', 'fontos', 'tanulmányhoz'), vagy utalhat akár a dokumentumhoz fűződő érzelmekre is ('tetszett', 'izgalmas', 'hatásvadász').

Mivel nincs semmilyen központi előírás, szabályozás, ezért a dokumentumok címkézését maguk a felhasználók végezhetik. Emiatt persze a címkézés erősen szubjektív lesz. Ez az elfogultság természetesen a címkézést végző felhasználók gondolkodásmódját tükrözi. A szubjektivitás hatása csökkenthető, ha az archívum felhasználóit érdekeltté tesszük abban, hogy a már mások által felcímkézett dokumentumokat maguk is ellássák új címkékkel (Speroni 2006). Minél több felhasználó címkézi a tartalmakat, annál több és relevánsabb címke fog rákerülni egy dokumentumra, megkönnyítve annak megtalálását. Ha ez sikerül, akkor a sokak által címkézett dokumentumokon nagy eséllyel lesz néhány olyan, amelyet kiugróan sokan ragasztottak rájuk. Ezek lesznek a dokumentumra valamilyen szempontból jellemző tulajdonságok. És persze akad majd nagyon sok olyan címke, amelyet csak a címkézők töredéke ragasztott fel. Ezek lesznek azok a tulajdonságok, amelyek csak kevesek számára fontosak, illetve ebbe a csoportba kerülnek a személyes érzelmek, emlékeztetők is. Adam Mathes már

az első folkszonómiák megjelenésekor megállapította, hogy a címkék eloszlása hatványfüggvény jellegűt követ (Mathes 2004). A címkék kisebbik hányadát nagyon sokszor rendelik hozzá a dokumentumokhoz, míg a címkék túlnyomó részét jóval ritkábban (legtöbb esetben nagyon kevés alkalommal) használják.

Mindkét fajta címke nagyon jól használható kereséskor. Az *erős címkék* a keresési eredmények rangsorolásakor nyújtanak nagy segítséget. Például egy sportkocsikat bemutató film esetén jellemző lehet az 'autó', és a 'film' címke. Ha ezen kulcsszavak alapján keresnénk dokumentumokat, akkor szeretnénk, ha ez a film a találatok között előkelő helyet foglalna el, és ehhez hatalmas segítséget jelent az erős címkék jelenléte. A *gyenge címkék*, amelyek az ún. „hosszú farok” részről származnak (Vander Wal 2005a), biztosíthatják azt, hogy az adott dokumentum a legváltozatosabb keresési feltételek esetén, különféle gondolkodásmódokhoz igazodóan is, megtalálható legyen. A fenti sportkocsis film példájánál maradván ilyen címke lehet például a 'nyolchengeres', 'izgalmas', 'sebességváltó'.

A dokumentumokhoz való hozzáférés is a címkék közti kulcsszavas keresésen alapul: a rendszer azokat a tartalmakat fogja megjeleníteni, amelyek a keresett címkét viselik. Kereshetünk az összes dokumentum, illetve csak az általunk megjelölték között, így amilyen könnyen megtalálunk egy 'ház'-at ábrázoló képet, amit sokan 'szép'-nek tartottak, olyan könnyen visszakereshető a mi 'ház'-unkról készült kép, vagy egy olyan, amit mi gondoltunk 'szép'-nek. Ezért a magunk számára érdemes megjelölnünk egy dokumentumot akár egy olyan, mások számára semmitmondó címkével is, mint például az 'elintéznivaló'. A címkézés azért működik, mert – a taxonómiába való besorolással szemben – a végtelenségig leegyszerűsített, intuitív művelet, és ezért a felhasználók hajlandóak használni. Sőt, nem csak hogy hajlandóak rá, de motiváltak is, hogy maguknak is felcímkézzék a számukra fontos dokumentumokat – hiszen azok később a saját fogalmi rendszerüket reprezentáló címkéken keresztül sokkal könnyebben megtalálhatóak lesznek.

A HP Lab két kutatója a Del.icio.us címkézési gyakorlatának adatait elemezve több szempontból is jellemezni próbálta a felhasználók aktivitását (Golder & Huberman 2006). Az elemzés során tipizálták a címkéket is, és az alábbi csoportokat (pontosabban címkézési funkciókat) különítették el egymástól.

- azonosítani, kiről vagy miről van szó (köznevek vagy tulajdonnevek, amelyek általános és egyes, illetve absztrakt és konkrét dolgokat, objektumokat, eseményeket stb. jelölnek: 'dog', 'computer')
- azonosítani, milyen fajtáról, típusról van szó (köznevek, melyek valamilyen szinten általánosítva jelzik azt, hogy milyen dolgokról van szó: 'article', 'blog', 'book' – általában az előző címketípus megadásával párhuzamosan)
- azonosítani, kinek a címkéjéről van szó (ki „birtokolja” a címkét: 'Peter')
- pontosítani a kategóriákat (már megadott címkék terjedelmét, pontosságát jelzőkkel, számokkal, értékekkel pontosítani: '12,500')
- értékelni a dolgokat (jelzőkkel kifejezni a felhasználó véleményét: 'scary', 'funny', 'stupid')
- önhivatkozás (kifejezni a személyes kapcsolatot: 'mystuff', 'mycomment')
- teendőket szervezni (utalás elvégzendő feladatokra: 'todo', 'toread')

Ezek a címketípusok abban az értelemben nem összevethetők, hogy különböző célokra lehet/érdemes őket használni, más tehát a funkciójuk. Később kitérünk a folkszonómiák gyakorlatával szemben felvethető kifogásokra is, előbb azonban nézzük meg a címkézési rendszerek előnyeit. A címkézés ugyanis számos előnnyel jár. A hagyományos taxonómia alapú archívumok esetén két dokumentumot akkor mondhatunk hasonlóknak, ha nagyjából egy kategóriába esnek a hierarchikus rendszerben. A hasonlóság fogalmát folkszonómia alapú archívum esetén sokkal természetesebben, ráadásul több szinten is értelmezhetjük. Címkék használata esetén két dokumentum akkor lesz hasonló, ha a rájuk ragasztott címkék között sok az átfedés. A hasonlóság természete ebben az esetben tehát nem a taxonómia megalkotójának fogalmi rendszere mentén nyilvánul meg, hanem azt a közösség által felállított címkék fogják meghatározni. Egy hagyományos fotóarchívumban ritkán kerülne egymás mellé egy piros sportkocsit és egy piros almát ábrázoló kép. Ez folkszonómia esetén teljesen természetes fogalmi társítás a 'piros' címkén keresztül. A hasonlóság ráadásul tovább általánosítható, hiszen nemcsak a dokumentumok között definiálhatjuk azt, hanem feltérképezhetők a gyakran együtt előforduló címkék is. Ezzel a rendszer felismerheti a szinonimákat, vagy az egy fogalomkörből származó szavakat, például az archívum valamiképp 'tudni' fogja, hogy a 'piros' és a 'vörös' címke hasonló dolgokat ír le. Ennek segítségével a rendszer megtalálja a valamiképpen kapcsolódó tartalmakat. És ezzel még mindig nincs

vége, hiszen az archívumból az is kideríthető, hogy kik azok a felhasználók, akik hasonló címkékkel láttak el dokumentumokat. Ők, szintén a hasonló fogalmi kör használata miatt valamiképp ugyanúgy gondolkodnak, azaz hasonlítanak egymáshoz. A 'hasonló' felhasználók könnyen megtalálhatják egymást, így nem csak a közösség építi a rendszert, hanem a rendszer is a közösséget. A címkézés segíthet megoldani az eltérő nyelvi környezetben élő felhasználók eltérő nyelvhasználatának problémáját is. A legnépszerűbb címkéket egyszerre megmutató címkefelhő a folkszonómiák egyik kedvenc szolgáltatása. Mivel a címkefelhő eligazító, navigációs ereje akkor „értékesül”, amikor a felhasználók ránéznek, könnyen szelektálni tudják a nem az általuk ismert nyelvből származó címkéket, ahogy ezt az alábbi ábrán láthatjuk.



többnyelvű elemek egy címkefelhőben

Az ábrán az angol címkék mellett szerepel néhány más (talán kínai) írásjel is, amelyeken az angolul olvasó felhasználó szeme könnyedén „átugrik”, és nem veszi figyelembe, illetve ugyanez igaz a kínai felhasználókra is, csak fordítva.

A címkék, a címkézés rugalmassága nagy előny. Ennek szemléltetésére hasonlítottunk össze a fotozz.hu szolgáltatását a Flickr.com által kínált lehetőségekkel. Mindkét rendszer képek feltöltését, megosztását, online fényképalbum létrehozását, valamint mások albumainak, képeinek böngészését, kereshetőségét biztosítja. A fotozz.hu szolgáltatás egy új kép feltöltésekor megköveteli, hogy a képet soroljuk be a megadott kategóriák egyikébe. A kategóriarendszer rögzített, nem lehet változtatni, bővíteni, és erősen fotószakmai elfogultságú kifejezésekből áll.

a fotozz.hu oldal legfontosabb kategóriái

| | | | |
|---------------------|---------------------|------------|--------------------|
| absztrakt | digit. illusztráció | légifotók | tárgyfotók |
| abszurd | divat | makró | természet |
| akt | életképek | panoráma | város, építészet |
| állatfotók | elkapott pillanatok | portré | vízalatti fotók |
| barlangfotók | glamour | riport | feldolgozott fotók |
| családi/emlékép | humor | sport | így készült |
| csendélet | infravörös fotók | szociofoto | egyéb |
| csillagászat/égbolt | koncert - színpad | tájkép | pályázat |

Lehet, hogy ez a kategóriarendszer fotószakmai szempontból teljes és kimerítő, de ezzel biztosan nem lehet a képeket kereső felhasználók szempontjait kielégíteni. Olyan keresési igény nehezen képzelhető el, amikor valaki 'pályázat' vagy 'így készült', netán 'egyéb' témakörben szeretne képeket találni magának. Ez a rendszer túl merev. A kép mellé a technikai metaadatokon kívül nem lehet mást felvenni, így kereséskor egyedül a képek kategóriáját lehet kiválasztani, majd arra bizonyos szűkítési feltételeket megadni, de akkor sem tartalmi, hanem egyéb szempontok szerint lehet csak szűrni (a feltöltés ideje vagy a képek értékelései alapján). Az értékelés mindig segíthet a relevánsabb találati listák létrehozásában, viszont a képek tematikus kereséséhez semmit sem képes hozzátenni. Ha kutyákról keresnénk képeket, akkor választhatjuk az 'állatfotók' kategóriát, de ezután már nem tudunk további szűkítést elérni, ezért ha 'briard' kutyafajtáról szeretnénk képeket, végig kellene lapoznunk az összes állatfotót, hogy találhassunk köztük nekünk megfelelő képet.

syil

Kreditek: 0,9

Új értékelések: 0

Új üzenetek: 0

Adatlap

Támogatás

KILÉPÉS

KATEGÓRIA: abszurd evelinda11 3,65 hirdetés

ELŐZŐ KÉP

KÖVETKEZŐ KÉP



Válassz géptípust!

- ▷ Feltöltve: 2009. március 18. 21:03
- ▷ Feldolgozható: nem
- ▷ Pontozható: igen

limóra várva
aki az élességet keresi, annak üzenem, hogy a kutyus segglyukán van. köszl.

KÖZEPES

EREDETI

PMA09 HELYSZÍNI TUDÓSÍTÁS LAS VEGASBOL

- 2009-es PMA: zárásó
- HP újdonságok
- Sigma stand
- Geolite
- Sony objektívek
- General Imaging stand
- Kodak stand
- Nikon stand
- Leica stand
- 45 gigapixel kép a Yosemite völgyről

PIXINFO.COM BLOG

- ▷ [Geekendő PS guruknak](#)
- ▷ [Fotók automatikus geotagolése](#)
- ▷ [Fényíró videók](#)
- ▷ [Fotó ágazatok - Ízlések és pofonok](#)
- ▷ [Vitorlásfotók a világ minden tájáról](#)
- ▷ [Három érdekes kép](#)
- ▷ [Fotóblog ajánló IV.](#)
- ▷ [Házi műterem tárgyfotózáshoz](#)
- ▷ [Dobjuk fel a panorámaképet](#)

LEGFRISSEBB HÍREK

FOTÓ ÉRTÉKELÉSE

A kép nem pontozható

ÉRTÉKELÉS:

Kredit átutalása:

keresés a 'briard'-ra a fotózz.hu oldalon

A flickr.com másként működik. Az oldal alapvetően ugyanazokat a funkciókat kínálja, mint a fotózz.hu, de itt a kép feltöltője nem egy merev kategóriarendszerbe sorolja be a képét, hanem szabadon, saját kulcsszavaival írhatja le azt. Ezáltal természetesen nagyságrendekkel több címkét rendelnek a képekhez (a fotózz.hu esetében láthattuk, hogy mindössze 32 kategória alá lehetett a képeket besorolni). A szabad és „bőséges” címkézésnek köszönhetően a keresési lehetőségek is jobbabbak lesznek. Szinte „mindenre” kereshetünk, mert a Flickr már elérte azt a kritikus méretet, amikor már a legextrémebb címkéket is használta valaki. Természetesen a 'briard' címke alapján is szép számmal kapunk találatokat.

Search


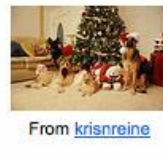
Photos

Groups

People

briard dog

 Full text
 Tags only

 We found **1,299 results** matching **briard** and **dog**.
View: [Most relevant](#) • [Most recent](#) • [Most interesting](#)Show: [Details](#) • [Thumbnails](#)From [Kathy's Pix](#)From [nils](#)From [per andersen](#)From [sandra juto](#)From [Martina V.](#)From [surferdog](#)From [surferdog](#)From [sandra juto](#)From [mybluemuse...](#)From [Johny Day](#)From [krisreine](#)From [surferdog](#)

keresés a 'briard'-ra a flickr.com oldalon

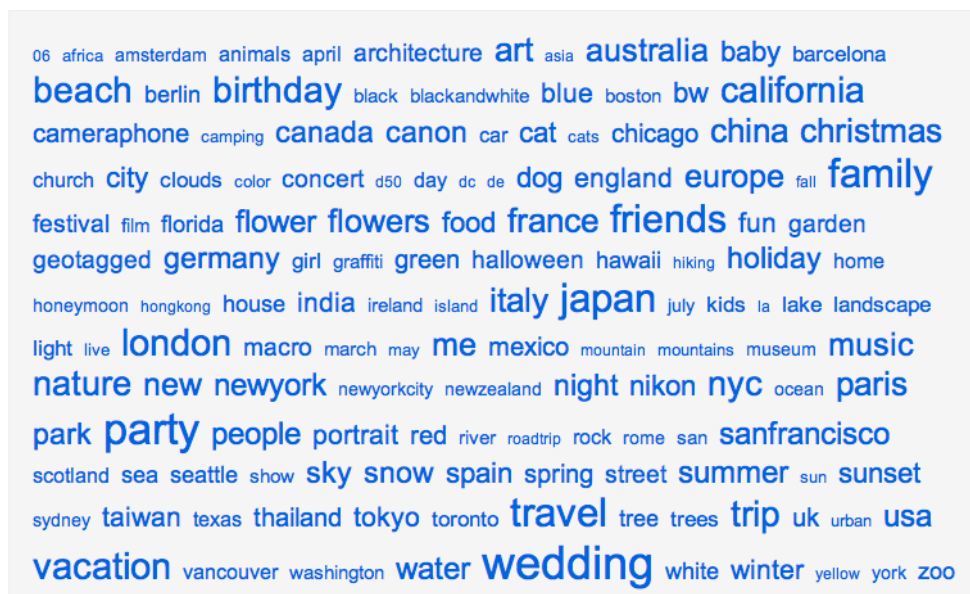
A címkézés tehát jóval rugalmasabb, mint a hagyományos osztályozási rendszer, a keresést sokkal jobban tudja támogatni. Sőt ezen felül egyfajta személyes „emlékezettechnikát” is biztosít a felhasználók számára. Említettük már, hogy a használt címkék egy része a kép tartalmát jelöli ('family', 'party', 'wedding', 'birds', 'dog', 'cat'), de a címkék utalhatnak a készítés módjára ('cameraphone'), idejére ('2006'), stílusára ('art'), helyszínére ('France') vagy akár egészen szubjektív dolgokra is ('cute', 'cool'). Vannak azonban olyan címkék is, amelyek annyira „személyesek”, hogy azok senki másnak nem hordoznak érdemi információt a címkézőn kívül ('friends', 'me', 'todo', 'toread'). Ezekkel a címkékkel nyilván senki sem akar keresni, hiszen ezek csak a feltöltő személyének ismeretében lennének értékelhetők. Ezt a kontextust csak maga a feltöltő személy ismeri, neki viszont értékesek lehetnek az ilyen címkék, amennyiben a célja nem a keresési lehetőség biztosítása bárki számára, hanem a kép személyes célú megjelölése a saját fotóalbumon belül.

A leggyakrabban használt címkék megjelenítését *címkefelhőnek* nevezik, amelyekben a címkék betűméretével, vagy egy számértékkel jelzik a címkék népszerűségét, azok gyakorisági értékeit. Már csak azért is érdemes kicsit alaposabban megvizsgálnunk ezt a jelenséget, mert közben bemutathatjuk a folkszónómiákkal kapcsolatban felmerülő problémákat is. Könnyen találhatunk ugyanis olyan példákat, amelyek azt mutatják, hogy a képzetlen és fegyelmezetlen felhasználók nehezen tudnak megbirkózni (ha meg tudnak egyáltalán) a tárgyszavazás olyan – régóta ismert – szemantikai alapproblémaival, mint:

- szinonimakezelés
- poliszémiakezelés
- homonímiakezelés

- granularitáskezelés

Kiinduló, szemléltető példaként vegyük a Flickr egy korábbi időszakból származó címkefelhőjét:



Flickr címkefelhő

Ha figyelmesen nézzük a fenti ábrát, a folkszonómiák több komoly problémáját is észre vehetjük rajta: a kontroll hiánya többféleképpen tetten érhető.

Több olyan címkepár van, amely ugyanazon fogalom egyes- és többesszámú alakjára vonatkozik ('cat' és 'cats', 'flower' és 'flowers', 'tree' és 'trees'), amelyeket a kontrollált szótár használatával természetesen ki lehetne szűrni. Az egyes- és többesszám kevert használatának problémáját általánosíthatjuk, és azt mondhatjuk, hogy a folkszonómiák nem képesek kezelni a morfológiai többalakúság egészének kérdéskörét. A nyelvi megnyilatkozásaink megformálásakor ugyanazt a szó-szintű nyelvi egységet (szót, kifejezést) sokféle morfológiai alakban használjuk a mondatok képzése során, ám ez a morfológiai többalakúság nem jelent szemantikai többértelműséget. Egy szó egyes- vagy többesszámú alakja – első megközelítésben – ugyanazt a jelentést hordozza csakúgy, mint egy szótó és annak bármelyik más, toldalékolt alakja ('ló' vagy 'lovat', 'lónak', 'lóval', 'lovon'). Ha viszont ezek az eltérő morfológiai alakok címkékké válnak (és nyugodtan válhatnak), akkor ezek mind különböző címkék lesznek anélkül, hogy bármilyen szemantikai különbséget ki tudnának fejezni. Ez a jelenség nyilván rontja a keresés hatékonyságát.

Persze, meg lehet kérdezni, hogy a felhasználók miért lennének olyan „buták”, hogy a 'ló' címke helyett a 'lovon' címkét adnák meg egy kép leírásakor. Nem arról van szó, hogy a felhasználók tényleg buták lennének (bár néha lehetnek azok is), hanem inkább arról, hogy a folkszonómiák sajátos működés módja eredményezi ezt a nem kívánt jelenséget. Ezek a rendszerek a címkéket automatikusan detektálják úgy, hogy két szóköz közti karaktersorozatot tekintenek egy címkének. Ebből következően a több szavas kifejezéseket „feldarabolják”, amit az emberi feldolgozás nyilván nem tenne meg. Példa lehet erre a 'New York' vagy a 'black & white' tárgyszó. Ezek ebben a formájukban nem szerepelnek (nem is szerepelhetnek) a rendszerben, hiszen több szóból álló „kifejezések”. A kifejezések tagjai, a 'new' és a 'york', illetve a 'black' és a 'white' címkék viszont előfordulnak külön-külön és egybeírva is: 'newyork' és 'newyorkcity', valamint 'bw' és 'blackandwhite'. Mivel a folkszonómiák az ún. „single-word” megoldást alkalmazzák a címkék behatárolására, elkülönítésére, a felhasználók – ezt tudva – az előző „egybeírós” technikát használják a több szóból álló kifejezések bevitelére.¹² Persze, nem mindenki alkalmazza ezt a megoldást, és emiatt megint csak pontatlanabbá válik a rendszer, amire a fenti címkefelhőben az utal, hogy a 'sanfrancisco' címke mellett ott szerepel a 'san' is, ami részben annak köszönhető, hogy sokan a 'San Francisco' kifejezést két szóval írták le, ezt pedig a rendszer felbontotta (ebbe persze a többi 'San' előtagú városnév (mint 'San Diego') is „belejátszhatott”). Ugyanígy jelenhetnek

¹² Előfordulhatnak más technikák is: például az alsóvonás jellel vagy a vesszővel elválasztás szóköz helyett.

meg azok a címkék is, amelyek az összetett szavak esetleges – helytelen – különírásával válnak „több szavas kifejezésekké”.

Az előző – New Yorkkal kapcsolatos – példa rávilágít a folkszonómiák egy másik problémájára is. Az ilyen rendszerekben nem lehet jól megragadni és kezelni a szemantikai hasonlóság, illetve a szemantikai azonosság problémáját. A 'New York' példából látszik, hogy a folkszonómiák nem tudják kezelni a *szinonimitás* kérdését, vagyis az azonos/hasonló jelentésű, de eltérő alakú szavak problémáját, hiszen a 'newyorkcity' és 'nyc' címkék ugyanarra a fogalomra mutatnak, mégis külön címkéként szerepelnek a rendszerben. A folkszonómiák ugyanilyen módon érzéketlenek a hagyományos szinonimitás megragadására, amelynek „klasszikus” példája a 'kutya'-'eb' szópár.

A folkszonómiáknak a szinonimák mellett kezelni kellene a *poliszémiákat* is, de ezzel a nyelvi jelenséggel sem tudnak mit kezdeni. A poliszéma azonos szóalakhoz több, valahogyan összefüggő, de azért részben eltérő jelentést rendel hozzá. A címke pontos jelentéséhez nyilván szükség lenne a különböző jelentésrétegek elválasztására, amely a folkszonómiák esetében – a felhasználói szabadság miatt – lehetetlen. A poliszémára a következő példák adhatók:

- (1) windows (mint falban lévő lyukak, amelyeken ki lehet látni)
- (2) windows (mint a falban lévő lyukakba mint keretbe helyezett üvegezett nyílászáró szerkezetek, amelyeken szintén ki lehet látni, de a levegőáramlást meg lehet akadályozni)
- (3) Windows (mint operációs rendszer, amely „ablakot nyit a világra”)

A fenti példa utolsó eleme természetesen már csak nagyon laza kapcsolatban van az eredeti jelentésekkel, de a kapcsolat léte nem vitatható, ezért poliszémiának tarthatjuk a harmadik értelmezést is.

Vannak esetek azonban, amikor két teljesen azonos szóalak jelentései már teljesen elválnak egymástól. Ekkor beszélhetünk *homonimáról*. A homonímia annyiban hasonlít a poliszéma problémaköréhez, hogy azonos szóalakok között fennálló viszonyt fejez ki, de abban már különböznek, hogy a szavak jelentésében már nincs semmi közös.

A fenti három nyelvi probléma a szóalakok és szójelentések között tetten érhető többértelműségről szól. Ezeket a többértelműségeket úgy lehet feloldani, ha kontrollált szótárakat és kezelési útmutatókat hozunk létre, amelyek pontosan szabályozzák a szinonimák, poliszémiák, homonímiák kezelésének feladatait. Ez az elvárás felkészültséget és fegyelmet követel meg, amely nem áll rendelkezésre a közösségi archiválás körülményei között.

Van azonban a tárgyszavazási tevékenységen belül még egy másik nehézség is. A jelenségek leírására alkalmas fogalmaink, szavaink, kifejezéseink eltérő pontosságúak, és mindig kérdés, hogy milyen pontossági szinten kell/lehet leírni az elemzett dolgokat. Ezt nevezik a *granularitás* problémájának. Amikor a saját kutyánkról készített fotót kell jellemeznünk, akkor ezt sokféle módon megtehetjük.

- (a) állat
- (b) kutya
- (c) briárd
- (d) briárdszuka
- (e) fekete szőrű briárdszuka

A nagy kérdés itt az, hogy mikor melyik szintet válasszuk. Ez azért különösen komoly probléma a folkszonómiák számára, mert ha a felhasználók egymáshoz képest eltérő pontossági, granularitási

nagyobb eséllyel kerülnek rá azok a címkék, amelyeket a nagy átlag is bejelölne (és így a legtöbben meg is találják a képet).

A folkszonómiák esetében a szemantikus vakság problémája sem jelent akkora gondot, mivel ha a keresési oldalon nem kapunk egyértelmű minősítést, megkülönböztetést a keresőfeltételek megfogalmazásakor (márpedig nem kapunk, hiszen a felhasználók nem adják meg az egyértelműsítéshez szükséges többletinformációt, amikor például beírják a 'cica' keresőfeltételt), akkor a keresések feldolgozása során sem tudjuk igazán feloldani a nyelvi többértelműségeket.

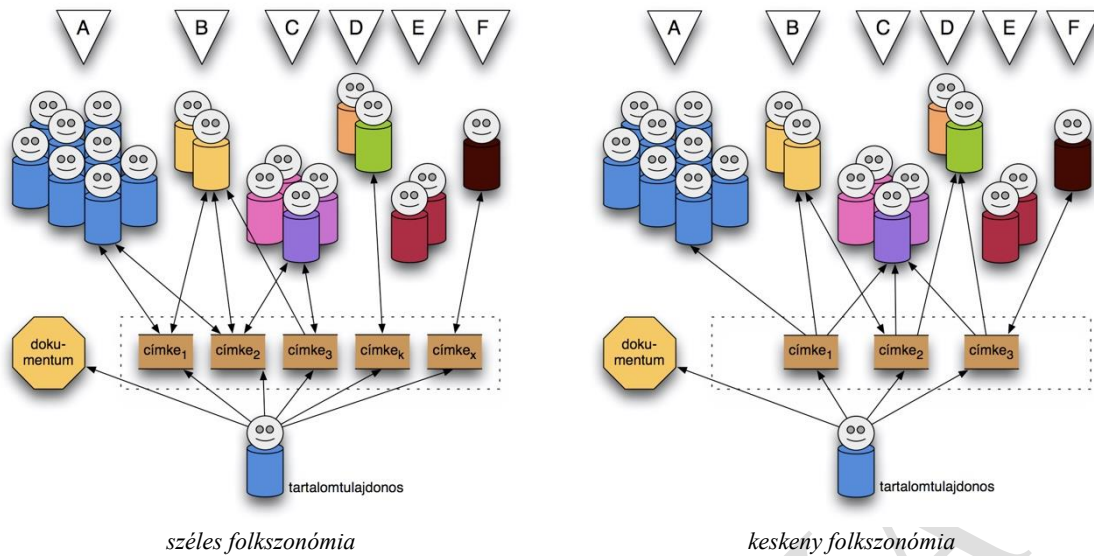
* * *

Ha megvizsgáljuk a folkszonómiák „termelési” és fogyasztási” mintázatait és sajátosságait, akkor észre vehetjük azokat a lehetőségeket, amelyek mentén tipizálhatjuk ezeket a rendszereket. Vander Wal elkülönítette egymástól a *folkszonómiák széles* és *keskeny* típusát (broad and narrow folksonomies) azon az alapon, hogy kik végzik (végezhetik) a dokumentumok, források címkézését, illetve a keresések során a felhasználók milyen „címkeszótárakat” használnak (Vander Wal 2005a). Mindkét folkszonómiatípus esetében abból indulhatunk ki, hogy azok a felhasználók, akik a folkszonómiák szkópjába tartozó dokumentumokat feltöltik, maguk is címkéket rendelnek saját dokumentumaikhoz. Minden felhasználónak van valamilyen címkekészlete (címkeszótára), amelynek elemeivel leírhatja a dokumentumokat, illetve keresheti azokat. A címkéket itt tehát kétféleképpen lehet használni. A címkék *aktív* használata a címkézést magát (azaz a címke valamely forráshoz való hozzárendelését) jelenti, míg a *passzív* címkehasználat fogalmával a keresésekhez használt címkék használatát jelölhetjük. Az egyedi felhasználókra fókuszálva kézenfekvő azt feltételeznünk, hogy az aktív és passzív címkeszótárak durván megegyezik. Azt is feltehetjük, hogy a felhasználók csoportokba rendezhetők címkeszótáraik hasonlósága (azonossága) alapján.

Az igazi kérdés az, hogy a „feltöltést végző” felhasználók mellett a többiek vajon címkézhetik-e mások erőforrásait vagy sem. Amennyiben a válasz 'igen', akkor adott dokumentum címkézését sokan végzik (végezhetik). Ebből fakad az, hogy ilyenkor kialakulhatnak nagy számosságok, amikor a címkék gyakoriságai egyre inkább hatványfüggvény szerinti eloszlást mutatnak, egyre inkább érvényessé válik a hosszú-farok hatás, valamint megjelennek az erős címkék (sokszor használt címkék) a rendszeren belül. Ekkor beszélhetünk *széles folkszonómiákról*. Ilyen esetekben az azonos/hasonló címkeszótárral (és ebből következően: azonos/hasonló érdeklődéssel) rendelkező felhasználók (felhasználócsoportok) saját (közös) szótáraik alapján címkézhetnek, illetve kereshetnek. Ez növeli a címkézés erejét és szubjektív hatékonyságérzetét.

Másként működnek a *keskeny folkszonómiák*. Az ilyen rendszerekben a címkézést csak kevesek végzik, végezhetik (a feltöltők és – esetleg – még olyan felhasználók, akiknek a feltöltők engedélyt adnak). Emiatt jóval kevesebb címke kerül be a rendszerbe, és természetesen kevesebb címkézési művelet valósul meg. Ez nehezíti a hatványfüggvény szerinti eloszlás kialakulását és mindazon előnyök hasznosítását, amely ebből fakadhatna. További fontos következménye ennek a helyzetnek az, hogy a keresések során nehezebben alakulnak ki közös szótárak a felhasználók között (mivel az aktív és passzív címkehasználat teljesen elválik egymástól), ezért a keresés is nehezebb lesz az ilyen rendszerekben. Persze a feltöltők számára nem jelentkezik ez a címkeszótár-probléma, ezért személyes használat esetében ezek a rendszerek továbbra is kiválóan működnek (a felhasználók egyéni címkézési tevékenységét *perszonómiának* (personomy) is nevezik – Vander Wal 2005b).

A kétféle rendszert az alábbi két ábra segítségével jellemezhetjük (Vander Wal 2005a).



a folkszonómiák típusai

A széles folkszonómiára a Delicious, a szűkre a Flickr szolgáltató példát.

Abból a tényből, hogy a címkéző felhasználók (prosumerek) a folkszonómiák építésekor, használatakor nem kontrollált szótárból választanak kulcsszavakat, az is következik, hogy nem lehet tudni, vajon ugyanabban (vagy legalább hasonló) értelemben használják-e egy adott címkét vagy sem. Látnunk kell azonban azt, hogy ez nem kizárólag a folkszonómiák esetében van így. A *szótárprobléma* (vocabulary problem) már a hagyományos kommunikáció vizsgálatoknál is felmerült a kutatókban, amikor azt vizsgálták, hogy a felhasználók milyen kulcsszavakat használnak ugyanazon objektumok leírására (Furnas et al. 1987). A felmérések eredménye az volt, hogy a hagyományos tárgyszóhasználat világában is gyakran elfordult, hogy a különböző felhasználók nagyon eltérő szavakat, kifejezéseket használtak, vagy ugyanazokat a tárgyszavakat nagyon eltérő jelentés mentén értelmezték. Tanulságos, ahogy Furnas és szerzőtársai a tanulmányukat zárják.

„A hatalmas, komplex rendszerek esetében nagyon sok – belépési pontként funkcionáló – alternatív kifejezésre szükség lehet ahhoz, hogy a felhasználók megkaphassák azt, amit keresnek.” (Furnas et al. 1987)

Erre hivatkozva persze fel lehetne tenni a kérdést, hogy – ebből a szempontból – mennyire volt hatékony a könyvtári, archívumi világ évszázados gyakorlata. Vajon a mindig is létező és ható szótárprobléma nem teremtett-e és nem tartott-e fenn párhuzamos értelmezési univerzumokat. Szerencsére azonban ezt a kérdést itt nem kell megválaszolnunk.

A folkszonómiák leírásakor a keskeny és széles folkszonómiák elkülönítése mellett figyelembe vehetünk más tipizálási szempontokat is (Marlow et al. 2006). Figyelhetünk arra, hogy milyen vonásai vannak a rendszer működésének magának, de értékelhetjük azt is, hogy a rendszerek milyen eszközökkel próbálják meg ösztönözni a felhasználókat arra, hogy aktív címkézési tevékenységet folytassanak. A folkszonómiák rendszerszintű leírásakor az alábbi szempontokat érdemes figyelembe venni:

- Milyen címkézési jogosultságai vannak a felhasználóknak
 - Ki címkézhet?
 - öncímkézés (self-tagging) – Technorati
 - szabad címkézés (free-for-all tagging) – Yahoo! Podcasts
 - megosztott címkézés (például: barátok, családtagok, ismerősök) – Flickr
 - Ki vonhat vissza címkét?
 - senki – Yahoo! Podcasts

- címkebejegyző (tag creator) – Last.fm
 - forrástulajdonos (resource owner) – Flickr
- Milyen támogatást adnak a címkézési tevékenységhez?
 - vak címkézés (blind tagging) – Del.icio.us
 - látható címkézés (viewable tagging) – Yahoo! Podcasts
 - ajánló címkézés (suggestive tagging) – Yahoo! MyWeb2.0
- Milyen aggregálás?
 - Zsákmodell: megengedi a címkék többszörös kiosztását (bag-model) – Del.icio.us
 - Halmazmodell: egy címkét egyszer enged egy forráshoz rendelni (set-model) – YouTube, Flickr
- Milyen típusú szociális tárgyról (objektumról, forrásról, dokumentumról) van szó?
 - weboldalak – Del.icio.us, Yahoo! MyWeb2.0
 - bibliográfiai tételek – CiteULike
 - blogposztok – Technorati, LiveJournal
 - képek – Flickr, ESP Game
 - felhasználók – LiveJournal
 - videók – YouTube
 - zeneszámok – Last.fm
 - podcast anyagok – Yahoo! Podcasts, Odeo
 - helyszínek, események – Upcoming, Yelp
- Ki biztosítja a szociális tárgyat (objektumot, forrást, dokumentumot)?
 - résztvevők – YouTube, Flickr, Technorati, Upcoming
 - rendszer – ESP Game, Last.fm, Yahoo! Podcasts
 - globális: bármilyen webforrás beemelhető – Del.icio.us, Yahoo! MyWeb2.0
- Milyen kapcsolat van a források között (resource connectivity)?
 - összekapcsolhatók (linked) – Web
 - csoportba rendezhetők (grouped) – Flickr
 - semmilyen
- Milyen kapcsolat van a felhasználók között (social connectivity)?
 - összekapcsolhatók (linked) – Flickr
 - lehet minősített (barát, ismerős)
 - lehet irányított vagy nem (szimmetricitás)
 - csoportba rendezhetők (grouped) – Del.icio.us (részben)
 - semmilyen

Ha a felhasználói ösztönzés eszközeit, technikáit keressük, akkor az alábbi motivációkat mindenképpen figyelembe vehetjük:

- jövőbeni visszakeresés (future retrieval)
- megosztás (contribution and sharing)
- figyelemfelkeltés (attract attention)
- játék, verseny (play and competition)
- önreprezentáció (self presentation)
- véleménykifejezés (opinion expression)

Nyilván lehetne még másfajta szempontokhoz is igazodni, de nincs most szükségünk arra, hogy a folkszonómiák alaposabb tipizálását elvégezzük. Érdemesebb az eddig leírtak alapján összehasonlítani és értékelné a tudásszervezési rendszereket. Említettük, hogy a terminuslisták nagyon egyszerűen használhatók, de a belső struktúra hiánya miatt nem igazán tudnak támogatást nyújtani az eligazodáshoz. Ehhez képest nagy előrelépést jelent a taxonómiák megjelenése, hiszen azokban egy hierarchikus reláció mentén egy már nagyon jól navigálható struktúrát képezhetünk, amely – elméletileg – sokkal inkább képes lehet a navigáció támogatására. Ennek azonban az az ára, hogy megfelelő tudású és fegyelmezettségű szakemberek kelljenek a működtetéséhez. Még inkább így van ez a teauruszok esetében, amelyeknek jóval bonyolultabb struktúrájuk van. A teaurusz sokkal rugalmasabb, pontosabb leírásra, barangolásra ad lehetőséget, de az építése, fenntartása „még többbe kerül”, a specializációt

keményen meg kell fizetni. A folkszonómiák más munkamegosztási séma szerint működnek, amelynek egyaránt vannak előnyei és hátrányai. Az előnyei közé sorolhatjuk a következőket.

- alacsony „humán erőforrás” költség: nincs szükség megbízható, szakmailag képzett, fegyelmezetten dolgozó, s ezért „drága” katalogizáló szakemberekre;
- erős önszabályozási képesség, demokratikus, decentralizált működésmód;
- rugalmas, adaptív, friss;
- azonnali visszacsatolási lehetőséget biztosít;
- könnyű használni;
- a „szabadpolcos felfedezés” támogatása (angolul: 'serendipitous discovery').¹³

A folkszonómiák hátrányai között azokat a vonásokat említhetjük meg, amelyek a felhasználók „képzetlenségéből” és „fegyelmezetlenségéből” fakadnak. A leggyakrabban emlegetett hátrányok:

- alacsony relevanciamérték (pontosság és felidézés) értékek – alapvetően a szinonimák és poliszemiák miatt;
- alapszint-hangolási, granularitási probléma: túl speciális, túl általános címkék együttes kezelésének nehézségei;
- a hierarchia hiánya, vezérlési gondok;
- túlérzékenység, védtelenség a rosszindulatú felhasználói tevékenységgel (csalással, rejtőzködéssel, spamtevékenységgel, tendenciózus tévedésekkel) szemben;
- gyenge kereséstámogatási képesség speciális keresési igények esetén.

A pro és kontra érvek más szempontokat emelnek ki illetve tartanak hangsúlyosnak, ezért nem könnyű az összehasonlításuk. A végérvényes döntést talán ma még nem érdemes meghozni, ha választani akarunk a kétfajta rendszer között, mert vélhetőleg még mindkét oldalon sokat fognak tanulni egymástól, és a jövőben inkább a valamilyen mértékig kevert rendszerek jelennek majd meg. Ezt erősítik azok az elképzelések is, amelyek a kétfajta elv ötvözését keresik, s már új terminust is generáltak erre. A 'collabulary' új terminust az alábbi "képlettel" magyarázzák.

folkszonomy + controlled vocabulary = collaborative vocabulary = collabulary

Látnunk kell azonban azt is, hogy a folkszonómiák terjedésével párhuzamosan válik egyre jelentősebbé az ún. *szemantikai rés* problémája. A folkszonómiák terjedésének, dominássá válásának legfőbb okaként azt adtuk meg, hogy a metaadat-hozzárendelés nehéz, fáradságos emberi munkáját sokak számára lehetővé téve, nagyszámú „szabad” és – ami a legfontosabb – „ingyenes munkát” várhatunk az önkéntes felhasználók seregétől. Ahhoz, hogy ez működjön, nem lehet semmilyen megkötést előírni a címkézési munkát végzők számára, tehát fel kell adni az előzetes szakképzésre, szakértelemre vonatkozó elvárást, a kontrollált szótárak fenntartásának igényét és a munka ellenőrzésének lehetőségét. Mindebből a fontos mozzanat most az, hogy kontroll hiányában nem lehet semmiféle szemantikai struktúrát elvárni és fenntartani a rendszer elemei között, ugyanakkor a struktúra hiánya miatt egyre nagyobb mértékben nő a szemantikai rés nagysága. E probléma megszüntetésére, csökkentésére megoldást jelenthet egyrészt a szemantikai relációk helyettesítése másféle (forgalmi-használati) relációk bekapcsolásával az egyik oldalon, másrészt a hiányzó szemantikai relációk pótlása speciális emberi szaktudás, valamint gépi intelligencia bevonásával. A kérdés az, hogy milyen módon lehet ez utóbbi hiányzó tudást a rendszerbe táplálni. A válasz pedig a számítógépes szemantikus technológiák, ipari ontológiák, következtető rendszerek felépítése, kifejlesztése és alkalmazása lehet, de azt tudnunk kell, hogy ennek az útnak csak az elején járunk.

A metaadatok dokumentumokhoz rendelésének, és ezáltal a visszakereshetőség biztosításának évszázados hagyománya olyan előfeltevéseken nyugszik, amelyek megkérdőjelezésével az egész építmény könnyen utópisztikus jellegűvé válhat (Doctorow 2001). Cory Doctorow írt egy elgondolkodtató esszét a 'metahadováról', a metaadat-utópia téveszméiről. Nézzük meg röviden, mit is

¹³ A “szabadpolcos felfedezés” kifejezés alatt azt a jelenséget értjük, amikor a könyvtárak olvasótermeiben kiteszik egymás mellé a frissen érkezett könyveket (vagy egy másik polcra az olvasóktól épp visszaérkezett könyveket), és a könyveknek ilyen egymás mellé kerülése gyakran érdekes „felfedezésekre”, „találatokra” ad alkalmat. A folkszonómiák ugyanezt az élményt képesek nyújtani azzal, hogy a különböző felhasználók által a dokumentumokhoz rendelt címkéket megmutatva érdekes kapcsolatokra „hívják fel” a figyelmet.

gondolt problémásnak Doctorow. Szerinte a nagy metaadat-elképzelések azért maradtak mindig utópiák, mert az ilyen víziók kidolgozásakor nem vették figyelembe az alábbiakat.

- az emberek hazudnak (a figyelemért folytatott versenyben „kiszínezik” a metaadatokat)
- az emberek lusták (ha egy kattintásnál többet kell tenniük)
- az emberek figyelmetlenek ("hülyék") (akkor sem figyelnek oda eléggé a metaadatok helyességére, ha elemi érdekük lenne, így sokszor elírják azokat)
- „lehetetlen küldetés – ismerd meg önmagad” (az emberek becsapják önmagukat is, másokat is, sokszor nem megbízhatóak az állításaik)
- a sémák nem semlegesek (a metaadatsémákban kialakított hierarchiák, struktúrák csak bizonyos kontextusokban érvényesek, máshol elvesztik a relevanciájukat)
- az értékelési-elemzési szempontok sokszor meghatározzák az értékelés kimenetét (a fontosnak tartott metaadatok kiválasztása jelentős részben meghatározza, hogy mit lehet és mit nem lehet leírni velük, így bizonyos jelenségek „eltűnhetnek” a metavilágból, vagy be sem kerülhetnek oda)
- mindig mindent többféleképpen írhatunk le (ugyanazt a jelenséget különféle szempontok szerint sokféle módon értelmezhetjük, értékelhetjük)

Erősen megfontolandó Doctorow listája, különösen a szótárprobléma kapcsán korábban írtak fényében. Persze nem feltétlenül kell/lehet minden megállapításával egyetértünk. Azt azért megállapíthatjuk, hogy ezek a kifogások alapvetően pragmatikai jellegűek, ez pedig arra ösztönöz, hogy a továbbiakban elsőbbséget adjunk ennek a szempontnak.

2.15.5.1 Formális jellemzés

A folkszonómiák – a web2-es paradigmának megfelelően – a metaadat-hozzárendelési tevékenységet teljes mértékben saját felhasználóikra bízák. Ezt természetesen csak akkor tehetik meg, ha nem várnak el semmilyen fegyelmet (és semmilyen speciális szaktudást) azoktól, akik a metaadatokat a dokumentumokhoz rendelik (azaz a felhasználói közösségük tagjaitól). Nem is tehetnek mást. Ha bármit megkövetelnének a felhasználóktól, akkor azok vélhetőleg tovább állnának olyan helyekre, ahol szabadon „mozoghatnak”. Ebből következően viszont nem számíthatunk arra, hogy a – „képzetlen és fegyelmetlen” – felhasználók bármiféle szemantikai relációt felvennének a rendszerbe, vagy ha mégis, akkor sem remélhetjük, hogy konzekvensen kezelnék azokat. Az ilyen rendszerekben a szemantikai relációk helyett tehát másfajta relációk alkalmazására van szükség, ez pedig azt jelenti, hogy a folkszonómiákat az eddigiekhez képest nagyon másként kell leírunk.

$$KOS_{folk} = \langle D, R_1, R_2, R_{10}, R_{11}, S_4 \rangle, \text{ ahol}$$

| | | | |
|----------|--------------------|-----------|---------------------|
| R_1 | lexikografikus | rendezési | reláció |
| R_2 | | | ekvivalenciareláció |
| R_{10} | címkegyakorisági | | reláció |
| R_{11} | címke-együttjárasi | | reláció |

A folkszonómiák megjelenésével a legfontosabb változás az, hogy azok az S_i munkaszervezési szabályok közül csak egyet vesznek figyelembe (az S_4 -es normát). Azt sem mindig, sőt talán még az is megkövethető, hogy nagyobb azoknak a folkszonómiáknak a száma, amelyek esetében ettől a szabálytól is eltérnek (amikor figyelembe veszik az S_4 -es szabályt, az akkor is „csak” annyit jelent, hogy a felhasználók kizárólag a saját maguk által feltöltött dokumentumokat címkézhetik, másokét nem). Azt mondhatjuk, hogy a folkszonómiák egyetlen igazi szabálya az, hogy „nincs szabály” – mindent szabad.

Ebből a „követlenségéből” sok minden következik. Ha nem követelünk meg semmilyen rendszerépítési szabályt, akkor egyrészt nem biztosíthatjuk a terminusok egyértelműségét a rendszeren belül (vagy másként mondva: nem lesz kontrollált szótárunk), másrészt szükségszerűen elveszítünk minden relációt a rendszerből, hiszen a szabad terminusfelvétel lehetősége mellett a címkézést végző személyektől nem követelhetjük meg azt, hogy az új címkéket hozzárendeljék a rendszerben már létező elemekhez. Ekkor viszont nem is tudunk komolyabb struktúrát értelmezni a címkehalmazon, és emiatt nem számíthatunk a szemantikai struktúra meglétéből fakadó – navigációs és következtetési lehetőségeket biztosító – előnyökre sem. Fontos újdonság azonban a folkszonómiákban az új – se nem szintaktikai, se nem szemantikai, sőt, nem is nyelvi, hanem *forgalmi* – relációk megjelenése (R_{11} és R_{12}). Az R_{11} reláció egy olyan, a címkékhez rendelt gyakorisági viszony, amely azt mutatja, hogy a felhasználók milyen gyakran használják az adott címkét a dokumentumok leírására. Ennek a relációnak azért van kiemelt jelentősége, mert ezáltal újfajta relevanciakezelési lehetőséget lehet biztosítani a folkszonómiák számára. A címkegyakoriság ugyanis megmutatja az adott címkének a felhasználói közösségen belüli „népszerűségét”, fontosságát, és ennek az értéknek a figyelembevétele már elég jó alapot nyújt a relevanciakezeléséhez. Másféle segítséget képes nyújtani az R_{12} reláció, amely azt mutatja, hogy a többi felhasználó korábban milyen más címkéket rendelt a dokumentumokhoz az éppen használatban levő címkével együtt. A címkéknek ezt a fajta együtt járását megmutatva segíteni, orientálni lehet a felhasználókat a megfelelő címkék megtalálásában.

2.15.6 Formális ontológia

Az ontológia tudományát a filozófia egyik ágaként művelték sokezer éven át, mígnem előbb a mesterséges intelligencia kutatások, majd később a szemantikus web kezdeményezés hatására megjelentek az ipari, formális ontológiák fejlesztésével kapcsolatos projektek. E tanulmány írásának időpontjában még nem történt átütő siker ezen a téren. A 2000-es évek elején indult szemantikus web kezdeményezés sokakban nagy reményeket ébresztett, de kezdettől fogva hallhatóak voltak azok a kritikus hangok, melyek a formális ontológiák építésének és alkalmazásának nehézségeit hangsúlyozták. A formális ontológiát is tudásszervezési rendszerként kell értelmeznünk, hiszen – ha eltekintünk az ontológia egyik fontos kérdésétől, a nyelvfüggetlen fogalmak és a nyelvfüggő lexikai egységek elkülönült kezelésének problémájától, és elmoszuk a fogalmak és terminusok közti határvonalakat, akkor – az ontológia is tárgyszavakból (fogalmakból) és a köztük értelmezett relációkból áll. A kérdés az, hogy milyen relációkat veszünk bele egy ontológiába, és erre egyelőre még nincs általánosan elfogadott válasz a világban az ezzel a kérdéssel foglalkozó szakemberek között. Pontosabban jelenleg az a válasz, hogy az ontológia belső relációi részben az ontológiát építő és használó emberek *ontológiai elkötelezettségein*, részben az *ontológiahasználati célokon* múlnak. Ebből következően nem, vagy csak nehezen beszélhetünk egyértelmű és széles konszenzuson alapuló ontológiarelációkról (Feilmayr & WöB 2016).

2.15.6.1 Formális jellemzés

Az ontológiák formális leírása azért nem lehetséges igazán, mert még nem alakultak ki széles körben elfogadott szabványok, megoldások.

$$KOS_{font} = \langle D, R_1, R_5, R_{16}, R_{17}, R_{18}, S_1, S_2, S_3, S_4, S_5 \rangle, \text{ ahol}$$

| | | | |
|----------|-------------------------------|-------------|--------------------------|
| R_1 | lexikografikus | rendezési | reláció |
| R_5 | generikus | alá- és | fölérendeltje relációpár |
| R_{16} | | tetszőleges | relációfogalom |
| R_{17} | | tetszőleges | osztályfogalom |
| R_{18} | tetszőleges attribútumfogalom | | |

A meghatározatlanságot az a tény fejezi ki, hogy az ontológiák általános leírásába csak egy teljesen általános, így praktikusán jelentés nélküli reláció-, osztály- és attribútumfogalmakat vehetünk csak fel.

Az ontológiák építésére, használatára sokféle kezdeményezés, projekt indult és indul világszerte, de a nagy áttörés egyelőre várat magára, az ontológiák fejlesztésére és használatára vonatkozóan egyelőre nem beszélhetünk sem igazán komoly eredményekről, sem széleskörű konszenzusról.

Érdemes viszont hangsúlyozni itt azt, hogy az ontológiák (formális ontológiák) esetében már nem emberi, hanem gépi használatban kell gondolkoznunk. Az ontológiaépítés még sokáig emberi közreműködéssel valósulhat meg (még ha vannak olyan elképzelések, amelyek szerint az ontológiaépítés feladatát is lehetne „gépesíteni”), de az ontológiahasználat már a számítógépek feladata. Emiatt az ontológiák kicsit ki is lógnak a tudásszervezési rendszerek sorából.

2.16 Dokumentumleíró rendszerek

Mind a névterek, mind a tudásszervezési rendszerek valamilyen struktúrának feleltethetők meg, melyek szemantikája részben a rendszerben alkalmazott relációktól, részben a terminuskészlet összeállításakor érvényesített ontológiai elköteleződésektől függ. Ezek a rendszerek önmagukban, „üresen” minden tudásterületen alkalmazhatóak, nincs olyan szakterület, ahonnan ontológiailag ki lennének zárva.

Vannak azonban olyan tudásrendszerek, amelyek valamilyen ontológiai elkötelezettség mentén adott szakterületre vonatkozó tudáskészletet akarják csak kezelni. A vizsgálódásunkat a könyvtári világra szűkítve, a példákat erről a szakterületről keresve hivatkozhatunk a MARC21, az RDA vagy a BibFrame rendszerekre. Ezeknek is van szemantikája, ezek is felfoghatók egy összetett struktúrának, de ezek funkcionalitása más az eddig tárgyalt rendszerekhez képest. Ezek is használnak, használhatnak tulajdonnévtereket, köznévtereket, tudásszervezési rendszereket, de a fő céljuk speciálisabb: valamilyen dokumentumtípusba tartozó dokumentumpéldányok tartalmi és formai leírása. Ezt a célt azzal érhetik el, hogy felvázolnak egy olyan – a szakma ontológiai elkötelezettségeit reprezentáló, leszűkített – világsémát, amelyben minden elmondható a kiválasztott dokumentumtípusról. Ebben a sémában benne vannak mindazok az entitások, amelyek szükségesek a terület szakszerű leírásához, és rögzítve vannak a modell entitásai közti relációk, illetve azok a pontok, ahol – ha szükséges – igénybe lehet venni "külső" névtereket, tudásszervezési rendszereket. Ezek a dokumentumleíró rendszerek kapcsolódhatnak (sőt, kapcsolódnuk kell) névterekhez, de más a céljuk, más a tartalmuk.

A névterek igazi értelme az általuk leírni kívánt ontológiai szegmensben felmerülő nevek egyértelmű azonosítása. Ha ezt a feladatukat sikeresen látják el, akkor azok a – könyvtári, múzeumi, levéltári, "tetszőleges nyilvántartási" – rendszerek, amelyek a névterek azonosítóit összerendelik saját névazonosítóikkal, átjárhatókká válnak,

2.17 Tulajdonnév és megnevezés

Korábban már elkülönítettük a névterek két fő típusát, a köznévtereket és a tulajdonnévtereket egymástól. Ebben a fejezetben a tulajdonnevek jellegzetességei közül egyet kiemelve rámutatunk arra, hogy tulajdonnevek esetében nem a nevet, tehát valamilyen nyelvi szimbólum által közvetített információt kell a névtér központi kategóriájának tekinteni, hanem a név valakire vagy valamire való vonatkoztatását mint eseményt. A tulajdonnevek esetében a névtér nem nevek mint megnevezések tereként értelmezhető. Ezt a fontos tézist Saul Kripke által bevezetett merev jelölő fogalma köré épített elméletre támaszkodva mutatjuk be (Kripke 2010). Ennek során tisztázni fogjuk, hogy a névterekre így tekintve miért fontos – elméleti és gyakorlati szempontból egyaránt – a megnevezői közösség kontextust biztosító szerepe, és kibontjuk azt is, hogy ennek az eseményszemantikai megközelítésnek az alkalmazásából milyen következmények fakadnak.

2.18 A névtér összetevői

A következő fontos kérdés az, hogy miket tartunk a névterek elemeinek. Az nyilvánvaló, hogy akárhogy is értelmezzük a névtereket, a neveknek mindig a névtér elemei közt van a helyük. A névtereknek a lehető legszűkebben vett értelmezése az lehet, amikor a névtereket nevek egyszerű gyűjteményeként, felsorolásaként definiáljuk megadva azt, hogy milyen típusba sorolható dolgokat

írhatunk le velük. Ezen értelmezés szerint a névtér a – valamilyen célra – kiosztható, használható nevek egyedi listája. Ebben az esetben egy földrajzi névtérben az 'Öreghegy' név csak egyszer szerepelhetne, és csak annyit állíthatnánk vele kapcsolatban, hogy ez a név rámutathat egy vagy több földrajzi helyre. A legszűkebben értelmezett családnévtérről is csak annyit várhatnánk el, hogy tartalmazza az összes lehetséges családnevet úgy, hogy minden név csak egyszer forduljon elő. Egy ilyen névtérnek egyszerre kell diszjunktnak és lefedőnek lennie. Diszjunktnak azért, hogy biztosítsa a névtér a nevek egyértelműségét, lefedőnek pedig azért, hogy a névtér garantálhassa, hogy minden lehetséges, kiosztható névalak szerepel benne. A névtér – ilyen leszűkítő értelmezés mentén – arra lehetne használni, hogy valamilyen adatrendszerben kontrolláljuk a névelemek választásának folyamatát azzal, hogy bizonyos adatbeviteli pontokon (például családnevek megadásakor) csak adott névtérből (családnévtérből) lehessen választani. Túl sok gyakorlati haszna azonban nem lenne az ilyen névtérnek, legalábbis abban az értelemben nem, hogy a névtér segítségével elérjük azt a célt, hogy a névtér elemeivel minden esetben egyértelműen azonosítani lehessen a névtérrel leírt dolgok halmazát. Erre a célra az ilyen névtér csak akkor lenne alkalmas, ha a nevek nagyobb számban állnának rendelkezésre, mint a hivatkozni kívánt dolgok, vagy ha a neveket valamilyen hierarchikus rend szerint újra lehetne hasznosítani. Személyek, testületek, földrajzi helyek és még sok minden más dolog esetében ezt nem tudnánk biztosítani, ezért a nevekkel ilyen kontextusokban nem garantálható a homonimitásmentesség. Visszatérve arra a kérdésre, hogy mik lehetnek a névtér – további – elemei, a válaszhoz figyelembe kell vennünk azt, hogy milyen célra tartjuk fent a névtérket. Ha a névtér elemeit valamilyen dolgok beazonosítására akarjuk használni (márpedig arra akarjuk), akkor a névtér elemei közé kell sorolnunk azokat a dolgokat is, amelyekre a nevekkel hivatkozunk. Ezeket nevezzük **névhordozóknak**. A fájlrendszerben az állományok azok a névhordozók, amelyeket az fájlnevekkel azonosítani lehet, a domainnév-rendszerben pedig az IP-címek (illetve az oda pakolt tartalomgyűttesek) azok a névhordozók, amelyeket a domainnevekkel azonosíthatunk. Névhordozó lehet egy személy, akire utalhatunk egy vagy több névvel (de mindegyikkel azonosítani kell tudni), névhordozók lehetnek a testületek vagy a földrajzi helyek is.

Ha a névtér nemcsak neveket, de névhordozókat is tartalmaznak, akkor tartalmazniuk kell még valami mást is, mégpedig a **nevek és névhordozók közti levő kapcsolatokat**. Voltaképpen e kapcsolat kezelése a névtér alkalmazásának igazi értelme.

2.19A tulajdonnévtér meghatározása

A névtérket a nevek és névhordozók közti – térben és időben dinamikusan változó – kapcsolatrendszer minősége alapján definiálhatjuk: **a névtér egy függvény, amely minden lehetséges megnevezésre egy és csak egy névhordozót ad vissza**. Ez a meghatározás egyszerűnek látszik, a kérdés inkább az, hogy miként lehet működtetni a gyakorlatban az így definiált névtérket.

A személynevek példája kapcsán már kiderült, hogy a nevek önmagukban nem képesek a személyek (névhordozók) egyértelmű azonosítására: több személynek is lehet 'Kovács János' a neve. A kérdés az, hogy ilyen feltételek mentén, hogyan képes a névtér függvényként működni. A továbblépés érdekében pontosítani kell az eddig használt fogalmainkon. Az első pontosítás már ott van a névtér fenti meghatározásában, amelyben nem "lehetséges nevek", hanem "lehetséges megnevezések" szerepelnek. Mi a különbség a két kifejezés közt, és miért van szükség az új kategória bevezetésére?

Az eddig hivatkozott névtérket mindegyikére igaz (még ha egy-két esetben talán meglepő is), hogy azok a nevek, amelyeket névtérelemként használunk, mind **tulajdonnévként** értelmezhetők. A személynevek, a földrajzi nevek, a testületi nevek esetében ez magától értetődő, a domainnevek esetében is könnyen belátható állítás, a fájlrendszer esetében viszont első hallásra meglepő lehet ez a minősítés. Pedig a fájlnevek ugyanazt a szerepet töltik be az állományok esetében, mint a személynevek a személyek, a földrajzi nevek a földrajzi helyek esetében. Egy konkrét létező dologra (fájlra, személyre, földrajzi helyre) mutatnak rá azért, hogy e megjelölés révén azt az egy konkrét dolgot meg lehessen különböztetni a többi hasonló, konkrét dologtól (az adott fájl a többi fájlától, az adott személyt a többi személytől stb.). Amikor neveket ilyen célra használunk, akkor speciális neveket, tulajdonneveket alkalmazunk.

A tulajdonneveknek az a lényegi, a köznevekhez képest megkülönböztető tulajdonsága, hogy a tulajdonnevek merev jelölők (Kripke 2010). A filozófusok azt ugyan vitatják, hogy a tulajdonneveknek

van-e jelentésük, de abban egyetértés van köztük, hogy a tulajdonnevekkel egyetlen individuumra referálunk, és ez a hivatkozás merev abban az értelemben, hogy minden elképzelhető feltételek között (minden lehetséges világban) megmarad ez a jelölő, referáló kapcsolat a tulajdonnév és a hivatkozott individuum között. Ez nem így van a köznevek esetében, amikor is eltérő feltételek közepette (más és más lehetséges világokban) változhat a nevek terjedelme, extenziója. Felmerül azonban a kérdés, hogy a nevek, tulajdonnevek hogyan tölthetik be ezt a funkciójukat, ha egy fontos filozófiai minőségükben nagyon különböznek a jelölt entitásoktól. Miről van szó?

Amikor – az azonosításra szolgáló – névterek kapcsán névhordozókról beszélünk, akkor mindig – valamilyen típusba sorolható – konkrét individuumról van szó (emberekről, emberek egyedi csoportjáról, földrajzi helyekről, könyvekről, IP-címekről stb.). Ezek létezését, konkrétságát és egyediségét nem nagyon lehet vitatni: ezek mind egyediek abban az értelemben, hogy lokalizálhatók a térben és időben.¹⁴ Amikor azonban ezekre a névhordozónak minősített konkrét létezőkre, példányokra utalunk a névtér névelemeivel, akkor a nevek térhez, időhöz nem köthető minősége miatt komoly problémával kerülünk szembe. A nevek ugyanis egy időben több helyen is létezhetnek. Egy személyre tudjuk azt mondani, hogy éppen most hol tartózkodik, de a személy nevére nem mondhatjuk azt, hogy "most itt vagy ott van". Pont ebből fakad a homonimitás problémája is. Amikor egy névvel azonosítani akarunk egy embert, azzal szembesülhetünk, hogy ugyanazt a nevet kiadhatjuk több személy számára is. Akkor viszont adott a kérdés: hogyan lehet a névtereket azonosításra használni, ha a tulajdonnevek – önmagukban – nem alkalmasak erre.

2.20A nevek egyértelműsítése

A kérdést feltehetjük kicsit másként is: hogyan lehet a (tulajdon)neveket úgy egyedivé tenni, hogy alkalmassá váljanak az individuális névhordozók azonosítására. Nos, a megoldás az lehet, hogy az azonosításra nem a (tulajdon)neveket, hanem **megnevezéseket** (megnevezési eseményeket vagy megnevezési állapotokat) használunk. Ez a megoldás Kripke tulajdonnév-értelmezésére vezethető vissza (Kripke 2010). Kripke szerint egy tulajdonnév referenciáját nem valamilyen hozzá társított leírás alapján határozhatjuk meg, hanem a névhez kapcsolódó, történeti láncba szerveződött megnevezések révén:

"... mikor kimondom a „Rubik Ernő” nevet, annak referenciája az a személy, akit a megnyilatkozásomhoz kapcsolódó névhasználatok történeti lánc (vagy inkább hálójá) meghatároz. A lánc kezdeténél a név bevezető használata áll, és ezt követik a név továbbadó használatai, amelyek révén a név végül eljutott hozzám. Nagyjából így foglalhatjuk össze a tulajdonnevek történetilánc-referenciaelméletét." (Zvolenszky 2015)

A (tulajdon)névadás és -használat tehát valamely megnevezői közösség praxisában kialakuló merev jelölő kialakítását és tartós fennmaradását jelenti. Az új elem itt a megnevezői és névhasználati gyakorlat mozzanata. "Rámutatások", névadási gesztusok révén egy közösség gyakorlatában egy névhordozóhoz tapad egy (tulajdon)név, és ez a megnevezési aktusok összefüggő hálózatán keresztül fenntartja a név és a referált dolog közti állandó kapcsolatot. Nem a (tulajdon)név azonosít tehát, hanem a névhasználó közösség megnevezési gyakorlata, a megnevezés(lánc), ami viszont mindig egyedi események láncolatát jelenti. Ez azért fontos, mert az egyedi névhordozókat az individuális megnevezéseken

¹⁴ Az individuumok definitív minősége, hogy azok mindig a térhez és időhöz kötöttek. Egy individuum esetében – elvileg – mindig meg kell tudnunk adni, hogy hol "helyezkednek el" a térben és az időben. Egy épületről, egy hegyről, egy személyről, egy könyvről, egy videokazettáról, egy CD-lemezről mindig állítható, hogy egyértelműen lokalizálhatók a tér és az idő dimenzióiban. A földrajzi entitásokat eleve azzal tudjuk meghatározni, hogy megadjuk, hogy az összetevői hol helyezkednek el a földrajzi térben egy adott időpontban. Elvileg bármely személyt azonosíthatnánk azon a módon, hogy megadjuk, a tér melyik szegmensét foglalta el egy konkrét időpontban, de ez a megoldás információs szükségletét tekintve nem lenne gazdaságos, hiszen minden személyről tudnunk kéne a mindenkori tér- és időlokalizációjukat. Praktikusabb ehelyett nevezetes eseményeket kiragadni az emberek életéből és azokat kötni a tér- és idő koordinátáihoz. Ilyen nevezetes esemény az ember születése és halála, amit "megfoghathunk" a két esemény tér- és időkoordinátáival.

keresztül azonosíthatjuk (ezzel eltűnik a nevek és névhordozók közti filozófiai különbség, hiszen a megnevezések mint események és a névhordozók egyaránt individuális entitások). Természetesen ugyanazt a (tulajdon)nevet más és más megnevezéseken keresztül különböző névhordozókra lehet alkalmazni – akár egy, akár több időben. Ez az értelmezés így kezelni képes azt a homonimitási jelenséget, amiről korábban említést tettünk. A megnevezést nem egyszeri aktusként, hanem egy folyamatos megnevezési eseménysorozatként (vagy épp egy megnevezési állapotként) érdemes felfogni. Ez a megoldás azért képes megoldani a konkrét névhordozók azonosítását, mert a megnevezést (mint eseményt vagy állapotot) már hozzáköthetjük valahogyan a tér és idő koordinátáihoz. A megnevezésnek ez a térben és időben történő lokalizálása úgy is értelmezhető, hogy ez teremti meg azt a kontextust az adott (tulajdon)név számára, amely révén utóbbi alkalmassá válhat az egyedi azonosításra.

Eddig még nem volt arról szó, de most már szükséges utalni rá, hogy nevekből és megnevezésekből többféle is lehetséges, vagyis mind a neveket, mind a megnevezéseket tipizálhatjuk. A személynevek esetében beszélhetünk becenevekről vagy álnevekről, hivatalos, születési, felvett, írói stb. nevekről. Ezek mind eltérő megnevezési gyakorlatokhoz és közösségekhez kapcsolódhatnak, és persze az is előfordulhat, hogy többféle névhasználói közösség ugyanazt a nevet tartja fent valakivel kapcsolatban. A megnevezési gyakorlat mint kontextus felfogás tudja azt a jelenséget is kezelni, amikor két megnevezés esetében a névalak ugyan teljesen megegyezik, de az névhasználói kontextuson belül más típusú nevekről van szó.¹⁵

A hosszabb elméleti felvezetés után térjünk rá arra a kérdésre, mi következik mindebből a névterek építésének konkrét gyakorlatára vonatkozóan. Az eddig bemutatott elméleti megfontolásoknak több fontos következménye van.

Abból a tényből, hogy a névtérnek két, egyenrangú komponense a név és a névhordozó (a jelölő és a jelölet), az következik, hogy a névtérben mindkét komponens kezelni kell tudni. Praktikusan ez annyit jelent, hogy mind a névhordozókat, mind a neveket azonosítani kell valamilyen módon, és – talán ez a legfontosabb – harmadik komponensként fel kell venni a névtérbe a köztük levő kapcsolatot is. Annál is inkább szükség van erre, mert szinte mindegyik névhordozótípus (a személyek, a testületek, a földrajzi helyek stb.) névhasználati gyakorlatában nagyon gyakori a szinonimitás jelensége, vagyis az, hogy egy névhordozóhoz több név is tartozik. Ezt úgy kell kezelni, hogy mind a névhordozót, mind a hozzá kapcsolható összes nevet önmagában azonosítjuk,¹⁶ és a megnevezési esemény (gyakorlat) leképezéseként ezek összekapcsolását is annyiszor elvégezzük, ahány névadási esemény (praxis) előfordul az adott névhordozó esetében (hogy erre miért van szükség, rövidesen bemutatjuk Biatorbágy példáján keresztül).

A megnevezést mint kontextust elég csak egyértelműen azonosítani, nem kell alaposabban feltárni (bár szükség esetén lehetséges). Azzal, hogy úgy kapcsoljuk össze a névhordozó (mondjuk a személy) egyedi azonosítóját (personid) és a hozzárendelt névelem egyedi azonosítóját (nameid), hogy ezt az összerendelést is azonosítjuk, magát a megnevezési kontextust ragadhatjuk meg. Nézzük meg Biatorbány példáját! Bia és Torbágy községeket 1958-ban egyesítették, az új településnek 1950 és 1958 között Biatorbágy volt a neve. 1958-ban újra szétvált a két település, és hivatalosan is visszakapták a Bia és Torbágy neveket. Végül 1966-ban újra egyesítették őket, és innentől kezdve megint a Biatorbágy nevet viseli a település. A két korábban önálló település, Bia és Torbágy az idők során kétszer is hivatalosan viselte a saját nevét, kétszer pedig a Biatorbágy hivatalos névvel jelölték az összevont települést. Ez három nevet, de hat megnevezést (névhasználati gyakorlatot) jelent. A nevek szintjén a Bia, Torbágy és Biatorbágy nevek keletkezésük óta léteztek, de a megnevezések/névhasználatok szintjén nem mindig volt így. Ez az eset arra is példát mutat, hogy amikor a megnevezésekkel hozzákapszolunk egy névhordozóhoz egy adott nevet, akkor megtehetjük, hogy – több megnevezés felvételével – ugyanazt a nevet többször is odakötjük ugyanahhoz névhordozóhoz.

¹⁵ Egy település hivatalos nevének megváltoztatása után a régi névhez két – olykor radikálisan eltérő, egymással versengő – névhasználati gyakorlat tartozhat: a hivatalos gyakorlat akadályozhatja, tilthatja a régi név használatát, míg a "laikus", hétköznapi praxis támogathatja azt.

¹⁶ Az azonosítás itt mesterséges azonosító hozzárendelését jelenti.

| | | | |
|----------------|---------------------------------|----------------|---------------------------------|
| 1950.01.01. | 1958.01.01. | 1966.01.01. | |
| Bia | <i>Bia</i> Biatorbágy | Bia | <i>Bia</i> Biatorbágy |
| Torbágy | <i>Torbágy</i> | Torbágy | <i>Torbágy</i> |

egy név, több megnevezés(i esemény)

Természetesen az emberek számára a nevek az elsődlegesek, és számukra a fenti nevek (Bia, Torbágy, Biatorbágy) többszörös használata nem kell, hogy elkülönüljön egymástól, de precíz, gépi és szaktudományi feldolgozás számára ezeket a különbségeket kezelni kell.

A névterekben tulajdonneveket kapcsolunk névhordozókhoz megnevezéseken keresztül. Eddig a névterekről úgy beszéltünk, mint amelyekben csak tulajdonneveket kell nyilvántartanunk. A tulajdonnevek kezelésével kapcsolatos problémák elemzése után érdemes kitekinteni a köznevek kérdéskörére is. Vannak ugyanis olyan nyelvi rendszerek, amelyek közneveket tartalmaznak, és hasznos lehet tisztázni ezek névterekhez való viszonyát.

Korábban elkülönítettük a névterek szűkebb és tágabb értelmezését egymástól, és a szűkebb értelmezés szerinti névtéren a tág értelmezésű névterekben használható tulajdonnevek egyedi felsorolását értettük. Ilyen gyűjteményt összeállíthatunk köznevekből is azért, hogy valamely ismeretterület azzal jellemezzünk, hogy megadjuk a vele kapcsolatos közneveket (fogalmakat) és a nevek (fogalmak) közti levő kapcsolatokat. Ha így teszünk, akkor egyrészt más jellegű feladatot kell elvégeznünk, másrészt eltérő eredményekhez juthatunk. Amíg ugyanis a tulajdonnevek szemantikailag lapos struktúrába szervezhetők, addig a köznevekből több relációval összefogott, olykor gazdag, összetett struktúrát lehet képezni. Olyan fogalmi struktúrát, amely alkalmas lehet a világról való tudásunk (egy részének) reprezentálására.

* * *

A tulajdonnévtér fogalmának meghatározása után érdemes egy további pontosítást tenni, és egy újabb megkülönböztetést tenni a névtér kategóriája kapcsán. Lehetséges ugyanis egy szűkebb értelmezést is adni a tulajdonnévtér fogalmára, aminek használhatósága kisebb hatókörű, mint amire nekünk szükségünk lehet, de bizonyos célok megvalósításához nagyon hasznos eszközt biztosíthatna az újonnan konstruált fogalom.

A fent megadott tulajdonnévtér-értelmezésre azt mondhatjuk, hogy az – nyelvészeti szempontok szerint – tágabb értelemben vett definíció volt. A **tágabb értelemben vett névtér** a tényleges egyedi megnevezések mint események listáját tartalmazza (a kiegészítő adatokkal együtt). Igazán pontosan nem névtérnek, hanem **megnevezéstérnek** kellene neveznünk az ilyen névtérket, de – a ritka és kivételes esetektől eltekintve – e terminus használatától tartózkodni fogunk.

Lehetőség van arra, hogy a fentihez képes szűkebb értelemben vett névtérfogalmat alkossunk. A **szűkebb értelemben vett névtér** a – megnevezési események során kiosztott vagy kiosztható – lehetséges nevek egyedi listáját tartalmazza. A számosságát tekintve utóbbi – bizonyos pontokon – lehet bővebb vagy szűkebb terjedelmű. A szűkebb értelmezés szerinti névtér a megnevezések során kiosztott neveket csak egyszer tartalmazza, ebből a szempontból kevesebb elemből áll a megnevezéstérhez képest, viszont tartalmazhat olyan egységeket is, amelyeket az adott névhasználati gyakorlat a megnevezési események során még nem vett igénybe, így azok nem szerepelnek a megnevezéstérben.

Érdeemes konkrét példával szemléltetni a fenti különbségtételt. Vegyük a személynevek példáját! A dolgokat kissé leegyszerűsítve tételezzük azt, hogy a személynevek két komponensből állnak, egy családnévből és egy keresztnévből. Vegyünk pár konkrét névelemet: 'Kovács', 'Molnár', 'Tóth' mint családnév, illetve 'István', 'Pál', 'József' mint keresztnév. Ezek a nevek egyszer-egyszer szerepelhetnek a szűkebb értelemben vett családnévtérben, illetve keresztnévtérben. akkor, ha külön névtérrel hozunk létre a család- és keresztnéveknek. Ha egyetlen – család- és keresztnévekből álló – személynévtérrel definiálunk, akkor ahány változatban előfordul vagy előfordulhatnak a példának vett család- és keresztnévek egy néven belül ('Kovács István', 'Tóth Pál', 'Tóth József' stb.), annyiféle személynevet kell rögzítenünk a szűknévtérben (persze, csak egyszer). A tágabb – és praktikus szempontból számunkra hasznosabb – értelmezés szerint viszont ahányszor a 'Kocsács István' nevet használták

megnevezésre, annyiszor fel kell vennünk ezt a személynevet a tágabb értelmű névtérbe. Ezzel szemben ha a névré által lefedett megnevezési kontextuson belül még egyszer sem használták a 'Tóth Pál' személynevet, akkor az nem lesz benne a megnevezéstérben, míg a szűknévtérben szerepelni fog.

2.21 Lokális, nemzeti és nemzetközi névtér

A névterek fogalmának elemzését el kell végezni egy eddig még nem említett szempont szerint is. A névtereket valamely névállomány elemeinek egyértelmű azonosítását biztosító rendszerként definiálva mindig valamilyen névhasználói kontextushoz kell kötnünk. Akkor viszont mindig kérdés lehet az, hogy kontextust adó névhasználói közösség milyen szinten létezik. Innen tekintve beszélhetünk lokális, nemzeti és nemzetközi névterekről. A továbbiakban ezek tulajdonságait, egymáshoz való viszonyait fogjuk elemezni.

2.22 A név fogalma, funkciója

Mielőtt a névterek fogalmát újabb elemzésnek vetnénk alá, érdemes összegezni, hogy mit tudunk mondani magáról a név fogalmáról.

A nyelvi konstrukciók konvencionális jelölések, ami azt – is – jelenti, hogy a nyelvi jelekhez önkényes módon kapcsoljuk a jelentéseket. A neveknek két típusa, a köznévi, illetve a tulajdonnévi gazdag, illetve lapos szemantikával rendelkezik. A különbség fő oka az, hogy a tulajdonnevek merev jelölőnek számítanak, míg a köznevek nem, mert utóbbiak referenciája változhat, az előbbieké meg nem. A tulajdonnevek adott kontextusban, adott névhasználói közösség rámutatásán keresztül „létrejött” megnevezések. A megnevezés esemény, minden esemény időhöz kötött (köthető), így a megnevezéseket is időhöz kell (lehet) kötni. A nevek alkalmazása kétféle értelemben sem egyértelmű. A többértelműség létre jöhet a nevek homoním vagy szinoním kapcsolódása révén. Hononima esetén ugyanazt a névalakot, többféle jelentéssel (tulajdonnév esetén többféle referátummal) használjuk, szinonima esetén ugyanazt a jelentést (tulajdonnév esetén ugyanazt a referátumot) többféle névalakkal is összekapcsoljuk. A névvel jelölt, referált dolgot nevezhetjük névhordozónak. A tulajdonnévtér kontextusában a jelölés valamilyen individuum (személy, hely, testület, film, könyv, periodika, cikk stb.). Ebben a kontextusban az – eredeti funkciója szerint azonosításra használt – név helyett ekvivalens módon szokás alkalmazni az olyan terminusokat, mint az azonosító, azonosítószám, cím, műcím, földrajzi cím, térbeli cím.

2.23 A névtér fogalma, funkciója

Először foglaljuk össze röviden azt az eddigiel alapján, mit is jelent a névtér fogalma, mi az alapvető funkciója. Kiderült, hogy kétféle névtérről kell beszélnünk, mert a köznevek és a tulajdonnevek számára egymástól részben eltérő logikájú, más felépítésű névtereket kell fenntartanunk. A két típus között sok a hasonlóság, de van különbség is köztük. Mivel a projektben a tulajdonnévtéren van a fókusz, ezért inkább ezekre koncentrálnunk. Amikor nem tüntetjük fel külön jelzők alkalmazásával, hogy milyen névtérről van szó, akkor tulajdonnévtérről beszélünk. Amennyiben szükség van arra, hogy a megállapításainkat névtértípusok szerint elkülönítve tegyük meg, akkor ezt jelezni fogjuk.

A nyelvi konstrukcióinkkal vagy egyedi dolgokra vagy dolgok halmazára akarunk utalni, és a két cél megvalósítására kétfajta nevet használhatunk. A **tulajdonnevekkel** individuumokra referálunk, a **köznevekkel** individuumok valamilyen gyűjteményére.

A nevekkel referált individuumokat (vagy azok gyűjteményét) **névhordozóknak** nevezzük.

A nevek mindig valamilyen névhasználói **kontextusban** nyerik el a jelentésüket, referenciájukat, minden nyelvi megnyilvánuláshoz szükséges valamilyen kontextus ismerete.

A nyelvhasználat során a nevek könnyen és gyakran többértelművé válnak, és az egyértelműséget csak az adott megnyilatkozás egyedi kontextusa képes biztosítani (a családon belül egyértelmű lehet, hogy a

'Jóska' név kire utal, de egy osztályban már nem. Az osztályban egyértelmű lehet az, hogy a 'Kovács Jóska' név kit azonosít, de az iskolában már nem feltétlen van így).

Minden névhasználat, minden névtér **kontextusfüggő**. A névhasználatnak, a névtérnek mindig van egy **szkópja**, amely azt fejezi ki, hogy a nevek milyen körére terjed ki a figyelem.

A tulajdonnevek másként működnek a köznevekhez képest abban az értelemben, hogy az előbbiek **merev jelölők**, míg utóbbiak nem azok. A tulajdonnevek segítségével bármilyen lehetséges világban ugyanarra az individuumba mutatunk, ez a referálás nem változik meg soha. Ez azért van, mert tulajdonnevekkel mindig olyan konkrét egyedre mutatunk, amely a térben és időben mindig elhelyezhető valahova. A köznevek jelentése, referenciája ezzel szemben mindig változhat a lehetséges világok között, így sosem lehetünk az adott kontextuson belül érvényes értelmezés, jelentés, referencia ugyanaz marad egy másik lehetséges világban, egy másik kontextuson belül.

A neveket nem lehet térhez is időhöz egyértelműen hozzákötni, hiszen a neveket térben és időben folyamatosan újra lehet használni. Egy nevet ugyan lehet példányosítani, de a név példánya (token) és a név általános formája (type) megegyezik. Ezért önmagában a név nem alkalmas egyértelmű referálásra.

A tulajdonnevet úgy tudjuk egyértelmű referálásra használni, hogy nem a tulajdonnevet magát használjuk erre a célra, hanem azt a **megnevezést** (megnevezési eseményt, folyamatot), amelynek keretében egy névhasználói közösség megnevezi egy individuumot, és ezt a nevet (névkiosztást) valamennyi ideig fenntartja. A megnevezési eseményt már lehet példányosítani azzal, hogy az eseményt az idő valamely pontjához kötjük.

A névterek alapfunkciója a nevek és megnevezések, valamint a névhordozók **egyértelműsítése**, vagy másként a nevek és megnevezések gyakorlatban létező többértelműségének megszüntetése, feloldása, és ezzel adott névhordozó (vagy névhordozók gyűjteményének) egyértelmű beazonosítása.

A névterek **nomonima- és szinonimamentes névállományok** abban az értelemben, hogy a hononim és szinonim névalakok mindig egyértelműen fel vannak oldva.

Az egyértelmű azonosításra nyelvi jelek nem alkalmasak (sem a köznevek, sem a tulajdonnevek), ezért helyettük **mesterséges azonosítókat** (számokat) kell használni. Természetesen egy mesterséges azonosító is kontextusfüggő, és ebben az értelemben nem egyértelmű, de a névtér szkópjába tartozó névállományt, illetve a megnevezések állományát mindig lehet egyértelműsíteni mesterséges azonosítókkal.

Az egyedi mesterséges azonosítók kiosztásához és folyamatos használatához szükség van „hagyományos” leíró adatokra is annak érdekében, hogy ezek segítségével az ember is egyértelműen azonosítani tudja a neveket, névhordozókat. Az emberi beazonosításhoz szükséges, hagyományos adatokat nevezhetjük névjegynek, ami a minimálisan szükséges és elégséges adatelemek körét tartalmazza. A **névjegy** tartalma névtértípusonként változhat.

Informatikai értelemben a névtér valamely entitástípus előfordulásainak és ezek neveinek (megnevezéseinek) **normalizálását** jelenti. A tulajdonnévterek – a rájuk jellemző – névjegyekkel kiegészített terminuslisták, a köznévterek leggyakrabban taxonómiák vagy teauruszok.

Amikor egy – neveket is kezelő – adatbázison, adatrendszeren belül normalizálják a névállományt, akkor saját névteret (adatrendszernévteret) hoznak létre, hiszen a normalizált névkezelés az egyértelmű azonosítást szolgálja az adatbázison, rendszeren belül. Ilyen esetben a névtér, az egyértelmű névkezelés kontextusát az adatbázis, adatrendszer határai adják.

Ha egy intézményen belül több névtérrel rendelkező adatrendszer is van, akkor ezek között, ezekre támaszkodva létre lehet hozni **intézményi névteret**, amely valamilyen módon egységesen kezeli az intézményen belüli, különböző adatrendszerek névtereit. Az intézményi névtér szkópja az intézmény összes, névtérrel rendelkező adatrendszerén belül kezelt névállomány egységesítése, egyértelműsítése. Mind az adatrendszerek normalizált állományait, mind az intézményi névtereket nevezhetjük **lokális névtereknek**.

A lokális névterek nevei, megnevezései csak saját határaikon, saját szkójukon belül egyértelműek. A nemzeti kulturális adatvagyon egymástól elkülönült intézmények adatrendszereiben egymástól elkülönülten van tárolva. Még ha mindegyik kulturális intézmény fel is építi saját lokális névterét, az adatrendszerek közti semmi sem biztosítja az átjárhatóságot, interoperabilitást mindaddig, amíg a lokális névtereket nem kötik össze egymással. Ennek a kapcsolatnak a megvalósulását biztosíthatja a **nemzeti névtér**.

A nemzeti névterek szkópját a nemzeti kulturális adatrendszereken belül létező névállományok adják, de a nemzeti névtérbe csak a kulturálisan jelentős neveket és névhordozókat érdemes „betenni”. Arra kell felkészülni, hogy a közgyűjteményekben mindig maradnak, maradhatnak olyan nevek, névhordozók, amik nem kerülnek be a nemzeti névtérbe.

A nemzeti névterek építésekor és fenntartásakor a nevek és névhordozók egyértelmű azonosításának feladata mellett fontossá válik a lokális/intézményi névterek nevei, pontosabban azok lokális azonosítói közti kapcsolatok megteremtése.

A **nemzetközi névterek** a nemzeti – ma még sok esetben: az intézményi – névterek azonosítói közti kapcsolat megteremtését végzik.

A **tulajdonnévtér** adott adatkörön belül adott névtípusba tartozó megnevezések és a nevekkal jelölt névhordozók egyértelmű azonosítását végző adatrendszer (megnevezések és megnevezettek egyértelmű azonosítását végző adatrendszer).

A **köznévter** adott tudásterületre vonatkozó általános fogalmak jelentését (ezáltal a világra vonatkozó tudásunkat), a fogalmak közti kapcsolatok struktúráján keresztül, adott nyelven kifejező adatrendszer

2.24 Lokális névterek

A lokális névterek lehetnek tulajdonnévterek és köznévterek egyaránt, utóbbiak struktúrája lehet többféle. A meghatározó minőség a névtér szkójában van. Akkor beszélhetünk lokális névtérről, ha a névtér egy adatkör vagy egy intézmény névállományát kezeli. Technikai értelemben az a kérdés, hogy egyetlen névtér létezik-e, vagy sem (azonos névtípus esetén). Egy adatrendszer számára nyilván csak egy névteret lehet építeni (adott névtípus esetén). Ha egy intézményen belül több adatrendszer is létezik, amelyek névállományaiból névteret lehet építeni (adott névtípus esetén), akkor a névtér fenntartására kétféle megoldás képzelhető el.

Lehetséges, hogy az egyes adatrendszerek saját névteret tartanak fent, és köztük egy összerendelési munka eredményeként teremtik meg az átjárhatóságot. Ekkor az intézményi névtér és az adatrendszer-névterek különböznek egymástól, és az intézményi névtér ugyanazt a szerepet tölti be, mint a nemzeti névtér a lokális névterek között.

A másik lehetőség az, hogy a különböző adatrendszerek egyetlen névteret építenek közösen. Ebben az esetben biztosítani kell azt, hogy a névtérépítés során minden pillanatban látható legyen a névtérben tárolt névállomány minden eleme, hogy teljesíthetők legyenek az elemek egyértelműségére és egediségére vonatkozó elvárások.

A ténylegesen létező lokális tulajdonnévtereket nem nagyon lehet látni, meghivatkozni, mert ezek az intézmények falai között léteznek. A könyvtárak besorolási állományai ilyen lokális névtereknek tekinthetők. Amennyiben egy nemzetközileg látható névtér nem igényli magának azt a státust, hogy a saját azonosítóit más, alsóbb szintű névterek elvigyék és saját rendszerükbe építsék, hogy az intézményközi interoperabilitás alapjává tegyék, akkor igazából csak lokális névtérről beszélhetünk. A földrajzi neveket kezelő Getty földrajzi nevek tezauruszát vagy az OpenStreetMap térkép- és névszolgáltatását – ebből a szempontból – ilyennek, nemzetközileg látható, lokális névtérnek tekinthetnénk.

A lokális köznévterekre példa lehet minden olyan, tartalmi feltárást támogató, tárgyszó-hozzárendelő munkát támogató névállomány, amelyet valamilyen kontroll alatt tartanak.

2.25 Nemzeti névterek

A nemzeti névtér fogalmának meghatározásakor több szempontot is figyelembe venni. Egyfelől a névterek egymáshoz való viszonyára fókuszálva a nemzeti névtereket szembe lehet állítani a lokális névterekkel. Másfelől a névterek által kezelt adatkörök tartalmára fókuszálva lehet szűkebb és tágabb értelemben vett nemzeti névterekről beszélni.

2.25.1A nemzeti névtér szűkebb értelmezése

Amikor adott adatrendszeren belül valamely névállományt egyértelműsítünk (normalizálunk), akkor ez a normalizált névállomány nem önmagában áll, hanem az adatbázis tartalmának adott szegmensként funkcionál. Egy lokális névtér mindig valamilyen adatrendszer kiegészítője és/vagy része. A lokális névtér fő funkcióját, a szkópjába tartozó nevek egyértelműsítését azáltal biztosíthatja, hogy minden névhez és minden névhordozóhoz mesterséges azonosítót rendel. Ez az egyedi azonosító egyértelműséget teremt a lokális névtéren belül, de ha kilépünk ebből kontextusból, akkor semmi sem biztosítja ezt az egyértelműséget tovább. Más adatrendszerek, más lokális névterek is kioszthatják ugyanazt az azonosítót, és akkor már csak azon az alapon lehetne megkülönböztetni a két – numerikusan teljesen egyező – azonosítót, ha tudjuk, melyik adatrendszerből, névtérből származik. Ha az intézményeknek, névtereknek is adnánk egyedi azonosítót, akkor a két azonosító (a névtér- és a névazonosító) összekapcsolásával már újra egyértelműséget teremthetnénk a mesterséges azonosítók szintjén, de ezzel a megoldással semmit sem tudnánk magukról a nevekről, névhordozókról, azok különbségeiről, azonosságairól. Ez abban a pillanatban problémát jelentene, ha feltételezhetnénk, hogy a névterek állományai közt lehetnek átfedések. Márpedig a kulturális adatrendszerek világában pont ez a helyzet. A közgyűjtemények által kezelt névállományokban gyakran komoly átfedések vannak, hiszen a kulturális élet szereplői, intézményei, helyszínei ismétlődően felbukkannak a kulturális dokumentumokat őrző, egymástól ilyen-olyan szempontok szerint elkülönített közgyűjtemények adatrendszereiben.

A kulturális értékek iránti érdeklődés nagyon sokszor intézményközi, amit az analóg világban kialakított intézményi különültség keretei közt nem lehet jól kiszorgálni. A kultúra szerteágazó jellege és az intézményi különültség merev határokat húzó logikája közötti inkompatibilitást a predigitális világban nem igazán lehet enyhíteni, megszüntetni. A hálózat elterjedése megteremti ugyan annak lehetőségét, hogy minél többen minél több kulturális dokumentumhoz hozzájussanak, ám még a közgyűjteményeink teljeskörű digitalizálása sem oldaná fel az adatrendszerek közti átjárhatatlanságot, amit a – predigiális korban kialakult, és akkor még szükségszerűnek mondható – intézményi különültség okoz.

Mi lehet a megoldás erre a problémára? A nemzeti névtér létrehozása, a nemzeti névtér azonosítóinak összerendelése a lokális névtér-azonosítókkal. Mi a nemzeti névtér, miért van rá szükség, hogyan működik? Nézzük végig, milyen válaszok vannak ezekre a kérdésekre.

A lokális névterek egyértelműséget teremtenek a lokális adatrendszereken belül, de az egyértelműsítésre használt mesterséges azonosítók intézményenként különböznek, és nem segítenek a két névtér közötti kapcsolat kialakításában.

A hálózaton keresztül hozzáférhető – kulturális, közgyűjteményi – adatbázisok egymást átfedő szegmenseiben a legnagyobb részt azok a dolgok, entitások (névhordozók) adják, amelyek névterekkel egyértelműsíthetők. A kulturális „közös részbe” tartoznak a kulturális dokumentumok (könyvek, filmek), a kulturális ágensek (személyek, testületek), a kulturális helyszínek.

Kulturális befogadói oldalról kívánatos cél, és ez alapján a közgyűjteményekkel szemben megfogalmazható elvárás az, hogy a kulturális adatbázisok legyenek átjárhatók (interoperábilisek). Az átjárhatóság legfontosabb eszköze a közös névterek kialakítása és használata. A lokális egyértelműsítés, lokális azonosítás nem segít ebben, csak egy rávezető lépés lehet, mert a különböző forrásokból származó, lokális szinten egyértelműsített névállományok nemzeti szinten átfedhetik egymást. A megoldást a nemzeti névtér közös használata jelenti.

A nemzeti névtér egymástól elkülönült – kulturális – adatrendszerek tartalmilag átfedő, adott névtípusba tartozó névállományainak együttes kezelésére alkalmas, nemzeti szinten egyértelműsítő, azonosítási rendszer. A nemzeti névtér kialakítása megtörténhet a lokális névterek, a lokálisan elkülönült névállományok, a lokális azonosítók összerendelésével, összefésülésével vagy közösen kezelt online névállomány működtetésével. Hosszútávon lehetnek nemzetinévtér-használó és részlegesen összefésült lokálisnévtér-használó együttműködő partnerek. A nemzeti névtér esetében nem működőképes a hagyományos értelemben vett adatcsere művelete, mert a különböző forrásokból származó, lokális szinten egyértelműsített névállományok nemzeti szinten átfedhetik egymást. Az új nevek és névhordozók azonosítása egyelőre csak emberi közreműködéssel végezhető.

Első megfogalmazásban – szűkebb értelemben – azt mondhatjuk, hogy a **nemzeti névtér** adott adatkörön belül, adott névtípusba tartozó megnevezések és a nevekkal jelölt névhordozók egyértelmű

azonosítását nemzeti szinten végző, és az adatokat publikáló adatrendszer (megnevezések és megnevezettek egyértelmű azonosítását nemzeti szinten végző adatrendszer).

Mire jó a névtér? A kulturális adatbázisok építése során felmerülő nevek, névhordozók (személyek, földrajzi helyek, testületek stb.) egyértelmű – lokális, intézményi szintű – azonosítására. Mire jó a nemzeti névtér? Egyfelől a különböző kulturális adatbázisokon belül használt nevek, megnevezések, névhordozók egyesített és egységesített azonosítására, másfelől a lokális névterek azonosítóinak összefésülésére. Ezek alapján a lokális és nemzeti azonosítókon keresztül mindkét irányban átjárhatóságot lehet teremteni azáltal, hogy a nemzeti névtér keretében a lokális névterek közt felépített kapcsolatok alapján fel lehet kínálni a lokális adatrendszerek felé mutató linkeket, ami a látogatók számára lehetővé teszi, hogy egy-egy kattintással további adatokhoz juthasson az őt érdeklő kulturális jelenséggel kapcsolatban.

2.25.2A nemzeti névtér tágabb értelmezése (névtár)

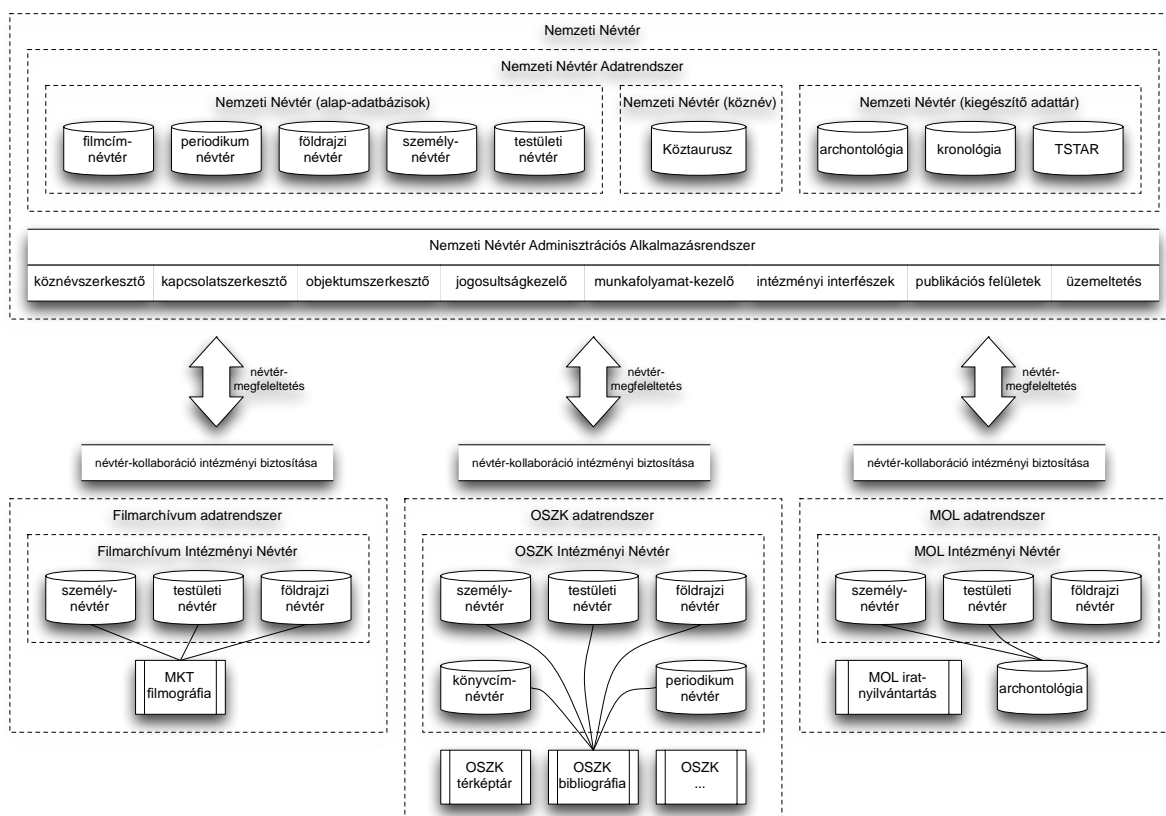
A fentiekben a nemzeti névtér fogalmának szűkebb értelmezéséről beszéltünk. A továbbiakban bemutatunk egy tágabb értelmezést, amely kitér a nemzeti névtér fogalmát a szolgáltatott adatkörök tartalma alapján.

A nemzeti névterekbe – első megközelítés alapján – a kiválasztott névhordozók, névtípusok névállományai tartoznak. A nemzeti névterekben együttműködő közgyűjtemények által épített kulturális adatrendszerek tartalma ennél természetesen sokkal bővebb. A lokális és nemzeti névterek építésekor, a kulturális tudás rendezésekor olyan – kulturális érdeklődésre számot tartható – adatok is keletkeznek, amelyek a különböző névtértípusok közti kapcsolatként értelmezhetők. Ilyenek – többek között – a könyvek és személyek közti kapcsolatokat leíró személyes bibliográfiai adatok, a filmek és személyek közt létező filmográfiai adatok, a személyek és testületek közti viszonyt megragadó archontológiai adatok, a testületek (iskolák) és személyek (diákok) közti tartós viszonyokat számon tartó iskolai adatok, de említhetjük még a földrajzi helyekhez kapcsolható térstatisztikai adatköröket, a testületek (pártok), személyek (jelöltek) és földrajzi helyek közötti kapcsolatokat ábrázoló választási adatokat. A kulturális adatbázisok, a névterek építése során ezek az adatok vagy „automatikusan” előállnak vagy gyorsan és könnyen előállíthatók, és így könnyen szolgáltatathatóvá is tehetők. Ezt a lehetőséget érdemes kihasználni és a nemzeti névtér fogalmát ebben a tágabb értelemben definiálni. Ezáltal sokkal több értékes adat kerülhet ki a nemzeti névterek felületeire. A nemzeti névterek értelmezését a névterek közti kapcsolatokkal kibővítve jelentősen növelni lehet a névterek, közgyűjteményi adatrendszerek iránti érdeklődést.

Második megfogalmazásban – tágabb értelemben – azt mondhatjuk, hogy a **nemzeti névtér** adott adatkörön belül, adott névtípusba tartozó megnevezések és a nevekkal jelölt névhordozók, valamint a különböző névterek közti kapcsolatok egyértelmű azonosítását nemzeti szinten végző, és az adatokat publikáló adatrendszer (megnevezések és megnevezettek, valamint a köztük levő kapcsolatok egyértelmű azonosítását nemzeti szinten végző adatrendszer).

A nemzeti névtér tágabb értelmezésű fogalmára használhatjuk a **nemzeti névtár** kifejezést.

A tágabb értelemben vett névtérfogalom alapján elképzelhető névtérrendszer nagyobb szegmenseit, jól elkülöníthető funkcionális egységeit, a potenciálisan együttműködő partnerek helyét mutatja az alábbi ábra.



a szélesebb értelemben vett nemzeti névterek kapcsolatrendszerének sematikus ábrája

2.26 Globális névterek

Korábban röviden bemutattunk több globális névteret is, bár akkor nem hangsúlyoztuk azt, hogy azok nemzetközi névterek lettek volna. Az említettek között voltak dokumentumokat azonosító rendszerek, mint az ISBN, ISSN vagy az ISAN, a dokumentumok létrehozásában, hozzáférhetővé tételében érintett szereplők, ágensek, helyszínek azonosítását végző rendszerek, mint az ISNI vagy a VIAF, a valós földrajzi helyek, földrajzi nevek világában való eligazodás segítő földrajzi névtérrendszerek, mint a Geonames szerver, a Getty földrajzi nevek tezaurusza, valamint a Google Maps vagy az OpenStreetMap mint térképszolgáltató. Ezek között vannak, amelyek nemzetközi szabványok is egyben (ISNI, ISBN), így egyfajta jogi támogatás is van mögöttük, de több példa is mutatja azt, hogy ez a jogi támogatás nem nélkülözhetetlen feltétele a nemzetközi elismerésnek, a nemzetközi használatbavételnek.

A globális névterek ugyanazt a koordinációs funkciót töltik be a nemzetközi szinten a nemzeti és nemzetközileg jelentős intézményi névterek világában, mint amit a nemzeti névterek képviselnek a nemzeti szinten, az „alájuk tartozó” lokális, intézményi névterek világán belül. A nemzetközi egyértelműsítő és azonosító rendszereknek az az alapvető célja, hogy egységesítsék, „magasabb szintre” emeljék a nemzeti, regionális szintek kontextusán belül már egyértelműsített nevek azonosítását. Amilyen mértékben sikeresek e tevékenységükben, olyan mértékben válhatnak egyre jobban összekapcsolttá, átjárhatóvá a nemzeti, regionális, lokális szinten létező adatrendszerek. Az persze nagy kérdés, hogy vajon a nemzetközi névterek milyen mértékben fogadják, fogadhatják be a nemzeti, regionális, lokális szintekről érkező adatokat. Bár nem elvileg lehetetlen, de gyakorlati megfontolások miatt nem tűnik valószínű opciónak az, hogy a nemzetközi névterekbe minden névadat bekerüljön a nemzeti névterek szintjéről. Ugyanolyan érvek alapján, mint ahogy a nemzeti névterek és a lokális névterek közt is feltételeznünk érdemes egyfajta szűrő létezését, amely nem enged be minden adatot a „központba” az alacsonyabb, lokális szintekről. Ez a szűrési, kanonizálási elvárás annak köszönhető, hogy minél magasabb szinten akarunk valamilyen közösségi élményt kialakítani, fenntartani, annál

biztosabban meg kell szűrünk, hogy milyen „elemeket” teszünk meg, engedünk meg közösségképző tényezőnek.

A globális névterekhez való csatlakozást, a nemzetközi névterekkel való érdemi kooperációt akkor érdemes komolyan venni, amikor a nemzeti névterek már elindultak, működnek, kialakították a lokális, intézményi névterekkel a kétirányú kapcsolataikat.

A személyneveknél és testületi neveknél az ISNI és a VIAF a potenciális partner, a földrajzi nevek esetében pedig az OSM.

2.27 A névtér fogalmának ontológiai elkötelezettségei

Amikor névterekről beszélünk, amikor névtereket építünk, fontos tisztázni, hogy milyen ontológiai elköteleződések teszik a rendszer modelljének felállításakor. Ez a feltárási munka egyfelől segíti a fejlesztési munkát azzal, hogy értelmezhetővé, egyértelművé teszi a modellben rögzített struktúrát, másfelől láthatóvá, megérthetővé és kezelhetővé teszi azokat a következményeket is, amelyek az alapvető ontológiai elköteleződéseinkből fakadnak.

Mielőtt a névterekkel kapcsolatos fejtegetésekbe vágnánk, be kell vezetnünk kell egy fontos fogalompart, amelyre a névterek értelmezésekor bizonyos pontokon hivatkoznunk kell majd.

2.28 A természeti és társadalmi tények különbsége

A társadalomtudományi diskurzusban régóta ismert az a megkülönböztetés, amely szerint a világról gondolkozva kétfajta tényről beszélhetünk, és rájuk utalva kétfajta tényítéletet fogalmazhatunk meg: vannak **természeti tények** és vannak **társadalmi tények** (Searle 1995). Mindkét típus esetében tényállításokat fogalmazhatunk meg, amiket lehet vitatni, módosítani, elutasítani vagy lehet elfogadni. Mindkét típusba tartozó tényállítást élesen el kell (el lehet) határolnunk az értékítéletektől. Mindkét fajta tény képes meghatározó erővel bírni az életünkre. A különbség köztük abban van, hogy a társadalmi tények emberi akartoktól függenek, míg a természeti tények nem. A társadalmi tények társadalmi konstrukciók, míg a természeti tények nem azok. Az a tény, hogy van valahol egy folyó, természeti tény, hiszen ha nem élnének a környékén emberek, akkor is ott folyana, ellenben az a tény, hogy a folyónak 'Duna' a neve, társadalmi tény, hiszen emberek nevezték el így. Az, hogy a folyó egyik oldalán egy hegy magasodik, míg a másikon síkság terül el, természeti tény, viszont az, hogy a helyvidékes oldal az egyik, a síkvidékes rész egy másik ország területének számít, már társadalmi tény.

A természeti, illetve társadalmi tények szembeállítását további terminuspárok segítségével is ki lehet fejteni. Vannak szakterületek, ahol erre a kettősségre a **SNAP** vs. **SPAN** minősítéseket (Grenon & Smith 2004), máshol és máskor a **bona fide** vs. **fiat** ellentétpárt alkalmazzák (Smith & Varzi 2000). Utóbbi fogalompart például a földrajzi entitások világában található kettősségre húzták rá. Eszerint a fiat objektumok (mint például a közigazgatási határok, a műutak vagy a településnevek) emberi akaratok termékei, míg a bona fide objektumok (mint például a folyók, a hegyvidékek, a hágók vagy a tavak) nem.

A tények kettős természetét figyelembe vevő fogalmi distinkció azért fontos, mert a névtereken belül is szembesülhetünk ezzel a kettősséggel, és ügyelni kell arra, hogy kezelni tudjuk az esetleg ebből fakadó problémákat.

2.29 Névterek és tudásszervezési rendszerek

A névtér fogalmának elemzésekor kiderült, hogy a köznévteret és a tulajdonnévteret el kell választani egymástól. Ez a különbség megnyilvánul abban is, hogy más és más választ adhatunk arra a kérdésre, hogy milyen tudásszervezési rendszernek minősíthetünk egy konkrét névteret. Amikor a tudásszervezési rendszereket bemutattuk, a rendszer elemeire építhető struktúrák különbségei alapján tipizáltuk őket. A névterek jellemzésekor is ezt a szempontot kell érvényesítenünk. A könnyebb kérdés (melyik tudásszervezési rendszer fogalma kell besorolnunk a tulajdonnévtereket) elemzésével kezdjük a

mondandónkat, majd ezután kezdünk bele a második kérdés megválaszolásába (melyik KOS-típusba tartoznak a köznévterek).

2.29.1A tulajdonnévterek struktúrája

A tulajdonnévterekre nagy általánosságban azt mondhatjuk, hogy azok a legegyszerűbb tudásszervezési rendszer, a terminuslista fogalma alá sorolhatók be, de típusonként változó módon lehet valamivel összetettebb struktúrájuk. A tulajdonnévterek elemei sokféle relációban állhatnak más tulajdonnévterek elemeivel, de ezek nem struktúráként, hanem a rendszerek közti dinamikus kapcsolatrendszerként értelmezhetők.

2.29.1.1 A személynévtér struktúrája

A személynévtér tiszta terminuslista. A természeti tények alapján nem igazán tudunk struktúrát érvényesíteni az elemei között. A nemi jelleg és a rassz alapján lehet tipizálni az embert, de ezen jegyek alapján nagyon lapos struktúrákat kaphatunk.

A társadalmi tények alapján többféle tipizálásra nyílik mód, és itt már előfordulhatnak olyan köznévsztruktúrák, amelyek alá a személyek besorolhatók a példány reláció révén. A foglalkozási besorolás kiváló példa erre. A foglalkozások listájából generikus hierarchiát lehet építeni és ezek végpontjaiba el lehet helyezni a konkrét személyeket. Látni kell azonban, hogy ebben az esetben nem a személyek, személynevek halmazán építhető fel a generikus struktúra, hanem föléljük rendelhető – a foglalkozási szerepeket leíró – köznevek halmazán.

- 0. Főcsoport: Fegyveres szervek foglalkozásai
- 1. Főcsoport: Gazdasági, igazgatási, érdek-képviselési vezetők, törvényhozók
- 2. Főcsoport: Felsőfokú képzettség önálló alkalmazását igénylő foglalkozások
- 3. Főcsoport: Egyéb felsőfokú vagy középfokú képzettséget igénylő foglalkozások
- 4. Főcsoport: Irodai és ügyviteli (ügyfélkapcsolati) foglalkozások
- 5. Főcsoport: Kereskedelmi és szolgáltatási foglalkozások
- 6. Főcsoport: Mezőgazdasági és erdőgazdálkodási foglalkozások
- 62. Csoport: Erdőgazdálkodási, vadgazdálkodási és halászati foglalkozások
 - 621 Erdőgazdálkodási foglalkozások
 - 6211 Erdészeti foglalkozású
 - 6212 Fakitermelő (favágó)
 - 622 Vadgazdálkodási foglalkozások
 - 6220 Vadgazdálkodási foglalkozású
 - 623 Halászati foglalkozások

7. Főcsoport: Ipari és építőipari foglalkozások
8. Főcsoport: Gépkészítők, összeszerelők, járművezetők
9. Főcsoport: Szakképzettséget nem igénylő (egyszerű) foglalkozások

részlet a foglalkozások egységes osztályozási rendszeréből (FEOR)

Hasonlóképpen lehet a testületi tagság relációja mentén kapcsolatokat találni a személyek és testületek között, és a testületre felépíthető egyfajta típushierarchia, de ebben az esetben egyfelől a konkrét személyek és konkrét testületek közti viszony individuális, partikuláris lesz, másfelől megint csak nem a testületek mint individuumok, hanem a fölük rendelhető testülettípológia tartalmazhat valamiféle generikus struktúrát.

A társadalmi tények önkényes konstruáltságából fakadóan olykor a személynévtér esetében is adódhatnak értelmezési problémák. Amikor valamely testületi tagságot mint társadalmi tényként kell rögzíteni a személynévtérben belül,¹⁷ akkor az emberi akaratokra való hivatkozás – mint a társadalmi tény alapja és ontológiai minősége egyben – olykor ellentmondásos eredményekhez, versengő interpretációkhoz vezethet. Szemléltető példaként hivatkozhatunk itt az akadémiai tagság intézményére. Politikai okok miatt a Magyar Tudományos Akadémia 1949-es átszervezésekor sok akadémikust megfosztottak akadémiai tagságától, akiket aztán az újabb politikai fordulat után, 1989-ben rehabilitáltak. Kérdés, vajon mit mondhatunk a kizárt, majd visszavett akadémikusok tagságáról az 1949-1989 közötti időszakra vonatkozóan. Ebben a negyven évben biztosan nem tartották őket akadémikusoknak a kortársaik, de vajon 1989 után mit mondhattak az 1989 előtti időszakra vonatkozóan? Lehet-e azt mondani, hogy „minden másként volt”, lehet-e a múltat átértelmezni? A válaszhoz ketté kell választanunk a kérdést.

Az természetesen mindig lehetséges, hogy a múltat átértelmezzük, amennyiben új információ birtokába jutunk. Ez mind a természeti, mind a társadalmi tények esetében elképzelhető. A most vizsgált esetben azonban nem erről van szó. Az akadémiai tagság társadalmi konstrukció, a „kinevezés” az erre hivatott emberek, testület akaratán múlik, és ebben az esetben az az igazi kérdés, hogy vajon lehet-e visszamenőlegesen akaratot érvényesíteni. A válasz nyilván az, hogy nem. Nem mondhatjuk tehát 1989 után, hogy a rehabilitált akadémikusok 1949-1989 között is akadémikusok lettek volna, mert visszamenőleg nem akarhatjuk ezt. Az akarat mindig csak a jövőre irányulhat. Az viszont már egy másik kérdés, hogy a rehabilitáció részeként visszamenőleges hatállyal is lehet rendelkezni, és – például – a szolgálati időbe bele lehet számítani a „kiesett” időszakot is. Ez azért lehetséges, mert ebben az esetben új akaratképzések mentén, „most” hozunk létre egy olyan – új – társadalmi konstrukciót, amely a múltat máshogy veszi figyelembe.

2.29.1.2 A testületi névtér struktúrája

A testületi névtér esetében egy fokkal komolyabb tipizálási lehetőség van a személynévtérhez képest, bár ez teljes mértékben a társadalmi tények világában érvényesíthető. A testületek egyértelműen társadalmi konstrukciók, így szinte minden, velük kapcsolatos tény társadalmi ténynek tekinthető. Talán csak a tevékenységük tér- és időfüggése minősíthető természeti ténynek, minden más adat velük kapcsolatban társadalmi konstrukció. A testület telephelye, megalakulásának időpontja, a nevének változtatása stb. a tér és idő dimenzió konkrét pontjára mutat, ami természeti tény. Az a fogalmi struktúra azonban, amelyben a testületek tipizálhatók, egyértelműen a társadalmi tények világába tartozik.

Itt is igaz, amit a személyekkel kapcsolatban már említettünk: a testülettípológiát generikus hierarchiaként is felfoghatjuk, de ez a struktúra nem a individuális testületek halmazán, hanem a köznevekkel kifejezhető testülettípusok halmazán értelmezhető.

¹⁷ Pontosabban a személynévtér és a testületi névtér közötti kapcsolati rendszeren belül.

A társadalmi tények már említett másik nagy halmaza a testületek és a személyek közti kapcsolatokat írja le, amelyek egy részét az archontológiai adatok, a másikat az iskolai adatok adják. Ezek mind partikuláris adatok, a tér és idő valamely szegmenséhez köthetők.

A testületek, intézmények, szervezetek nyilvántartását testülettípusonként más helyen, más módon végzik. Más szabályozási rend vonatkozik a civil társadalom társadalmi szervezeteire, a politikai, érdekképviselői szervezetekre, a gazdasági szervezetekre, az államigazgatási szervezetekre vagy az oktatással, neveléssel, illetve kulturális értékeink megőrzésével foglalkozó szervezetekre.

Mivel a testületek lényegük szerint társadalmi konstrukciók, létezésük emberi akaratokra vezethető vissza, ontológiai elköteleződés szempontjából meghatározó, hogy milyen választ adunk arra a kérdésre, mi biztosítja a testületek identitását, mire hivatkozva lehet az azonosságukat tételezni a való életben. A testületeket jellemzi az a tevékenység, amit a tagjai végeznek, azonban mind a testület tagjai, mind a testület által végzett tevékenység változhat az időben. Meddig „él” és meddig azonos önmagával egy testület? Ha egy focicsapat, mondjuk, a Barcelona minden tagja kicserélődik, attól még nyilván ugyanarról a testületről beszélhetünk. De vajon igaz-e ez a könnyűzenei zenekarokra is? Elképzelhető-e, hogy a Beatles minden tagja más, mint az induláskor volt, és még mindig ugyanarról a zenekarról beszélünk?

Másfajta kérdés merül fel a testületek időhöz való viszonyával kapcsolatban. Ha egy testület egy ideig szünetelteti a tevékenységét, majd újrakezdi, akkor mennyi idő eltelté után mondhatjuk azt, hogy már más vagy még ugyanazon testületről van szó? Magyarországon 1948-ban egyesült a Kommunista Párt és Szociáldemokrata Párt, 1949-ben beolvadtak a Hazafias Népfrentba az akkor még létező politikai pártok, és négy évtizeden keresztül nem szerepeltek a politika színpadán, majd 1989 környékén egyre-másra bukkantak fel a régi neveiket használva, a jogfolytonosságot deklarálva a régi/új politikai formációk. Megjelent a színen a Szociáldemokrata Párt vagy a Független Kisgazdapárt. Ezek a testületek ugyanazok voltak, ugyanazok maradtak annak dacára, hogy negyven éven át semmilyen életjelt nem adtak magukról? Milyen szempontok alapján, hogyan lehet ilyen kérdésekben dönteni?

Másféle problémákba ütközhetünk a testületek tevékenységi körében, hatáskörében beállt változások és a testületek azonosságának kérdésével kapcsolatban. Magyarországon 2010 után az EMMI-be olvasztották be a korábban önállóan létező Oktatási és Kulturális Minisztériumot, akkor utóbbi létezik 2010 után is, csak éppen egy nagyobb szervezet (az EMMI) része, vagy 2010 óta nincs ilyen testület? Az oktatás területéért 1848 óta sokféle elnevezésű és sokféle feladatkörű minisztérium volt felelős, ahogy azt az alábbi áttekintés mutatja.

- Vallás- és közoktatásügyi Minisztérium (1848–1919)
- Közoktatásügyi Népbiztosság (1919-1919)
- Vallás- és közoktatásügyi Minisztérium (1919–1951)
- Közoktatásügyi Minisztérium (1951–1953)
- Oktatásügyi Minisztérium (1953–1957)
- Művelődésügyi Minisztérium (1957-1974)
- Oktatásügyi Minisztérium (1974–1980)
- Művelődési Minisztérium (1980-1990)
- Művelődési és Közoktatási Minisztérium (1990-1998)
- Oktatási minisztérium (1998–2006)
- Oktatási és Kulturális Minisztérium (2006-2010)
- Nemzeti Erőforrás Minisztérium (2010-2012)
- Emberi Erőforrások Minisztériuma (2012-)

Azt talán még mondhatjuk, hogy az Oktatási Minisztérium és az Oktatásügyi Minisztérium vagy a Művelődési Minisztérium és a Művelődésügyi Minisztérium csak a nevükben különböznek, de ugyanazt a testületet jelölik, de igaz-e az Oktatásügyi Minisztérium és a Művelődési Minisztérium vagy a Nemzeti Erőforrás Minisztérium esetében?

Ezekre a kérdésekre nem tudunk előzetesen, mindenre kiterjedő, általános érvényű döntéseket, elköteleződéseket hozni, az ilyen – szerencsére azért nem túl gyakori – kérdéses esetekben egyedileg kell majd dönteni a névtérépítés során.

2.29.1.3 A földrajzi névtér struktúrája

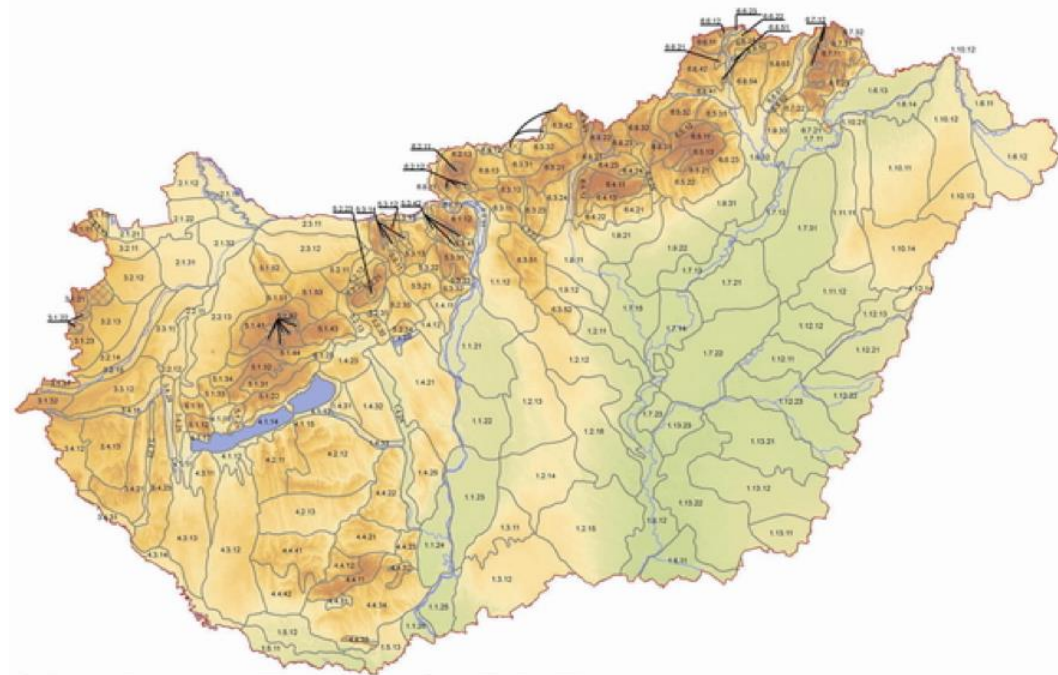
A földrajzi névtér az a tulajdonnévtér, amelynek „komolyabb” struktúrája van (lehet). Egyfelől a földrajzi helyekre többfajta tipizálási lehetőség létezik (a generikus reláció mentén). Első körben – a természeti tények szintjén – meg lehet adni, hogy milyen geometriával jellemezzük az adott földrajzi helyet (lehet pontszerű, vonalszerű, poligonszerű). Még mindig a természeti tények szintjén, de egy másik fogalmi dimenzióban jellemezve a geoindividuumokat ezt a geometriai besorolást alá lehet bontani természetföldrajzi és vízrajzi fogalmak segítségével. Ezek között a fogalmak között fel lehet építeni egy – nem túl mély és nem túl széles, de azért létező – generikus hierarchiát.

A földrajzi helyek tipizálhatók a társadalmi tények alapján is. A geometriai jellemzők egyértelműségével szemben azonban itt megszűnik az egyértelműség. A társadalmi tények alapján definiált földrajzi helyek többértelművé teszik köztük levő kapcsolatrendszerét, hogy megengedik többszörös hierarchiák létezését.

Ha veszünk egy adott geometriával leírt – tehát természeti tényekkel jellemzett – lokációt, akkor arra mondhatjuk azt, hogy önálló falu, mondhatjuk azt, hogy településrész, miután beolvadt a szomszéd városba, de lehetséges az is, hogy ugyanaz a geometria választási szavazóközvetet, egyházközséget vagy filiát jelöl. Ezek a minősítések már társadalmi tényeknek felel meg. Nagyon sok szempont szerint feloszthatunk egy adott földrajzi területet, és a különböző felosztási logikák eltérő szerkezeteket eredményeznek. Vegyünk számba pár lehetséges szempontot (néhány alájuk tartozó fogalommal együtt), és szemléltessük a az eltérő szempontokból adódó felosztások területi következményeit.

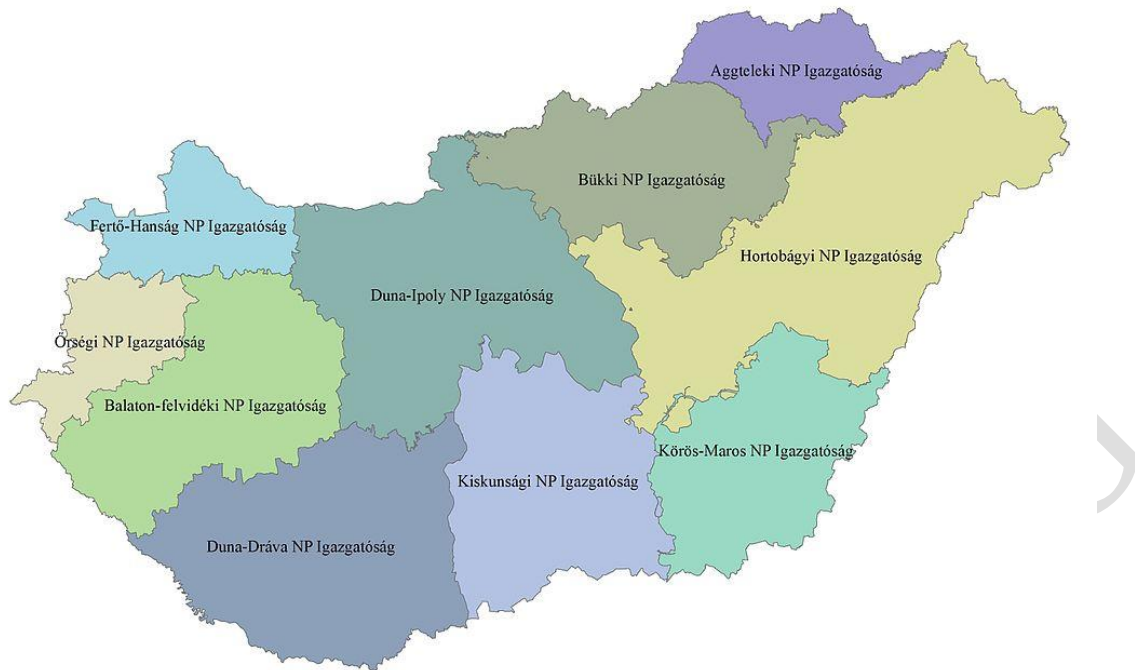
Magyarország területi felosztása:

- földrajzi tájak szerint (nagy-táj, középtáj, kistáj)



Magyarország kistájai

- természetvédelmi körzetek szerint (nemzeti park, tájvédelmi körzet, természetvédelmi terület, tanösvény)



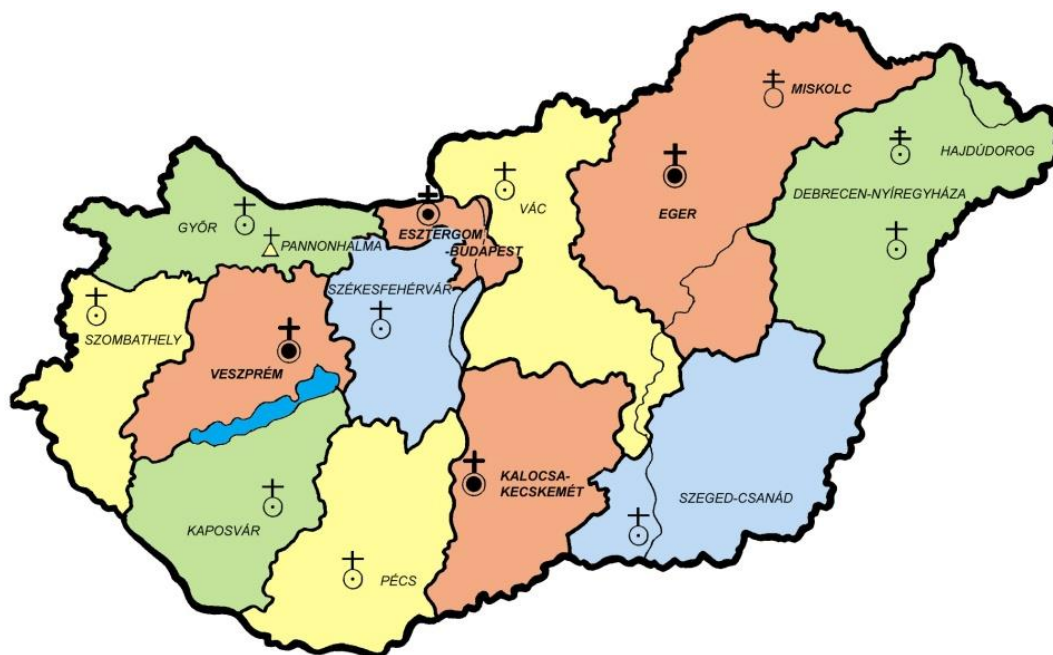
Magyarország nemzeti parkjai

- államigazgatási körzetek szerint (ország, országrész, megye, járás, település, településrész)



Magyarország megyéi

- egyházi körzetek szerint (világegyház, egyháztartomány, egyházmegye, esperesség, egyházközség/plébánia, leányegyház/filia)



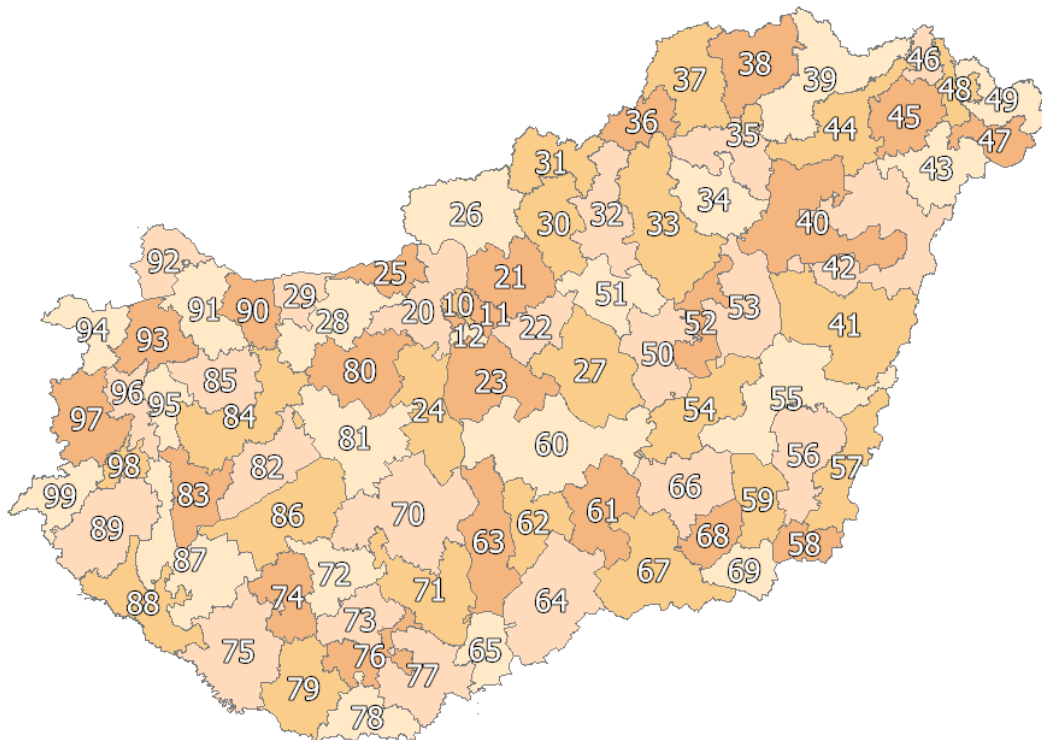
a katolikus egyház egyházmegyéi

- választási körzetek szerint (ország, megye, választókerület, település, szavazókörzet)



Magyarország választókerületei (2012-től)

- postai körzetek szerint (postai régió, két-szamos régió, budapesti kerületek, irányítószám-körzetek)



Magyarország 2-szamos postai irányítószám-körzetei

Mindegyik dimenzióban egymás alá sorolhatók be a használt fogalmak, de nem a generikus, hanem a partitív reláció mentén. Ebből következően a kiválasztott szempont szerint egybetartozó fogalmak között nem generikus, hanem partitív hierarchiákat lehet építeni.

Miért nem a generikus reláció köti össze az egy szempont szerint definiált fogalmakat? Nézzük meg ezt az egyházkormányzati fogalmakat példján keresztül! A generikus kapcsolat egyik fontos minősége az, hogy érvényes rá az IS-A reláció. Amikor azt mondjuk:

kutya IS-A állat

akkor minden tulajdonság, amit az állatra érvényesnek tartunk, érvényes lesz a kutyára is. Ezt fejezzük ki azzal, hogy a kutya fogalmát besoroljuk az állat fogalma alá. Az egyházkormányzati (és az összes többi, itt tárgyalt) fogalmak esetében azonban ez nincs így. Nem mondhatjuk ugyanis azt, hogy:

egyházközség IS-A egyházmegye

A közigazgatási fogalmak esetében sem igaz, hogy

település IS-A járás IS-A megye IS-A ország

Ezeket a fogalmakat nem a generikus reláció kapcsolja össze, hanem a partitív reláció¹⁸ (IS-PART-OF), ami egyfelől tranzitív láncba, másfelől – legtöbb esetben – monohierarchiába köti az azonos szempont szerint definiált fogalmakat.

egyházközség IS-PART-OF egyházmegye

¹⁸ Ezek a partitív tartalmazási viszonyok nem-atomos mereológianak felelnek meg.

település IS-PART-OF járás IS-PART-OF megye IS-PART-OF ország

Fontos azt is rögzítünk itt, hogy ezek a relációk univerzálék közti relációk, amelyeket figyelmesen el kell választanunk majd a partikulárek közt létező individuális partitív relációktól. Az az állítás, hogy

település IS-PART-OF¹ megye

két univerzálé közti kapcsolatot ír le, míg az a kijelentés, hogy

Dunaújváros IS-PART-OF² Fejér megye

két partikuláre kapcsolatáról szól. Ezt a megkülönböztetést a földrajzi névtérben érvényesíteni érdemes azzal, hogy kétféle partitív relációt engedünk meg.

Említést kell még tennünk azon ritka esetekről, amikor a partitív reláció mentén alárendelt földrajzi entitások nem monohierarchiába rendeződnek, hanem polihierarchiába. A társadalmilag konstruált (bona fide) földrajzi helyek esetében előfordulhat, hogy a geometriájuk (tehát a természeti tények) alapján több szülőjük is van. Ennek legnyilvánvalóbb példája az, amikor egy település nem egyetlen, összefüggő poligonból áll, hanem több részből. A település bona fide objektum (társadalmi konstrukció), míg a határát jelző geometria fiat objektum (természeti tény). Ha egy település területe diszjunkt poligonokból áll, akkor valamelyik része egy másik település poligonjába ágyazódik be, és fizikai-geometriai értelemben a beágyazott poligon a befogadó – másik – település része (ilyen magyarországi település Géderlak, Foktő, Fajsz).



Géderlak: több poligonból álló település

A partitív reláció pontosítását azonban nem csak az itt jelzett ontológiai különbségek miatt érdemes elvégezni, hanem azért is, mert ha a többféle fogalomképzési szempont miatt egy-egy földrajzi hely több „felettséssel” rendelkezhet, akkor valahogyan érdemes jelölni azt, hogy a szóban forgó konkrét földrajzi hely konkrét felettsége milyen szempont mentén van kapcsolatban vele. Fentebb bemutatuk, hogy egy települést sokféle szempont szerint része lehet egy másik földrajzi helynek (járásnak,

180

megyének, egyházmegyének, tájvédelmi körzetnek stb.). Ha ezeket a rész-egész viszonyokat egyetlen (ugyanazzal a) relációval reprezentáljuk, akkor az egyik adott szempont szerint felépíthető partitív hierarchiát (például a közigazgatási hierarchiát) csak úgy tudjuk összeállítani, hogy a felettesek (partitív fölrendelt kategóriák) minősítését is mindig figyelembe vesszük (lekérdezzük). Ez a művelet elvégezhető, csak sokkal több erőforrást (számítási kapacitást) köt le. Ezen a ponton tehát nincs szükség új ontológiai elköteleződésre, de a rendszer működtetésével kapcsolatos – hatékonysági, performancia – szempontok miatt elképzelhető új – egyébként redundáns – entitások (adatmezők) felvétele a rendszerbe.

2.29.1.4 Az írásműcímter struktúrája

A nemzeti névterek körébe olyan név- vagy címtereket is be lehet sorolni, amelyek adott típusú kulturális dokumentumok azonosítására alkalmasak a dokumentumok címei és további adattípusok alapján. A filmes világ számára ilyen névtérsgemmens lehet a **filmcímter**, a zene műveknek a **zeneműcímter**, a könyvtári világ számára pedig az írásos dokumentumok címtere, az **írásműcímter**. Utóbbira nehéz jó címet adni, és nehéz a határait is megvonni, vagyis azt a kérdést megválaszolni, hogy milyen írásos dokumentumok kezelését kell megoldani egy ilyen írásműcímter keretein belül és milyenként nem. Nem minden írásos dokumentum tartozik a könyvtári világ fennhatósága alá, hiszen a levéltárakban is alapvetően írásos dokumentumokat őriznek. Ha megtesszük azt a leegyszerűsítést, hogy a könyvtári kompetenciába – legalábbis első körben – a sokszorosított írásos dokumentumokat soroljuk be, akkor talán kicsivel könnyebb dolgunk lehet. Persze akkor olyan komoly filozófiai, modellezési problémákkal szembesülünk, amelyekkel nem találkozánk, ha a másolás jelenségétől eltekintenénk. Az írásbeliség keretein belül létrehozunk **írásos dokumentumokat**. Az írásos dokumentum adott gondolati **tartalom** rögzítése valamilyen írástechnikával valamilyen **hordozóra**.

Hogy milyen és mennyi tartalmat szükséges rögzíteni ahhoz, hogy dokumentumról beszéljünk, illetve milyen hordozó esetén beszélhetünk egyáltalán írástechnikáról, nehéz kérdés, nincs egyértelmű válasz rá. Még inkább bizonytalanok lehetünk annak a kérdésnek a megválaszolásakor, hogy a dokumentumok közül miket tarthatunk (egyéni jellegű, valamilyen kreativitást megjelenítő) műveknek és miket nem. A dokumentumok közül valamilyen társadalmi/értékelési mechanizmus révén kiválogatódnak azok a **művek**, amelyekhez szerzőket, kreatív alkotókat rendelhetünk, és szerzői jogokat tulajdonítunk nekik, illetve a művek azonosítása végett címetek adunk nekik. Egy következő lépésben pedig, annak érdekében, hogy ezek a művek minél több befogadóhoz eljuthassanak, sokszorosítjuk őket. Mivel ezek a sokszorosított írásművek bekerülnek abba a kulturális térbe, amelyekben a kulturális dokumentumok megőrzését, fenntartását közgyűjtemények (is) végzik, ezért ezen a ponton megjelenik az az igény, hogy az írásműveket is azonosítani tudjuk valamilyen írásműcímter segítségével. A kérdés már csak annyi, hogy pontosan mi is az, amit itt azonosítanunk kell.

Ha csak egyedi művek lennének a könyvtárakban (ahogy – kis túlzással – mondhatjuk, hogy a levéltárakban ez a helyzet), akkor könnyebb lenne megválaszolni azt a kérdést, miket és hogyan kell azonosítanunk az írásos dokumentumok világában. Minden egyes dokumentumot külön-külön kellene leírni és egyedi azonosítóval ellátni. A könyvtárak azonban sokszorosított műveket kell, hogy feldolgozzanak, leírjanak, amikor mondhatjuk, hogy a – gépi úton – sokszorosított írásművek (könyvek) nem különböztethetők meg egymástól, ezért ezeket egy csomagban kellene leírni valahogyan, nem pedig külön-külön. Természetesen a könyvtári világ már rég kitalálta a megoldást erre a problémára, hiszen a könyvtárak mindig is elkülönítették a könyvek kiadását, illetve az egyes kiadások konkrét, individuális példányait egymástól. Ilyen módon a könyvtárban mindig is elég volt egyszer felvenni a könyv egy adott kiadásának adatait: a könyv címét, szerzőjét, kiadóját, kiadási idejét stb. Ha egy művet több kiadásban is kiadtak, akkor az nem jelentett különösebb pluszmunkát, ha az új kiadás esetében újra fel kellett venni az alapadatokat, hiszen ezekből nem volt olyan sok. Az új kiadások esetében voltak adatok, amelyek nem változtak az új kiadással (például a mű szerzője vagy címe), de a könyvtári világ számára vállalható "fölsleges teher" volt a már ismert adatok újbóli rögzítése. Azt a veszteséget is vállalhatónak tartotta a könyvtári világ hosszú időn keresztül, hogy a kiadvány-példány alapú modellben nem tudták igazán jól kifejezni a különböző kiadványok közti kapcsolatot.

Már az 1990-es évek végén megszületett az az új elgondolás, az FRBR-modell, amely a dokumentumok többszintű leírását javasolta (FRBR 2008). Az FRBR négy szint elkülönítését javasolja, hogy egy egységes modellben lehessen kezelni a **példányok (instance)** és kiadványok vagy **megjelenési formák**

(**manifestation**) mellett a **műveket (work)** és a művek különböző **kifejeződési formáit (expression)**. Utóbbi szint behozatalával lehet remélni azt, hogy össze tudjuk kapcsolni egy írásos mű (mondjuk: egy regény) hagyományos könyvkiadványait a mű más formátumú, más médiumokon keresztüli (például hangoskönyv vagy képregény alakban való) megjelenéseitől.

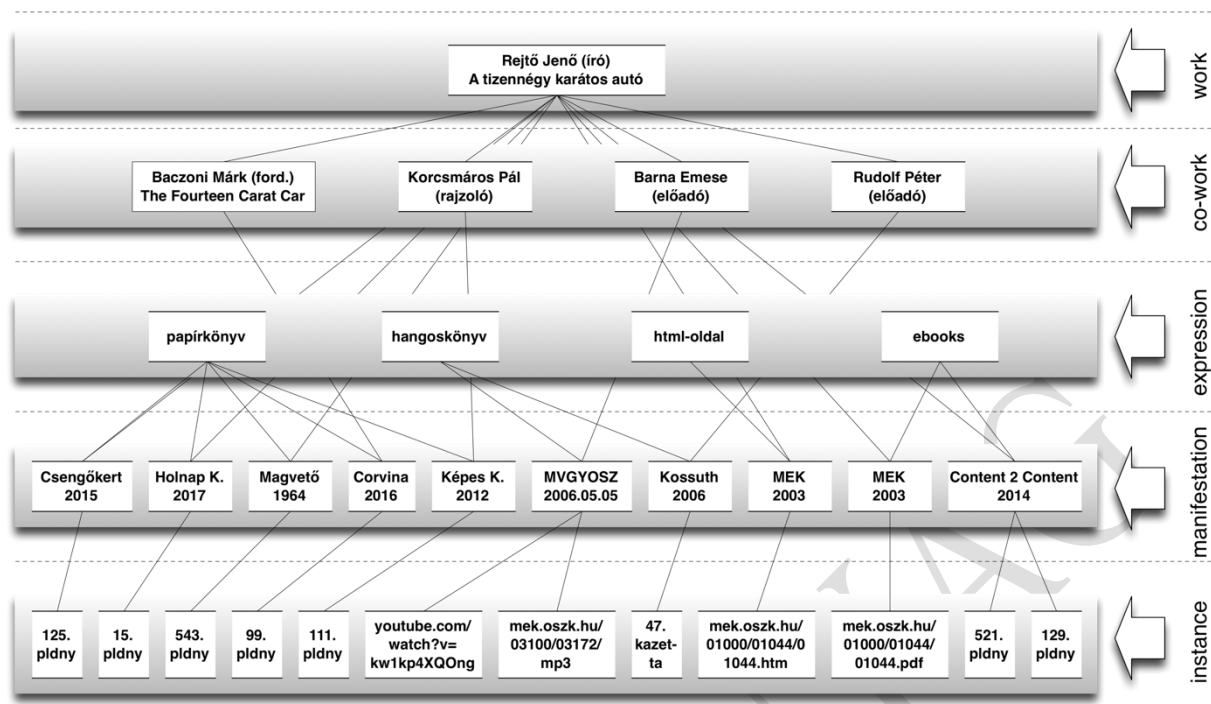
A négy szintű FRBR-modell sokkal jobban alkalmas arra, hogy megragadjuk vele a művek (tartalmak) többféle médiumon keresztüli nyilvánosságához közvetítéseket, illetve a különböző sokszorosítási eseményeken keresztüli megjelentetéseket, de azzal, hogy a modell elméletileg képes lekezelni a világ bonyolultságát, az adatrögzítés gyakorlati feladatát is bonyolulttá teszi. Sok más tényező mellett ez a nehézség is hozzájárulhatott ahhoz, hogy a könyvtári világ sehol sem fogadta be és nem implementálta azonnal a modellt a hétköznapi munkájába. Az új, feltörekvő szabvány, a Bibframe ugyan elfogadta az FRBR-t a dokumentumok modelljeként, de csak három szintet (a work, a manifestation és az instance fogalmait) alkalmazza.

Pedig az igazán pontos modell felállításához nemhogy elvenni kéne belőle, hanem inkább még egy szintet bele kéne tenni a modellbe. Az FRBR ugyanis nem tudja elég finoman kezelni azt, hogy az eredeti műveket olykor – ugyancsak kreatív alkotói tevékenység révén – átalakítják, valamilyen értelemben más formátumra hozzák. A legnyilvánvalóbb példa erre az írásművek más nyelvre történő lefordítása és kiadása. Ilyen esetben azt sem mondhatjuk igazán, hogy az eredeti mű jelent volna meg egy új kiadványban, hiszen könnyen lehet, hogy a mű szerzője nem is tudná elolvasni, nem értené az idegen nyelvre lefordított művét, de nyilván azt sem mondhatjuk, hogy a fordítással egy új mű keletkezett volna. Ebbe a problémahalmazba tartozik az a jelenség is, amikor egy művet több fordító is lefordít egy másik nyelvre, és ilyenkor kellene tudni jelezni a kiadványok esetében, hogy melyik fordító fordításáról van szó. Ez a megkülönböztetési igény világosan jelzi a fenti problémát. Nyilván ugyanarról a műről van szó, hiszen mindkét fordítás/kiadvány esetében az eredeti mű megegyezik, de a két kiadvány mégis kétféle fordításban (valamilyen értelemben eltérő tartalommal, akár eltérő címeikkel) jelenteti meg az eredeti művet, kétféle szerzői jogi igény keletkezik a kétféle fordításért. Másik példa lehet a hangoskönyvek esete, amikor a szöveget különböző előadók olvassák fel, és ezzel másféle hozzáadott érték keletkezik. Az így kiadott hangoskönyvek nyilván eltérnek egymástól (bizonyos értelemben tartalmilag is), hiszen más színész (felolvasó) hangján szólnak meg. Nem elég tehát azt mondani, hogy ezeket hangoskönyvként elválasztjuk a papírkönyvektől (az expression szinten), és megengedjük, hogy a hangoskönyveknek legyenek egymást követő kiadásai, mert így nem tudjuk elválasztani egymástól a felolvasás révén keletkezett kreatív hozzájárulásokat, az eltérő színészi munka révén keletkező, eltérő hozzáadott értékeket.

A megoldás az lehet, hogy a mű (work) szintjéhez kapcsolódóan bevezetünk még egy második, **társ-mű (co-work)** szintet, ahol már a származtatott, kiegészítő jellegű alkotói tevékenység eredményét is meg tudjuk ragadni (az eredeti műhöz hozzáragasztva, de egyben el is választva tőle). Javaslatunk szerint az FRBR eredeti négy szintje helyett helyesebb lenne öt szintet elkülöníteni egymástól (ez a megoldás pontosabb modellt eredményezne).

- work
- co-work
- expression
- manifestation
- instance

Az öt szintű modell értelmét (értelmességét) egy olyan ábrával szemléltetjük, ahol a megfelelő szinteken tüntetjük fel egy konkrét mű (Rejtő Jenő 'A tizennégy karátos autó' című regényének) legfontosabb adatait.



az ötszintű dokumentummodell szerkezete egy Rejtő-mű példáján keresztül

Az FRBR modell ahhoz nyújt segítséget, hogy milyen szinten kell megragadni a dokumentum fogalmát. Az írásos dokumentumokkal kapcsolatos másik fontos kérdés annak tisztázása, hogy milyen **szerkezetű dokumentumról** van szó. Szerkezetét tekintve a dokumentum lehet **összetett**, illetve **egyszerű** (vagy gyűjteményes), ami annyit jelent, hogy egy dokumentum (egy könyv, egy periodika adott száma) vagy tartalmaz több dokumentumot vagy nem, tehát vagy többedmagában vagy egymagában áll. Előbbi esetben összetett vagy gyűjteményes dokumentumról beszélhetünk, utóbbi esetben pedig egyszerű dokumentumról. Ha ezt a megkülönböztetést megtesszük, akkor összetett dokumentumként értelmezhetjük egy szerző tanulmánygyűjteményét, egy költő verskötetét és egyszerű dokumentumként az egyedi tanulmányokat vagy verseket, és természetesen meg kell engednünk (és a modellnek kezelnie kell tudni) az olyan eseteket, amikor egy tanulmány vagy egy vers több gyűjteményes dokumentumban is megjelenik. Ebből a szempontból tekintve a periodikumok mind gyűjteményes dokumentumoknak tekinthetők.

A harmadik dokumentumtipizálási szempont szerint arra is érdemes figyelni, hogy bizonyos dokumentumok úgy is összekapcsolhatók egymással, hogy nem egy kiadványban válnak egy összetett dokumentum részeivé, hanem időben, **sorozatként** kapcsolódnak össze egymással. A könyvsorozatokba foglalt egyedi kiadványok ilyenek, de a periodikumoknak is fontos tulajdonsága ez a sorozati jelleg. Fontos megjegyezni, hogy a sorozatiság is egyfajta gyűjteményességet jelent, csak itt az idő tengelyen kapcsolunk össze dokumentumokat egymással, míg az összetett dokumentumokban egy időben, egyszerre megjelent műveket fogunk össze. Mondhatjuk, hogy a gyűjteményesség kétféle értelme abban a különbségben ragadható meg, hogy a sorozatiság elve mentén az idő eltérő pontjain megjelent dokumentumokat kapcsolunk össze, míg az összetett dokumentumok esetében az idő ugyanazon pontján kapcsolunk össze különböző dokumentumokat. A világot így értelmezve a periodikumok egyszerre összetett és sorozati dokumentumok, a gyűjteményes kötetet összetett dokumentumok, és a könyvsorozatok tipikusan sorozati dokumentumok.

Az írásművek névterében maga az írásmű mint dokumentum áll a középpontban. Az eddigi terminológiánkat használva az írásmű felel meg a névhordozó fogalmának. Az FRBR-modell alapján a dokumentum fogalmát több ontológiai szintre „szétterítve” tudjuk csak jól megragadni. Ezen a ponton elég sok új entitást kell a modellünkbe beemelni. Ezzel szemben az **írásműcím** a nyelvi konstrukciók világába tartozik, amelyek kezelése végett nincs szükségünk további ontológiai elköteleződésekre, hiszen a nyelvi konstrukciókat minden névtér esetében egységesen kezeljük. Ügyelni kell viszont arra, hogy a modell biztosítsa azt, hogy a címeket mindig a megfelelő dokumentumszinthez rendeljük hozzá.

Az írásművek metaadatainak leírásához szükség van arra, hogy rögzítsük az írásművek és a létrehozásokban, terjesztésükben, megőrzésükben stb. szerepet vállaló ágensek (személyek és testületek), valamint a kulturális termelés és fogyasztás helyszínei közti kapcsolatokat, illetve a tartalom hordozójára vonatkozó adatokat. Az írásműcímtéren belül azonban ezekre a metaadatokra (pontosabban a metaadatok egy kisebb körére) csak azért van szükség, hogy segítségükkel egyértelműen azonosítani lehessen az írásműveket. Ha van két 'Alkonyat' című könyv, akkor abban dönteni, hogy ugyanarról a műről van-e szó vagy sem, nagy segítséget jelenthet, ha tudjuk, ki a szerzője, kiadója, mikor jelent meg. Az írásműcímtéren belül definiálni lehet azt a minimálisan szükséges (és elégséges) adatelemkészletet, amely az egyértelmű azonosítást segíti, és a többi metaadatot nem szükséges „beengedni” ebbe a névtérbe.

Mivel az írásművek azonosításához szükségesnek látszanak a dokumentum-ágens viszonyt leíró metaadatok, ezeket be kell emelni a névtérbe, ami a személynévtér és a testületi névtér felé mutató kapcsolatok kezelését igényli. Ennek a kapcsolatnak az inverze viszont lehetővé teszi azt, hogy mind a személynévtérbe, mind a testületi névtérbe elhelyezhessük a személyes, illetve intézményi bibliográfiákat.

2.29.1.5 Az audiovizuálismű-címterek struktúrája

Az audiovizuális művek esetében nagyon hasonlóak, még ha nem is ugyanazok a modellezési kérdések, a meghozandó döntések, a szükséges ontológiai elköteleződések, mint amelyeket az írásműcímtérrel kapcsolatban bemutatunk. Mivel mind a tartalom, mind a hordozó minősége más az írásos és audiovizuális művek esetében, ennek nyilvánvalóan vannak következményei a modell egészét tekintve, de nagyon fontos felismerni, hogy az FRBR-modellt alkalmazhatjuk az audiovizuális dokumentumok esetében is. Itt is ugyanúgy hasznos az ötödik szint (co-work) felvétele az FRBR-modellbe (a zeneművek előadói esetében kiváltképp fontos lehet ez a kiegészítés). Vannak sorozatok ebben a világban is (gondoljunk csak a rádió- és tévésorozatokra), és előfordulhat a gyűjteményes jellegű egymás mellé rendelés is, bár kisebb mértékben, mint az írásos művek esetében. A művek azonosításához itt is szükség van a fontosabb dokumentum-ágens kapcsolatok ismeretére, és itt is értékes „melléktermék” lehet a személyes filmográfiák, diszkográfiák létrejötte a személynévtérben, testületi névtérben.

Mivel a kulturális érdeklődések igen nagy hányada az audiovizuális művekre irányul, ezért a **zeneműcímtér** és a **filmcímtér** létezése valós közönségigényt elégíthetne ki. Ezeket az íráscímtér alapján igen nagy részben fel lehetne építeni, és csak kisebb részben lenne arra szükség, hogy az audiovizuális dokumentumok specialitásainak kezelésére további ontológiai elköteleződések tegyünk.

2.29.2A köznévterek struktúrája

Arra a kérdésre, hogy milyen struktúrája van a köznévtérnek, melyik tudásszervezési rendszer alá sorolható be, nincs egyértelmű válasz. Sokféle köznévtér hozhatunk létre, és hogy melyet, az jelentős mértékben függhet attól, hogy milyen célokra akarjuk használni. A tulajdonnévtér célja a névhordozók és nevek egyértelmű azonosítása, és ezáltal annak a lehetőségnek a megteremtése, hogy a nemzeti névtér azonosítóit átvevő kulturális adatszolgáltatások megnyílnak egymás és a közönség előtt abban az értelemben, hogy a közös ID-kel azonosított személyek, testületek, földrajzi helyek, művek révén sok irányban, sok szálon keresztül bejárhatókká, átjárhatókká válnak ezek az adatrendszerek. Elméletileg a nemzeti köznévtér is betöltheti ezt a funkciót, amennyiben a kulturális adatszolgáltatók ugyanazt a köznévtérstruktúrát veszik igénybe az általuk kezelt tartalmak formai és tartalmi jellemzésére. Erre azonban kevésbé számíthatunk, mert a különböző archívumok dokumentumai eltérő adatelemkészlettel írhatók le pontosan, ami megmutatkozik a nemzetközi szabványok különbségeiben is. A tartalmi leírás területén nem kell eltérésekre számítanunk az egyes szakmák, intézmények, archívumok, műtípusok között, hiszen tartalom rövid jellemzésében nincsenek a dokumentumok típusától függő szempontok, ennek ellenére ebben a tevékenységi körben sem alakultak ki közös archívumi gyakorlatok.

A köznévtér modelljét úgy kell megtervezni, és a köznévtér működését úgy kell felépíteni, hogy mindenféle köznévtérrel kezelni, építeni, reprezentálni lehessen benne. Ez nem tűnik igazán nagy kihívásnak, hiszen ha belegondolunk, hogy milyen tudásszervezési rendszereket képzelhetünk el

egyáltalán, ezek a tudásszervezési rendszerek milyen struktúrával rendelkeznek, és mindezek fölül milyen általánosított modellt és struktúrát lehet megtervezni és elhelyezni, akkor elfogadhatjuk azt az álláspontot, miszerint egy általánosított, tetszőleges relációt kezelni képes köznévtér megtervezve, felépítve és működtetve minden köznévtérrel kapcsolatos igényt ki lehet elégíteni.

Az ETO egy taxonómia, a Köztaurusz egy tezaurusz, a személynévtér, a testületi névtér, a földrajzi névtér, az íráscímtér mint tulajdonnévtér egyben egy terminuslista. Korábban bemutattuk, hogy ezek a tudásszervezési rendszerek egységes modellben kezelhetők, és „csak” a rájuk jellemző relációk típusában és számosságában különböznek egymástól. Ha a köznévtérrel úgy építjük fel, hogy az általános tudásszervezési rendszer modelljében felsorolt összes relációt felvesszük a modellbe és implementáljuk az adatrendszerbe, akkor az összes tudásszervezési rendszer igényét képesek lehetünk kiszolgálni. Praktikusan, a fejlesztés első időszakára nézve mondhatjuk, hogy a köznévtérnek valahol a tezauruszok és a formális ontológiák között van a helye. Kezelnie kell minden tezauruszrelációt, sőt, fel kell tudni venni benne bármilyen új relációt, ebben az értelemben el kell indulni teljesen formális és általános ontológia felé, de arra még nincs szükség, hogy a formális ontológiák következtető képességét is elvárjuk a rendszertől.

A tulajdonnévtérben szükség lehet arra, hogy a köznévtérből listákat, olykor hierarchikus listákat vegyenek igénybe akkor, amikor a névhordozókat, a neveket, a névhordozók, nevek közti kapcsolatokat kell tipizálni, de ezek a köznévtérstruktúrák sosem lesznek bonyolultak.

A dokumentumleíró rendszerek más adatelemeket, más relációkat is megkövetelhetnek, ám ezeknek az igényeknek már nem kell megfelelnünk a névtér tervezésekor, fejlesztésekor.

2.30A névtérdefiníció fejlesztésre gyakorolt hatásai

A névtéren belül megkülönböztetjük a névhordozókat (akiket vagy amiket megnevezünk, akikre vagy amikre a nevekkel referálunk) és a neveket (amelyekkel a referálást végrehajtjuk). E kettősségből fakadóan a névtérben nem csak a névhordozókat, de a neveket is – egyedileg – azonosítani kell.

Névtérről névtérre változhat az, hogy névhordozót vagy nevet azonosít-e a névtér. A KSH Helységnevtára nevet azonosít (külön azonosítóval szerepel benne 'Dunaújváros', 'Sztálinváros' és 'Dunapentele', holott ezek ugyanazon település különböző nevei). A FÖMI Földrajzi Névtére viszont névhordozót azonosít, és nem azonosítja külön-külön a neveket.

A tulajdonnévtéren belül az eseményeknek kulcsszerepe van. Egyrészt azért, mert a tulajdonnévtérben kezelt névhordozók mindig individuális entitások (partikulárok), amelyeket – elvileg – mindig el kell tudnunk helyezni a tér és idő valamely pontján. Ebben segítenek az események: a névhordozók azonosításához, változásainak nyomon követéséhez mindig eseményeket kell felvennünk az adatbázisban. Az események tulajdonnévtéren belüli központi jelentőségének másik oka az, hogy itt nem neveket, hanem megnevezéseket, megnevezési eseményeket kell kezelnünk. A nevek mint tértől és időtől független entitások (univerzálék) sosem lehetnek alkalmasak önmagukban az egyedi azonosításra. Az individuális névhordozókhöz mint a típusuk instanciáihoz csak úgy tudunk névpéldányokat rendelni, ha a megnevezési aktuson (gyakorlaton), azaz egy eseményen keresztül hozzákötjük a referáló nevet az időhöz (és természetesen a jelölt entitáshoz).

A névtér modelljében minden időfüggő adatot „eseményesíteni” kell, vagyis nem valamely entitás attribútumaként kell felvenni, hanem „önálló” eseményként, és az eseményt kell összekapcsolni az adott entitással (névhordozóval, névvel stb).

A megnevezések kontextusfüggősége, megnevezőiközösség-függése miatt a nevekhez kapcsolt tipológiában valahogyan jelölni kell, hogy milyen megnevezési kontextusa, közösségi státusa van a névnek.

Az események központi kezelése lehetővé teszi azt is, hogy egyfelől egy helyen, egységesen oldjunk meg minden időkezeléssel kapcsolatos problémát (az időpontosság, az összetett események vagy a balról vagy jobbról nyitott végű események kérdését), másfelől egységes módon, egy helyről építhessük fel az idővonal-szolgáltatásokat, amelyek az egyes névtérek részeként is, de önállóan is működhetnek.

Minden nyelvi adat az elkülönült nyelvi modulban (és csak ott) van. A köznévtérek, a szűkebb értelemben vett tulajdonnévtérek minden adata a nyelvi sémán keresztül érhető el.

Szükségesnek látszik, hogy a partitív relációkat, vagy általában a hierarchikus relációkat kifinomultabban kezeljük. Egyfelől el kell különíteni az univerzálék (köznevek) és partikulárék (tulajdonnevek) közti partitív relációkat egyástól, másfelől külön kell kezelni az atomos és nem-atomos partitív relációkat. E két szempont figyelembe vétele már azt követeli meg, hogy legalább négyféle partitív relációt tudjunk kezelni a rendszeren belül. Arra is fel kell készülni, hogy – szükség esetén – különbséget tudjunk tenni a természeti, illetve társadalmi tényeken alapuló (fiat, illetve bona fide) partitív relációk között is.

A földrajzi helyek világában egy-egy entitást több entitással is összekapcsolhatunk a partitív reláció segítségével (egy hely többféle „körzetnek” lehet a része). Ilyen esetekben a partitív reláció ontológiai minősége lehet ugyanaz, és csak a partitív fölérendelt entitás eltérő minősége mentén tudjuk elkülöníteni a kapcsolatokat egymástól: ha egy település része egy megyének is, meg egy egyházmegyének is, akkor ugyanarról a partitív relációról beszélünk, és a két kapcsolat különbsége a megye és az egyházmegye eltérő minőségéből vezethető le. Megfontolandó, hogy az ilyen különbségek jelzésére is önálló relációt vezessünk be (különben olykor nagy és talán fölösleges számításokra lehet szükség az adatok megjelenítésekor).

A geoentitásokat – elvileg – egyértelműen azonosítani lehet a földrajzi térben való pontos elhelyezkedésükkel. A gyakorlatban ezt a lehetőséget akkor tudjuk kihasználni, ha rendelkezésre állnak a megfelelő pontosságú geometriaitartalmazási adatok. Amíg nem ez a helyzet, addig szükségünk van másfajta azonosítási megoldásokra is. Már csak azért is, mert az ember számára nem ez a legbefogadhatóbb azonosítási megoldás. A geoentitások „földrajzi térben való elhelyezésében” kiemelt szerepet játszik az az útvonal (path), amely az „origo”-tól, a Földtől kiindulva megmutatja a kiválasztott geoentitás helyét egy partitív tranzitív lánc segítségével. Ennek során meg kell oldani azt, hogy vannak több felettel rendelkező geoentitások (a folyók szolgáltathatnak példát erre, hiszen azok több településen, akár több megyén, több országon keresztül folyhatnak, és így több geoentitásnak is részét képezik).

A partitív relációval mindig a közvetlen fölérendeltje/alárendeltje viszonyt rögzítjük, de a tranzitivitás kihasználásával természetesen mindig meg tudjuk mondani, hogy egy adott geoentitás milyen más entitást tartalmaz a többszörös partitív kapcsolódások mentén. A fejlesztés során meg kell oldani, hogy valamely pont összes alatta levő egyedét egyetlen függvénnyel ki lehessen listázni (klánlistázás).

A tágabb értelemben vett nemzeti névterek, a névtárak tartalmaznak olyan adatszégmenseket, amelyek az egyes névterek entitásai közti kapcsolatokat fejezik ki (archontológiák, bibliográfiák stb.). Ezek egy része olyan egyedi kapcsolat, amelyet egyedileg lehet csak felvenni, ezért az ilyen adatokat csak a nemzeti névtér szerkesztőségi rendszerén belül menedzselni. Vannak olyan adatkapcsolattípusok is, amelyek kezeléséhez nincs szükség szerkesztőségi rendszerre, anélkül is nagy tömegben menedzselhetők (választási, térstatisztikai adatok stb.). Az egyedi kapcsolatok szerkesztésére az adminrendszeren belül kell megfelelő felületeket biztosítani, a nagytömegű adatkapcsolatok menedzselését SQL-konzolon keresztül is biztosítani lehet. A fejlesztés során viszont az adminrendszert fel kell készíteni arra, hogy lehetővé tegye azt, hogy a nagytömegű adatokat megjelenítő oldalakat be lehessen fűzni a publikus szolgáltatás menüjébe.

A tulajdonnevek egyszerű terminuslisták. A szűkebb értelemben vett tulajdonnévterek a lehetséges tulajdonnevek listáját tartalmazzák. Ezek a listák a rendszeren belül elkülönített nyelvi modulban vannak. A tágabb értelemben vett tulajdonnévterek, azaz a megnevezések az individuális névhordozókat kapcsolják össze a tulajdonnevekkel, és a nevek „példányosítását” a megnevezési események hozzárendelésével végezhetjük el.

A névtér általános modelljébe a tudásszervezési rendszer (KOS) általános modelljét kell beemlíteni, és ezt kell implementálni a fejlesztés során. Ha a névtér-projekt bármely más, későbbi fázisában új igény jelenik meg egy speciális tudásszervezési rendszer alkalmazására, akkor a köznévtéren belül ezt mindig meg kell tudni oldani a modell, az adminrendszer bármilyen változtatása nélkül.

Az adminrendszer használata szempontjából nagyon fontos biztosítani, hogy a rendszer bármely pontján a köznévtér elemeire támaszkodva új, szakmai kontextusokat biztosító nézeteket lehessen definiálni és a rendszerbe felvenni. Ennek példái lehetnek az olyan nézetek, amelyeket annak érdekében kell definiálni, hogy általuk biztosítani tudjuk az együttműködő intézmények szakmai elfoglaltságait (kinék mi számít kiemelt névalaknak, névváltozatnak stb.) vagy, hogy csak az általuk használni kívánt adatszégmenseket láthassák, ne kelljen „kerülni” az összes adatot.

3 KOS-alapú keresés

3.1 Bevezetés

A **nemzeti névterek** alapvető funkciója az, hogy egyfelől a különböző – kulturális és más tartalmú – adatbázisok közös részét alkotó információk elemeket (személyeket, testületeket, földrajzi helyeket, műcímeket stb.) **egyértelműen és közösen azonosítsák**, másfelől a közös azonosítók segítségével **átjárhatóvá** tegyék a különböző adatrendszereket. Ez utóbbi cél megvalósítása képes kiszolgálni azt az igényt, amely a felhasználók oldalán jelentkezik, amikor valamilyen jelenségről mindent egyszerre és könnyen szeretnének megszerezni akkor is, ha azok szét vannak szórva a különböző adatbázisok között. A kulturális és más jellegű információk igényeket ugyanis ritkán lehet kielégíteni az intézményi szegmentáció logikája által meghatározott szegmentált szolgáltatásokon keresztül.

Az adatrendszerekben tárolt információkhoz való **hozzáférésnek** (vagy másként: **navigációnak**) többféle módja lehetséges. Ezeket áttekintjük a későbbiekben. Kiemelkednek ezek közül a **keresőszolgáltatások**, amelyek a felhasználók által megfogalmazott keresőfeltételekre adnak válaszokat valamilyen keresési algoritmusok segítségével. Ezért nagy szükség van a keresés fogalmának, a keresési lehetőségek tipizálásának minél pontosabb és minél kimerítőbb tárgyalására.

Az – információhoz való – hozzáférés, illetve az – információban való – keresés legáltalánosabban azt jelenti, hogy valamilyen információk tartományban „elhelyezett” „specifikus” információt szeretnénk megtalálni. Ennek tisztázáshoz legelőször arra a kérdésre kell választ adnunk, hogy milyen információról, illetve milyen információk tartományról beszélhetünk egyáltalán. Ezután kell tisztázni azt, hogy az információ átadásának milyen technikái lehet, amelyek – a projekt szempontjából vett – legfontosabb kérdése az, hogy a teret vagy időt áthidaló eszközöket alkalmazunk-e a kommunikáció során. Az időt áthidaló kommunikációt az teszi lehetővé, hogy az átadni kívánt információt rögzítjük, ami azt is jelenti egyben, hogy dokumentumokat hozunk létre. Az információ tipizálása mellett ezért arra is szükségünk lehet, hogy legyen egy dokumentumtipológiánk is. Az **információ- és dokumentumtipológia** felállítása után fel lehet vázolni azt, hogy milyen hozzáférési és keresési lehetőségek vannak.

A keresőtechnológiák tárgyalását érdemes a **nyelvi**, szövegalapú **információkra** szűkíteni, ezt a szűkítést viszont érdemes azzal árnyalni, hogy a nyelvi, szöveges információk kezelésnek milyen módjai lehetségesek. Ennek során tisztázni kell azt, hogy mit jelent a **szöveg** és az **adatbázis** fogalma, mi a különbség és mi az azonosság köztük. Az elméleti tisztázó munkához használni lehet Ferdinand de Saussure nyelvelméletét, amely felkínálja azokat az alapfogalmakat, amelyek segítségével nemcsak az adatbázis fogalmát lehet elkülöníteni a szövegtől, de rá lehet mutatni arra, is hogy mi az értelmük és mi a hasznuk a **tudásszervezési rendszerek** felépítésének és alkalmazásának.

Az információhoz való hozzáférés, a navigáció mindig valamilyen információk tartományban valósul meg, vagy másként: mindig valamilyen információk térben keresünk valamilyen információt. A hozzáférési, navigációs technológiák mindig valamilyen **szűrést** hajtanak végre az adott információk térben, eredményül mindig a teljes tartomány egy részét adják vissza (ami lehet egy szó is vagy egy mondat). A hozzáférés, navigáció és a szűrés így ugyanannak a jelenségnek két oldalát, két nézetét jelenti.

A névterekben közneveket, illetve tulajdonneveket kezelünk abból a célból, hogy azokat egyértelműen azonosíthatóvá tegyük úgy, hogy egyedi azonosítókat rendelünk hozzájuk. Ezzel egyben lehetővé tesszük azt, hogy azon intézmények adatbázisai közt sokszálú kapcsolatrendszert építhessünk fel.

Ahhoz, hogy a névterekben való keresés lehetőségeit minél jobban kihasználhassuk, szükség van arra, hogy tisztázzuk, milyen viszonyt tetelezhetünk a névtereknek és a tudásszervezési rendszerek között.

3.2 Alapfogalmak

Az információtipológia bemutatásával együtt röviden áttekintjük azt a gondolatmenetet, amelyen végig menve – második menetben – szabatosan definiálni tudjuk az adatbázis, a keresetőség, a tudásszervezési rendszer fogalmát. A fogalmi építkezés menetének felvázolása után példákon keresztül, szisztematikusan kibontjuk azt az elméletet, amelyre támaszkodva a nemzeti névtér építéséhez szükséges legfontosabb fogalmak jelentését tisztázni lehet.

Az információ többféle típusáról beszélhetünk aszerint, hogy milyen érzékszervünkön keresztül fogadjuk be, illetve nyelvi vagy nem-nyelvi kommunikációról van-e szó. Ha csak a vizuális és auditív csatornákon keresztül érkező információkra szűkítjük a vizsgálódás fókuszát,¹⁹ akkor az alábbi típusokat kapjuk (Syi 2007, 2017a).

| | auditív jel | vizuális jel |
|----------------|-------------|-----------------------|
| nem-nyelvi jel | zene, zörej | mozgóképek, állóképek |
| nyelvi jel | beszéd | írás, jelnyelvelés |

További tipizálási lehetőséget jelent, hogy az információ rögzített-e, vagy sem. Egy nyelvi üzenetet kifejezhetünk jelnyelvi mutogatással és elmondhatjuk szóban is. Ekkor az üzenet nem marad fenn az időben, azt csak a kibocsátásával egy időben fogadhatják be mások. Ha viszont rögzítjük a tartalmat (leírjuk papírra, felvesszük diktafonnal, kamerával), akkor nem vagyunk a kibocsátás pillanatához kötve, időben később is hozzá lehet férni a tartalomhoz. Mind a négy fenti helyzetben (a zene, a kép, a beszéd vagy az írás esetében) beszélhetünk rögzített és nem rögzített kommunikációs formáról.

| | auditív | | vizuális | |
|---------------|----------------|------------------|-------------------|--------------|
| | nem-nyelvi | nyelvi | nem-nyelvi | nyelvi |
| nem-rögzített | élő zene | élő beszéd | táncszínház | jelnyelvelés |
| rögzített | rögzített zene | rögzített beszéd | fénykép, mozifilm | írás |

Végül a nyelvi információ fogalmát érdemes még egy szempont szerint alábontani, hogy megragadható legyen a digitális korszak talán legfontosabb vonása. Ferdinand de Saussure nyomán a nyelvi megnyilatkozásainkat két dimenzió mentén értelmezhetjük (Saussure 1998). A szintagmatikus tengely mentén az akusztikus és/vagy vizuális térben jelenlevő, látható, hallható, szintaktikailag jól-formált mondatok interpretációját végezhetjük el, míg az asszociatív/paradigmatikus dimenzióban a kommunikatív térben nem jelenlevő (az akusztikus térben nem hallható, a vizuális térben nem látható), de a fejünkben meglévő, szemantikailag rendezett (vagy rendezhető) tudást mozgósíthatjuk. Másként fogalmazva: a világról szóló tudásunkat nyelvi mintázatok, sémák alapján általánosíthatjuk, amelynek kifejezésére nem alkalmas a narratív logika, a mondatok lineáris, szekvenciális sorba rendezése – amely a nyelv szintagmatikus dimenzióját jellemzi. Ehhez egy sajátos grafonyelvi technikára van szükség (Goody 1998), amely csak írásban lehetséges, amelynek ősképe a táblázat, a modern formája pedig az adatbázis. Ez a technika, formai értelemben, biztosítja a mondatainkba foglalt tudás két- vagy többdimenziós elrendezését azáltal, hogy egyszerre teszi lehetővé a nyelvi üzenetek sorok és oszlopok szerinti olvasását, tartalmi szempontból pedig lehetőséget ad a világról való egyedi tudásunk

¹⁹ A jelen kutatás számára irreleváns a szaglási (olfaktorikus), az ízlelési (gusztatorikus), a tapintási (taktilis), valamint a testérzékelési érzékszervi csatornákon keresztül befogadható információ.

általánosítására az osztályzási műveleteken keresztül (Syi2017b). Az adatbázis is rögzített, nyelvi információ, de strukturált szöveg, ami annyit jelent, hogy egyfelől a mondatok komponensei – osztályozási szempontok szerint – szegmentálva, tipizálva vannak, másfelől minden mondat új sorban (de legalábbis a többi mondatnál egyértelműen elválasztva) van rögzítve. A beszéd – ebben az értelemben – nem strukturált, és az írás is csak akkor és úgy válik azzá, ha rendezett formába alakítjuk át a szöveget. Ehhez a szöveg önmagában nem elegendő, további információra van szükség, amely a szöveg belső tagolását, elrendezését adja meg. Ezt nevezhetjük sémának, ami nélkül a táblázat, adatbázis értelmezhetetlen lenne. Azt már említettem, hogy a szemantikailag jól-strukturált mondatok egyik előnye az, hogy megadják az általánosítás – és azon keresztül az új, érvényes tudás megszerzésének – lehetőségét. A strukturálva osztályozás másik előnye az, hogy az ilyen mondatok sematikuságukból fakadó „egyszerűségük” révén könnyen kezelhetők, ami – idővel – megteremti annak lehetőségét, hogy szenzorokkal gyűjtsük, gépekkel növeljük a világról szóló, sémákba írható egyedi állítások számosságát, ezáltal növeljük a tudásunk mennyiségi, statisztikai értelemben vett megbízhatóságát. Van tehát az egyszerű, strukturálatlan, rögzített, nyelvi információ, amit szövegnek szokás nevezni, és van a rendezett, strukturált szöveg, ami az adatbázis alapját képezi. Ezen kiegészítéssel immár rendelkezésünkre áll egy információtipológia, amelynek segítségével pontosabb állításokat fogalmazhatunk meg.

| | auditív | | vizuális | |
|---------------|----------------|------------------|-------------------|----------------------------------|
| | nem-nyelvi | nyelvi | nem-nyelvi | nyelvi |
| nem-rögzített | élő zene | élő beszéd | táncszínház | jelnyelvelés |
| rögzített | rögzített zene | rögzített beszéd | fénykép, mozifilm | írás sima szöveg adatbázis |

Az információtipusok meghatározásának egyik alapvető – a fenti táblázatban már érvényesített – szempontja az, hogy az információ **rögzítve** van-e vagy sem. Ha az információt nem rögzítjük, akkor csak egyidejű kommunikációra lehetünk képesek. Az **időben eltolt kommunikációhoz**, vagyis ahhoz, hogy a kommunikáló felek úgy cserélhessenek információt, hogy elérő időben történjen meg a kibocsátás és a befogadás aktusa, az szükséges, hogy az átadni kívánt információt rögzítsék valamilyen hordozóra, médiumra. A rögzítés biztosíthatja, hogy időben később is befogadhatóvá váljék a kibocsátott információ.

A **tartalom** (információ) és a **hordozó** (médium) egységét nevezzük **dokumentumnak**. Dokumentum lehet egy papírra írt szöveg, egy fénykép, egy mozifilm, egy CD-n levő zenemű, egy – valamilyen tartalommal bíró – számítógépes állomány, egy weboldal stb. Ennek megfelelően a dokumentumok tipizálását a két alkotókomponens szerint lehet elvégezni. Más típusú dokumentum egy papírra írt szöveg és egy rajz vagy egy DVD-n tárolt beszéd és egy zeneszám, hiszen a tartalmuk más típusba tartozik. De más típusú dokumentumról beszélhetünk akkor is, ha a tartalmuk megegyezik ugyan, de más hordozójuk. Ugyanazt a mozifilmet tárolhatjuk celluloidszalagon, a tévés infrastruktúra eszközein, DVD-n, és nyilván másfajta dokumentumként kell ezeket kezelnünk, hiszen a hozzáférésüket egészen eltérő technikák segítségével biztosíthatjuk (és megnézhetjük a moziban, a tévékészülék, a számítógépünk monitora előtt ülve, vagy az okostelefonunk képernyőjén).

A dokumentumok alapvonásának, hogy ti. a tárolt információ addig elérhető rajtuk, ameddig a tároló médium létezik, van két fontos következménye. Az egyik az, hogy a médiumok, a tárolóeszközök maguk könnyen sokszorosíthatók, másolhatók. Ez a **többszörözési** tevékenység sokszor jóval kisebb erőfeszítést igényel, mint az eredeti dokumentumpéldány előállítás, viszont a többszörözés lehetővé teszi, hogy a tárolt információ több befogadóhoz eljuthasson akár egy időben. Ez a lehetőség a kommunikált üzenetek terjesztésének hatékonyságát növelheti. A médiumok időben tartós létezésének másik következménye az, hogy a dokumentumaink **halmozódnak**, egyre több lesz belőlük.

A dokumentumok halmozódásával párhuzamosan értelemszerűen felmerül az az igény, hogy a létező dokumentumokat **gyűjteményekbe** rendezzük egyfelől azért, hogy egy helyen tegyük elérhetővé őket, másfelől azért, hogy biztosítsuk azt, hogy minél tovább elérhetők legyenek az időben. Sokféle gyűjtemény létezik: azok a közgyűjtemények (könyvtárak, levéltárak, múzeumok, audiovizuális archívumok), de gyűjtemény egy vállalati dokumentációs osztály vagy egy egyetemi tanszék által őrzött dokumentumhalmaz, de gyűjteményként értelmezhetjük egy magánlakás polcain tárolt könyveket vagy bakelitlemezeket is. A gyűjtemények méretének növekedésével egyre fontosabbá válik az a kérdés, hogy milyen módon lehet hozzáférni a keresett dokumentumhoz a gyűjteményen belül. Ez rámutat arra, hogy az információhoz való hozzáférés sok esetben egyet jelent a dokumentumokhoz való hozzáféréssel (bár a két fogalom nem teljesen azonos).

Amikor a dokumentumainkat gyűjteményekbe rendezzük, akkor új probléma jelentkezik: az egyre bővülő gyűjteményekben meg kell találni valahogy azokat az dokumentumokat, bizonyos esetekben azt az információt, amire a felhasználónak szüksége van. Az alaphelyzet az, hogy tudjuk (feltételezzük), hogy valahol a dokumentumgyűjteményen belül ott van az a dokumentum, vagy ott van az az információ, amit szeretnénk megkapni, de nem tudjuk, hol. Ez az alaphelyzet további kérdéseket vet fel. Az első kérdés az, hogy milyen módon lehet biztosítani a dokumentumhoz, információhoz való hozzáférést. A válaszok feltárása érdekében meg kell vizsgálnunk, milyen információhozzáférési, **információelérési technikák** léteznek, ezeknek mik az előnyeik, mik a hátrányaik. Az információelérés problémája a felhasználók és dokumentumok közti viszony vizsgálatával elemezhető. Ebből vezethetjük le a következő két kérdést, és a válaszok kereséséhez szükséges feladatokat.

A második kérdés az, hogy a felhasználók milyen igényekkel lépnek be abba az információs térbe, ahol válaszokat remélnek maguknak. Ehhez elemeznünk kell az **információs igény** fogalmát, és meg kell adnunk, hogyan tipizálhatjuk a felhasználók információs igényeit. Ez az elemzés az információelérés problémakörét a felhasználó oldaláról közelíti meg.

Az információelérés dokumentum oldalára fókuszálva merül fel a harmadik kérdés: hogyan lehet a dokumentumokat, az elérhető információt rangsorolni. Elképzelhető ugyanis, hogy több ponton is megtalálható a szükséges információ, elképzelhető, hogy több dokumentum is megfelelhet a felhasználó számára, és az is elképzelhető, hogy a valamilyen mértékben megfelelő dokumentumokból, információból jóval több van, mint amennyit a felhasználó be tud, be akar fogadni, fel tud dolgozni. Ennek a problémának a kezelése veti fel annak szükségességét, hogy a vizsgálati szempontok közé felvegyük a **relevancia** problémáját.

3.2.1.1 Információ

A későbbiekben használni kívánt fogalmak vázlatos áttekintése után belekezdünk a projekt számára szükséges elmélet részletesebb kifejtésbe. Fontos rögzíteni itt azt a tényt, hogy a bemutatott információ típusok közül nem fogjuk az összeset használni. A projekt számára elsősorban a szöveg általános és speciális fogalmát, valamint az általános fogalomból levezethető származtatott fogalmakat, a hipertext, az adatbázis és a tudásszervezési rendszerek fogalmát kell minél pontosabban definiálnunk.

3.2.1.1.1 Szöveg

Ha meg akarjuk érteni a digitális világ olyan új jelenségeit, mint a hipertext vagy az adatbázis, akkor a szöveg fogalmából kell kiindulnunk.²⁰ Ferdinand de Saussure elmélete szerint a nyelvi jelenségek két dimenzióban értelmezhetők, kétfajta tengelyre képezhetők le (Saussure 1998). Ez a két dimenzió a szintagmatikus és az asszociatív (vagy paradigmatiszós dimenzió).²¹ A szintagmatikus dimenzióban (a szintagmatikus tengely mentén) ragadhatjuk meg azt, ahogyan a beszélő (író) kifejezésre juttat valamilyen üzenetet a nyelvi elemek egymás után rendezésével, a szavak lineáris összekapcsolásával. Amikor kimondunk (vagy leírunk) egy mondatot, például ezt:

²⁰ Szövegen itt egyaránt érthetünk beszédben vagy írásban létrejött mondatokat.

²¹ Saussure az 'asszociatív' terminust vezette be, de a követői már a 'paradigmatikus' fogalmat kezdték el használni. A két kategória közt – jelen tanulmányban – csak annyi különbséget tételünk, hogy amíg az asszociativitás bármilyen kapcsolatot megenged a nyelvi egységek között, addig a paradigmatiszós minősítés az osztályozási, tipizálási relációt engedélyezi az elemek között. Az előbbi reláció szemantikailag tágabb jelentésű az utóbbihoz képest.

s1) *A Kovács család István fia Monoron lakik, foglalkozása kovács.*

akkor a mondatot alkotó szavak fizikailag „megjelennek“ a térben. Amit kimondunk, egy hangfolyamban hallható, amit leírunk, egy darab papíron (vagy egy képernyőn) látható. Az egymás mellé helyezett nyelvi elemek egyszerre vannak jelen a – vizuális vagy akusztikai – térben, van kiterjedésük. A – kimondott vagy leírt – mondat explicit, valós, jelen van, – Saussure kifejezésével élve – „in praesentia” létezik, amit a szintagmatikus dimenzióban tudunk megragadni. Különbéféle szintagmatikus műveleteket végzünk el, miközben a szavainkból egy teljes mondatot építünk fel: időt, módot, számosságot fejezünk ki, ragokat illesztünk a szótövekhez, egyeztetjük az ige, az alany és tárgy közti viszonyokat, felsorolásokkal, elhagyásokkal élünk stb. A nyelvi kompetenciánk része, hogy szintagmatikai szinten jól-formált mondatokat tudunk képezni (és ezeket a mondatokat képesek is vagyunk megérteni).

Van azonban egy másik nyelvi képességünk is, amely révén a nyelvünk elemeit, a szavakat másfajta módon is össze tudjuk kapcsolni egymással: „azok a szavak, amelyekben van valami közös, összekapcsolódnak az emlékezetben, és így csoportokká alakulnak, amelyekben különféle viszonyok uralkodnak“ (Saussure 1998). Amikor egy mondatot képzünk, akkor a mondat mindegyik új elemét egy nagyobb csoportból választjuk ki, és mondatba végül is bekerül a kiválasztott elem, de a fejünkben ott marad a csoport többi eleme, miközben tudjuk azt is, hogy a csoport elemei közül választhatunk volna mást is, és akkor is ugyanolyan jól-formált mondathoz jutottunk volna. Sokféle szempont szerint csoportosíthatjuk a nyelvi elemeinket: minden főnév, egy adott szó összes szinonimája, a férfi keresztnév, a családnevek, a foglalkozásnevek stb. mind egy-egy egységet alkot. Saussure alapján ezeket nevezzük asszociatív (paradigmatikus) egységeknek, amelyekhez kizárólag az olvasó (hallgató) ember elméjében létező szavak tartoznak. A paradigmaticus dimenzióban a nyelvi elemeket nem a valóságos, hanem a képzelt térben társítjuk egymáshoz, vagyis ez a kapcsolódás implicit, „in absentia egyesít elemeket egy virtuális emlékezeti sorban” (Saussure 1998). A mondatok elemei közt fennálló paradigmaticus kötődések nem látszanak, nem hallatszanak, de ettől még léteznek, hatást gyakorolnak a nyelvi megnyilatkozásainkra. A nyelvünkben rejlő asszociatív (paradigmatikus) kapcsolatot a következő két példamondattal szemléltetem.

s2) *A Gyula család Gyula fia Gyulán lakik, foglalkozása gyula.*²²

Ez a mondat formailag helyes, szemantikailag értelmes, bár – a 'gyula' szóalak négyszeri használata miatt – kicsit szokatlan. A megelőző mondatához képest alakilag is (a szavak szintjén), jelentését tekintve is komoly eltérést mutat, ám nyilvánvalóan érezzük, hogy van valami szerkezeti hasonlóság köztük. A második mondat úgy állt elő, hogy az első mondat bizonyos szavait kicseréltük valami másra, miközben bizonyos szabályokhoz azért igazodtunk, nem lehetett volna bármit bármire kicserélni. A következő mondatot mind alakí, mind tartalmi szempontból helytelennek tartjuk.

s3) *A miért család szélessáv fia megnézhetné lakik, foglalkozása kimosták.*

Az első mondat ugyanazon pontjain cseréltük ki a szavakat a második és a harmadik mondatban, de mást eredményre jutottunk: az s2)-ben értelmes, az s3)-ban értelmetlen mondatot kaptunk. Azért, mert a második mondatban a paradigmaticus kapcsolat mentén hajtottuk végre a szócsereket, míg a harmadik mondat esetében nem. A „Kovács” helyére bármit beírhatunk a családnevek halmazából, értelmes változtatást hajtunk végre (ilyenkor a családnév paradigmacsoportot használjuk). Hasonlóképpen a „Monor” kicserélhető bármelyik más településnévvel, egy foglalkozásnév helyett vehetünk egy másikat stb., és az ilyen változtatások után értelmes mondatokat kaphatunk. A mondatok bármely elemét ki lehet

²² A gyula a magyar történelem korai időszakában méltóságnév volt. A mai értelemben vett foglalkozások közé nem igazán lehet besorolni, csak a későbbi mondanivaló kedvéért használom itt – kicsit sután – a 'méltóságnév' megjelölés helyett a 'foglalkozás'-t.

cserélni, és ha megfelelő paradigmikus kapcsolat mentén tesszük ezt meg, akkor értelmes, szemantikailag jól-formált mondatokhoz juthatunk.

A nyelv szintagmatikus elrendezése során szintagmatikus relációkat alkalmazunk. Jurij Šrejder öt szintagmatikus relációt azonosított (Šrejder 1975), amelyek segítségével felírható minden mondat szintagmatikai szerkezete.

- következés
- közvetlen függőség
- egyeztetés
- egyneműség
- összetevőjének eleme

A fenti relációkat az alábbi módon értelmezhetjük.

1) A **következési** reláció egyszerűen azt a tényt fejezi ki, hogy a mondat szóelőfordulásai szigorú, rögzített sorrendben következnek egymás után. Ez a reláció egy lineáris elrendezés.

2) A **közvetlen grammatikai függőségi** relációt a mondat elemei közötti nyelvtani, grammatikai viszonyok gyűjtőfogalmaként használhatjuk, éppen ezért konkrét tartalmat csak altípusainak tulajdoníthatunk. A mondatok konkrét elemzéséhez, értelmezéséhez szükség van a közvetlen grammatikai függőségi reláció altípusainak részletes kibontására, de a szöveg meghatározásához elégséges a közvetlen függőségi relációnak ez az általános, altípus nélküli megközelítése is (olyan konkrét függőségek tartoznak ide, mint például az alany vagy a tárgy függése az állítmánytól).

3) A **egyeztetési** reláció a mondat szóelőfordulásai, kifejezések előfordulásai között teremt kapcsolatot azzal, hogy a cselekvések, történések módját, idejét az eseményekben, történésekben érintettek számosságát, személyét stb. igazítja egymáshoz

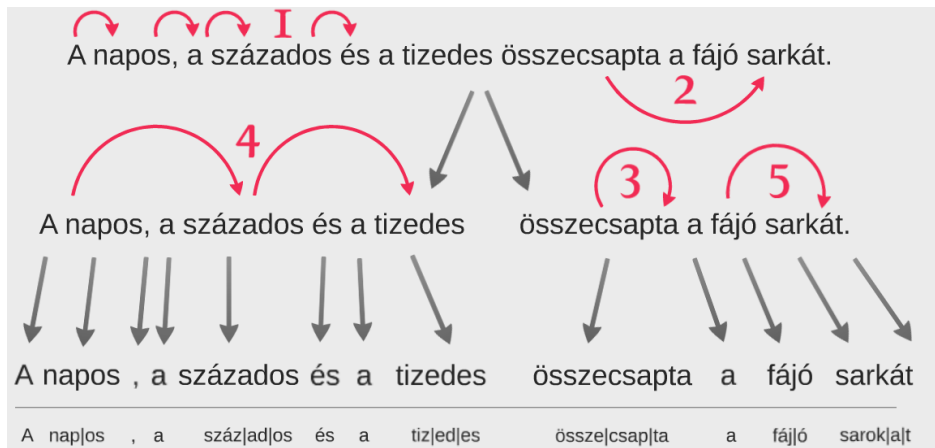
4) Az **egyneműség** reláció az azonos mondatrészeket kapcsolja egymáshoz valamilyen szintaktikai megoldás segítségével. Például ha több személy társas alanyként vesz részt egy cselekvésben, akkor ezt az egyneműség relációjával úgy lehet kifejezni, hogy a személyek neveit, egymástól vesszővel elválasztva, egyszerűen felsoroljuk.

5) Az **összetevő része** reláció a mondat szintaktikai szerkezetében nyilvánul meg. Azáltal, hogy a mondatot felbonthatjuk igei és névszói szerkezetekre, majd mindkét ágon egyre kisebb és kisebb egységekre oszthatjuk a mondat összetevőit, olyan részekre szedett mondatokhoz jutunk, amelyben minden értelmesen elkülöníthető szövegrész (mondatrész, kifejezés, szó) valamely nagyobb szövegrész, összetevő eleme

Vegyünk egy példamondatot, és próbáljuk meg elemezni a szintaktikai szerkezetét.

s4) *A napos, a százados és a tizedes összecsapta a fájó sarkát.*

A mondatban alkalmazott szintagmatikai relációkat tüntetjük fel a következő ábrán.



- 1) A **következési** reláció kapcsolja össze az egymást követő szavakat a mondaton belül. Az 'a' névelő után következik a 'napos', majd egy ', ' tagolójel jön, aztán egy újabb 'a' névelő és így tovább.
- 2) A **közvetlen grammatikai függőségi** reláció segítségével a mondatrészek, szavak közti kapcsolatot ragadhatjuk meg. A példamondatunkban az 'összecsapta a ... sarkát' kifejezésen belül a 't' tárgyrag alkalmazásával fejezzük ki, hogy az ágens (ágensek) milyen tárgyhoz kapcsolódva cselekszik (cselekszenek). A tárgyrag jelzi, hogy az 'összecsapás' cselekvése a 'sarok'-ra irányul.
- 3) Az **egyeztetési** reláció a mondat összetevőit illesztik egymáshoz a cselekvések időbeliségét, módbeliségét, a cselekvők személyét, számosságát illetően. Az 'összecsapta' szóban található 't' múlt időt
- 4) Az **egyneműség** reláció 'a napos, a százados és a tizedes' kifejezésen belül érhető tetten, amikor is az állítást három cselekvőre vonatkoztatjuk egyszerre. A mondaton belül van egy csoport ('a napos, a százados és a tizedes'), amely elemei azonos szerepben vannak a mondaton belül. és mindhárom ágensre egyszerre (egyetlen mondaton belül) fejezzük ki, hogy milyen cselekvést végeznek el (ontológiai értelemben egyébként külön-külön). Ennek gazdaságossági és/vagy stilisztikai oka van. Az egyneműségi reláció hiányában a fenti mondat tartalmát a következő három mondattal adhatnánk meg.

- s5) *A napos összecsapta a fájó sarkát.*
- s6) *A százados összecsapta a fájó sarkát.*
- s7) *A tizedes összecsapta a fájó sarkát.*

A három mondatos megoldás helyett takarékosabb, ha az egyneműségi reláció alkalmazásával ugyanannak az üzenetnek a közvetítését rövidebben oldjuk meg.

5) Az **összetevő része** reláció a mondat egészén belül a belső szerkezeti egységekre bontás során érvényesül, amikor feltárjuk a mondat igei és névszói szerkezetét. A mondaton belül létező szerkezeti hierarchia bármely pontján előfordulhat, hogy az adott rész több összetevőre bontható, illetve az is fontos, hogy az adott rész beletartozik valamely fölérendelt szerkezeti egységbe. Ezt a –partitív – tartalmazási viszont ragadhatjuk meg az összetevője része szintagmatikus relációval. Példaként hivatkozhatunk itt a 'fájó sarka' kifejezésre, amely egyfelől egy szerkezeti egységet alkot, másfelől két további részre osztható. A 'fájó' és a 'sarkát' szavak külön-külön a 'fájó sarkát' szerkezeti egység részét képezik.

* * *

Lehet vitatni, hogy a Šrejder által javasolt szintagmatikus relációk megfelelők-e, kielégítőek-e, elégségesek-e, de ez a jelen kontextusban nem igazán lényeges kérdés. A kutatás számára ugyanis nem a szintagmatikusan rendezett mondatok elemzése a fontos. Sokkal inkább arra van szükségünk, hogy minél pontosabban értelmezzük a saussure-i paradigmatisz dimenzió igazi jelentését.

Ehhez az értelmezési kísérlethez vegyünk a következő példamondatot!

s8) *A kutya megugatja az utcán elhaladó embert.*

Ez a mondat a szintagmatikus tengely mentén (vízszintesen) jól-formált módon van elrendezve. A szintagmatikus relációk alapján feltárható a struktúrája. Pontosan ez a mondatstruktúra azonban kitölthető másfajta szavakkal is, amikor a mondat jelentése nyilván megváltozik, de a szerkezete nem. Az alábbi táblában látható mondatok mindegyikének hasonló a formátuma, de eltérő jelentése.

| | | | | | | | |
|---------------------|----|----------|------------|----|-------------|----------|----------|
| ↑ paradigmatikus | a | macska | megfogja | az | asztalon | talált | húst |
| | a | fiú | megtanulja | az | iskolában | feladott | leckét |
| | az | oroszlán | megeszi | a | bozótosban | elfogott | gazellát |
| | a | buldózer | lerombolja | a | szomszédban | álló | házat |
| | a | kutya | megugatja | az | utcán | elhaladó | embert |

Milyen szintagmatikus mondatoknak azonos a szintaktikai szerkezete? Azt, hogy ugyanazokkal a szintagmatikus relációkkal írhatjuk le őket. Ebből következik az is, hogy a formai szerkezetben azonos helyen álló szavakat ki lehet cserélni anélkül, hogy a szintaktikai jól-formáltságot elveszítenénk. Természetesen szemantikailag könnyen értelmelenné válhat az így átalakított mondat, de fontos tudnivaló az is, hogy nem minden esetben. Kicsit furcsa, de értelmes mondatok lehetnek a következők.

s9) *A kutya megeszi a bozótosban talált húst.*

s10) *Az oroszlán megeszi a szomszédban talált embert.*

s11) *A fiú lerombolja a szomszédban álló házat.*

s12) *A kutya megeszi az utcán elfogott egeret.*

A következő mondatokban viszont már nem látunk értelmet.

s13) *A bulldózer megugatja az iskolában feladott gazellát.*

s14) *Az oroszlán megtanulja az iskolában feladott leckét.*

s15) *Az oroszlán megeszi az iskolában feladott gazellát.*

s16) *A fiú megeszi az asztalon elhaladó húst.*

Nem tartjuk értelmesnek őket, mert a bulldózer nem ugat, az oroszlán ugyan megtanulhatna sok mindent, de nem tanulja meg az iskolában feladott leckét, az oroszlán megehetné a gazellát, de a gazellát nem szokták iskolában feladni, a fiú pedig ehetne húst, de a húsok nem szoktak az asztalon elhaladni. Persze, lehet olyan értelmezési kontextust találni, teremteni, amelyben még az ilyen mondatok is jelentéstelivé válhatnak. Az 'iskolában feladott gazella' kifejezés alatt érthetnénk azt is, hogy egy gazellát az iskolában feldolgoztak, becsomagoltak, postakocsira tettek, és elindították valamely irányban, ám útközben egy oroszlán ezt a csomagot megszerezte magának, és megette a benne talált gazella húsát. A következő mondat – a szavak elsődleges, hétköznapi értelme alapján – szemantikailag értelmetlennek tűnhet:

s17) *A macska lerombolja az asztalon álló házat.*

de ha a házon egy lego-kockákból, kártyákból vagy akár csokiból tortaként felépített házra gondolunk, akkor máris értelmessé válhat. A nyelv rugalmasága, a szavaink átértelmezhetősége, átvitt értelmű, metaforikus használata nagyon gazdag szemantika kezelésére tesz képessé minket.

Fontos kérdés itt az, hogy milyen módon, milyen szabályok alapján cserélhetjük az azonos szerkezetű mondatokon belül elhelyezett szavakat. Saussure eredetileg az **asszociatív** kapcsolatok kifejezést használta erre, a követői csak később kezdték el a **paradigmatikus** jelzővel felváltani az asszociativitás terminusát. Utóbbi talán egy kicsivel jobban kifejezi azt a minőséget, hogy ebben a nyelvi dimenzióban, ezen láthatatlan tengely mentén tetszőleges kapcsolatot tételezhetünk a mondatokon belül egymással ekvivalens szerepben levő szavak között.

3.2.1.1.2 Hipertext

A Saussure által jelzett kapcsolat asszociatív jellegű, ami megengedi tetszőleges reláció alkalmazását. Kutatásunk szempontjából szigorítanunk kell majd ezen a szabadabb értelmezésen, de érdemes előtte megemlítenünk azt, hogy a digitális világ egyik forradalmian új minősége, a **hipertext** voltaképpen ezt a lehetőséget testesíti meg. A dolgokat kissé leegyszerűsítve mondhatjuk, hogy a hipertext olyan szöveg, amely az egymást követő, szintagmatikusan rögzített propozíciók rendszerében két tetszőleges propozíció két tetszőleges eleme között tetszőleges asszociáció mentén kapcsolatot teremt.

Ahhoz, hogy a hipertext fogalmát értelmezzük a nyelv – Saussure által feltárt – asszociatív dimenziójára hivatkozva, arra a kérdésre kell felelnünk, hogy milyen módokon tudjuk biztosítani azt, hogy az adott írástechnikák segítségével haladni tudjunk az írástartalomban akkor is, ha akkora az írás mennyisége, hogy nem fér ki az írást megjelenítő felületen (papírlapon, képernyőn). Ilyenkor mind a hagyományos, mind a digitális írás esetében tagolni kell valahogy a szöveget. A hagyományos írás korszakában ezt úgy tesszük meg, hogy az írást tagoljuk, és az egymást követő szakaszokat egymás mellé, egymás alá és egymás mögé rendezjük a tér három dimenziójában. A mondatainkat sorokba írjuk, majd ha már nincs hely, a sorokat egymás alá kezdjük írni, és ha a sorokkal betöltöttünk egy lapot, akkor újat nyitunk, a lapokat pedig egymás mögé helyezzük. Így csinálunk könyvet, ami nem más, mint egy háromdimenziós íráshordozó. Az írás befogadására annyi használati útmutatót kell csak kiadnunk, hogy „ha az oldal aljára értél, lapozz”. Ez az ősrégi technika könnyen kezelhető, de van egy hátránya: nem engedi meg, pontosabban nagyon nehezé teszi az oldalak közötti ugrálást, és még inkább az oldalakon belüli pozicionálást.

A digitális írásszöveget képernyőkön keresztül fogadhatjuk be és a hosszú szöveg megjelenítésére hasonló megoldást kell használnunk, mint a hagyományos szöveg esetében. A képernyő felel meg a hagyományos papírlapnak, és ha itt is több lapnyi/képernyőnyi olvasnivaló van, akkor lapoztatni kell az olvasót. Ehhez valamilyen lapozógombot kell kitenni a képernyőre, amelyre kattintva az olvasó megkaphatja a következő oldal tartalmát. A digitális szövegkezelés azonban lehetővé teszi azt is, hogy ne csak a következő oldalra lehessen lapozni, de a szöveg tetszőleges pontjára el lehessen ugrani, amit megint csak könnyedén, egyetlen kattintással meg lehet oldani. A hipertext annyit tesz, hogy a szöveg bármely pontján elhelyezett valamilyen jelzéssel a szöveg bármely más pontjára el lehet ugrani.

A hipertext jelenségét úgy értelmezhetjük, hogy a hipertext azt valósítja meg, amit Saussure szerint a nyelvünk felkínál az asszociatív dimenzióban. A mondat valamely összetevőjéhez hozzákapcsoljuk a szöveg egy másik pontját, valamilyen asszociatív kapcsolatot tételezve köztük. Saussure pont ezért nevezte el asszociatívnak ezt a viszonyt. A digitális szövegkezelés pedig azt képes biztosítani, hogy a nyelvünkben létező elvi kapcsolódási lehetőségeket a valóságban is könnyen realizálni lehessen. Saussure szerint az asszociatív kapcsolat értelme bármi lehet, és ez igaz a hiperlinkre is: a szöveg két pontja között bármilyen értelmezés mentén létrehozhatunk ugrási lehetőségeket.

• * * *

A tanulmány számára azonban nem a hipertext, hanem az adatbázis fogalma a fontos. Utóbbi értelmezésében Saussure elmélete továbbra is hasznos lehet, csak a terminológián kell változtatni. Úgy, ahogy ezt Saussure követői is megtették, amikor a 'paradigmatikus' jelzőt kezdték el használni az 'asszociatív' helyett.

Az adatbázis fogalmának értelmezéséhez első lépésként egy olyan technikára kell rámutatni, amelyet csak írásban lehet alkalmazni (ezt nevezzük grafonyelvi technikának), mert csak ennek segítségével lehet értelmezni azt, hogy mi az adatbázis lényege.

3.2.1.1.3 Grafonyelvi technika

Walter Ong Jack Goody egyik könyvére (Goody 1977) hivatkozva nagyon erős állítást tesz:

„Goody ... részletesen vizsgálta a táblázatok és listák – köztük a naptár – poétikai szerepét. Ezen eszközök használatát az írás teszi lehetővé. Valójában az írást nem utolsósorban éppen listához hasonló szövegek előállítására találták ki: a lehető legősibb írás, melyet ismerünk, az i.e. 3500 körül feltalált sumer ékírás pénzügyi elszámolások rögzítésére szolgált.” (Ong 2010: 89)

Ong könyvének egy másik fejezetében részletezi, milyen is volt az első írás, és miért valószínűsíthető, hogy gazdasági, adminisztrációs érdekek kielégítésére jött létre – legalább részben (Ong 2010: 79). Goody az Ong által hivatkozott könyvében vezette be a **grafonyelvi technikák** kifejezést, amivel az íráskultúra azon jelenségeire utalt, amelyek nem lehetségesek a szóbeliség keretei között.²³ Goody a következő példákat hozta fel az általa teremtett fogalomra.

„Először is létezik olyan elkülönített SZÓ, betű, illetve szám, amelyet skatulyába vagy rubrikába zárt logogramként foghatunk fel. . . .

Második helyen említeném a LISTÁ-t, vagyis azoknak a nyelvi, numerikus és más grafémikus feljegyzéseknek (pl. ikonogramok) valamely oszlopát (illetve sorát), amelyeket néha sorszámmal vagy betűjellel is ellátnak. A listáknak, akár csak a szavaknak, van elejük és végük, és bármit vegyenek is lajstromba, bizonyos értelemben véve 'dekontextualizáló' jellegűek.

Harmadik helyen kell szólnunk a TÁBLÁZAT-ról, ami rendszerint nem más, mint valamilyen kételemű lista. Negyedikként – és világosan elhatárolva a többi technikától – szólnunk kell még a MÁTRIX-ról is, amit az oszlopok és sorok pluralitása jellemez.” (Goody 1998)

A lista, a keresztretjvény vagy a táblázat szerepel Goody kiemelt példái között.

²³ A szóbanforgó fejezetet lefordították magyarra is, innentől kezdve erre hivatkozom (Goody 1998).



grafonyelvi technika: lista

| | | | | | | | | | | | | |
|---|------|------------------------|----------------------------------|---------------------------|-------------------------------|--------------------------|---------------------------|----------------------------------|-------------------------|--------------------|--------------------|---|
| AZ ELSŐ MAGYAR NEMZETKÖZI EREH MEGSZERZŐ JE (1993) | META | ÜVESKE-DESRE-DKTÁT | ... JÜ-TANG: KI-NAI IRO | CSINOS-TISZTA | TALAJT-FELTÜR | TENSZ-ÜTES | NIT-ROGEN | KENYAI-AUTÓJEL | LEN-VEDEKI | LENGYEL-FOCISTA | KOCS-MÁBÁ-JÁRO | AZ ELSŐ MAGYAR FERTI SOFT VILÁGBAJNOK SZOKNYAREDD |
| AZ UTCÁRA HULLÓ ELSŐ MAGYAR STEEL EUROPA BAJNOKOK (FÁRSI) | | | | | | | ... FRANCE: FR. IRO-ÖTLET | | | | | PROMETIUM VEGYÜ-LETFELE |
| HÁZLAT-ELTÜR | | | | | POLGARI-TÖRVÉNY-KÖNYV-A LENTI | | | | BÉR-AUTÓS-ÁLLOMÁS-RESZE | | | |
| JÓSLAT-RESZET | | | FRANCIA-RENDEZŐ-DUGI ...: SZIGET | | | | | OSZTRÓG-VÁROS-ÁLLAM-KACER | | | | |
| CSONTHÉ-JAS GYÜ-MÜLCS | | | | SZEP-RÓVÁR-KANADAI-ENEKES | | | | | | KORTÁRS-IRO-KÖRZET | | |
| BÉRÜ-GOTT | | | | | KUTYA-VEZETŐ-RÓNA | | | | | | | TOVÁBBÁ-POLINE-ZIAI NÉP |
| NAHÁT | | CINE-ZETT-APRÓ-DISZHAL | | | | | FEKETE-ÉDES-TERMÉK-GABONA | | | | | NITRO-GEN-ERDELYI-FILMKEM |
| SZINESZ-MARKA | | | | | IDEGEN-NŐI NEV-ABCÖG | | | | | | AZ ÁFA-IS EZ-TASAK | |
| ALÁÍRÁS-UTÁNI-RÓV | | | LÉTRE-HOZ-BORSODI-HELYSÉG | | | | | GODOL-MŰVE-FR. RAJZ-FILM-RENDEZŐ | | | | |
| KIS-OLIVÉR | | | | NŐI-RENEV-ESZAKI-FFINEV | | | | | | SEMI-NÉP-SZENT-KEP | | |
| MEGADÁS | | | | | | | | MÁRIA-BEKEZVE ... SHOP-BOYS | | | | |
| ITTRIM-VEGYJELE | | ORSZ-FFINEV-ÜRES-NEV: | | | | | ATOM-VÁROS-MENET-BEN VAN! | | | | | BURK-BAN VAN-SORVÉD: |
| TÓNINAK-BECEZIK | | | | | | ÉRVEK-KEL-I-GAZDOLT-KILO | | | | | | |
| | É | | | | | | | | | | | |

grafonyelvi technika: keresztrejtvény

| Cross-Tabular | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---------------------|-------|--------|---------|---------|-------|---------|---------|
| Daniel Wood | car | boat | bus | truck | van | bike | moped |
| Stephen Baker | cat | dog | bird | mouse | rat | emu | cow |
| Bevan Campbell | blue | green | white | black | brown | red | grey |
| Vincent Hsu | up | down | left | right | back | forward | stop |
| Paul Hutton | one | two | three | four | five | six | seven |
| Catherine Dillimore | rugby | soccer | cricket | netball | touch | hockey | bowls |
| Robert Sinton | gin | vodka | beer | wine | rum | whiskey | bourbon |
| Wendy Wethey | run | walk | sprint | jog | amble | gallop | skip |
| Phil Taylor | nba | nfl | nhl | mib | wwe | mls | mma |
| Sam Sehnert | oak | fur | redwood | pine | maple | beech | hickory |

grafonyelvi technika: táblázat

Goody példáin túl említhetjük még a jogi szövegekben gyakran előforduló „jelölt szerkezeti egyégek” jelenségét is, amikor egymás alá pontokba szedve írnak be egy jelölt (számozott) listába olyan hiányos mondatokat, amiket írásos formában ki tudunk egészíteni teljes mondatokká, míg szóban erre nem volnánk képesek (Syi et al. 2015). Például:

„3:239. §(1) Az alapszabály az alaptőke húsz százalékát meg nem haladó mértékben rendelkezhet olyan részvény kibocsátásáról, amely alapján a részvényre vonatkozóan

- a) a részvénytársaságot vételi jog;
- b) a részvényest eladási jog; vagy
- c) a részvénytársaságot vételi jog és a részvényest eladási jog

illeti meg.” (2013. évi V. tv.)

Ezt a jogszabályrészletet olvasva értelmezni tudjuk a benne rejlő három mondatot, tudjuk, hogy a pontokba szedett hiányos részmondatok elé kell gondolni a bekezdés felvezető mondatát, illetve mindegyik részmondat végére kell illeszteni a legutolsó sorban látható szövegrészt. Ezekre a műveletekre szóban nem lennénk képesek. Írásban „szabályosan”, teljes mondatokat alkalmazva az alábbi módon kellene kifejezni a fenti jogszabályrészletet:

(1a) Az alapszabály az alaptőke húsz százalékát meg nem haladó mértékben rendelkezhet olyan részvény kibocsátásáról, amely alapján a részvényre vonatkozóan a részvénytársaságot vételi jog illeti meg.

(1b) Az alapszabály az alaptőke húsz százalékát meg nem haladó mértékben rendelkezhet olyan részvény kibocsátásáról, amely alapján a részvényre vonatkozóan

a részvényest eladási jog illeti meg.

(1c) Az alapszabály az alaptőke húsz százalékát meg nem haladó mértékben rendelkezhet olyan részvény kibocsátásáról, amely alapján a részvényre vonatkozóan a részvénytársaságot vételi jog és a részvényest eladási jog illeti meg.

A jelölt szerkezeti egységek egyébként Goody listájának felelnek meg, de elterjedtségük és fontosságuk miatt érdemes őket saját címkével illetni.

Goody grafonyelvi technikái közül azonban a táblázat az igazán fontos. Mi a tábla? Olyan szövegelrendezési mód, amely nemcsak sorokba helyezi el a szövegegységeket, de a szövegegységek oszlopokba való rendezésére is figyel. Lássuk, hogy mit is jelent ez pontosan, hogy lehet értelmezni, és mi köze ennek az adatbázishoz.

3.2.1.1.4 Adatbázis

A hipertext esetében feltételezett szabad asszociáció nem engedi meg azt, hogy szigorú(bb) struktúrát képezzünk a mondataink szemantikai tartalmára támaszkodva. Márpedig sokszor erre lenne szükségünk, és ezen igény kiszolgálására van is megfelelő technológia: **adatbázisok** létrehozása strukturálatlan nyelvi szövegeink alapján. A továbbiakban ezzel a lehetőséggel foglalkozunk alaposabban, mert az így nyerhető megfontolások segítséget nyújthatnak a későbbi fejlesztési munka során is.

A kérdés tehát az, hogy mit jelent az adatbázis fogalma, milyen feltételekkel definiálhatjuk a saussure-i elméletre, különösen a paradigmikus tengely fogalmára támaszkodva.

Vegyük a következő példamondatokat!

- s18) *A napos összecsapta a bokáját.*
- s19) *A tizedes jelent a századosnak.*
- s20) *A százados megfenyítette a napost.*
- s21) *A napos eltörte a vállát.*
- s22) *A százados kitüntette a tizedest.*

A korábbi példáinkhoz hasonlóan most is kijelenthetjük, hogy ezek a mondatok hasonló szintagmatikai szerkezettel rendelkeznek. A kérdés az, hogy milyen közös jelentéssel írhatjuk le a mondatokon belül azonos funkciókat betöltő szavakat. Ha a szintaktikai jól-formáltság, a szintagmatikai illeszkedés miatt szükséges nyelvi konstrukciókat (jeleket, ragokat) lefejtjük a szavakról, akkor a fenti mondatokat az alábbi szemantikai fogalmakkal írhatjuk le egységiesen.

| ágens | cselekvés | páciens | páciensrész | |
|----------|-----------|----------|-------------|----------------------|
| napos | összecsap | | boka | <= predikátum |
| tizedes | jelent | százados | | <= szintagmatikailag |
| százados | megfenyít | napos | | <= "nem-teljes" |
| napos | eltör | | váll | <= prozódiciók |
| százados | kitüntet | tizedes | | <= |
| ↑↑ | ↑↑ | ↑↑ | ↑↑ | |

osztályba tartozás

A táblázat fejlécébe írtuk fel azt, hogy milyen általános fogalmak alá sorolhatók be az elemzett mondatok egyes – a paradigmikus tengely mentén összetartozó – komponensei. A fejlécben voltaképpen egy predikátumot írunk fel, amit természetes nyelven így mondhatnak el: *egy ágens cselekvő valamit cselekszik, ami hatással van egy páciens és/vagy egy páciensrész „szerepet” betöltő cselekvőre.* A táblázat következő öt sorában az elemzett mondatok összetevői jelennek meg a megfelelő – predikátum-argumentumok alapján „kijelölt” – helyükön.

Ha ezt a szemantikai keretezést fogadjuk el, akkor újabb sorokba további, más és más jelentésű mondatokat írhatnánk be. Az egyetlen szemantikai megkötés az lenne, hogy az egyes mezőkbe írt szavak (fogalmak) generikus alárendeltjei legyenek a fejlécbe írt predikátum megfelelő argumentumaiként értelmezett szavaknak (fogalmaknak). Ha így teszünk, akkor voltaképp egy olyan szerkezetet hozunk létre, ami kielégíti az adatbázisokra adható meghatározás minden kritériumát. Az adatbázis alapja ugyanis az, hogy egy előre rögzített predikátumszerkezet (séma) minden komponensére megengedünk egy-egy osztályba sorolási relációt, keressük, hogy milyen elemekkel lehet kicserélni az egyes komponensek "tartalmát", majd összegyűjtjük azokat az elem-együtteseket, amelyekre igaznak véljük az adott predikátum-összekapcsolást – a táblázatba rendezés/osztályozás műveletével.

Lev Manovich az új média nyelveként határozta meg az adatbázisokat (Manovich 2001), amit nem nagyon vitathatunk, mégha az adatbázis fogalmát több szempontból másként is értelmezzük.

A tábla grafonyelvi technika, tehát csak írásban alkalmazható. A tábla is szöveg, csak épp – valamilyen módon – strukturált szöveg. A tábla előnye a sima szöveghez képest az, hogy nem-lineáris abban az értelemben, hogy a szöveget (szövegelemeket) nem csak „vízszintesen”, a sorok szerint, de függőlegesen, az oszlopok szerint is lehet olvasni. Ha egy táblában vezetem azt, hogy egy közösségen belül mikor születtek az emberek, milyen magasak, milyen a szemük színe, akkor a 'magasság' fejlécű oszlopba írt szövegrészek alapján ki tudom számolni, hogy milyen az emberek átlagmagassága vagy mennyi a közösség átlagéletkora stb. A tábla ősrégi grafonyelvi technika, és pont az oszlopba rendezésből fakadó előnyök miatt alkalmazzák már évezredek óta. Bizonyoság erre az a rengeteg ősrégi nyilvántartás, leltártábla, leltárkönyv, katalógus, amelyekben táblázatos formában (proto-adatbázisszerűen) kezeltek szöveges nyelvi egységeket elődeink.

A digitális szövegkezelés előnyei között említettem a könnyű manipulálási képességet. Ez az előny jelentkezik a táblakezelés világában is. A számítógépek széles körben használt alkalmazásai között vannak a táblázatkezelők (amelyek „csak” egy táblát kezelnek) és az adatbázis-kezelők (amelyek több tábla összekapcsolására is alkalmasak). Már a táblázatkezelők is ezerszer többet tudnak, mint amit a hagyományos táblában megvalósíthattunk papíron, de az adatbázis-kezelők még inkább képesek kiaknázni a tábla formátum előnyeit azáltal, hogy lehetővé teszik több tábla összekapcsolását. De még mindig nem mondtam el, mit is jelent az adatbázis fogalma.

Az adatbázis fogalmának értelmezéséhez vissza kell térni ahhoz a gondolatmenethez, amelyet a szöveg Saussure-féle értelmezésekor kezdtem el. Idézzük fel a korábban használt két példamondatot.

s23) *A Kovács család István fia Monoron lakik, foglalkozása tanár.*

s24) *A Gyula család Gyula fia Gyulán lakik, foglalkozása gyula.*

A régiek mellé mellé vegyünk fel két további mondatot.

s25) *Gyula gyula, a Gyula családból, lakhelye Gyula.*

s26) *Gyula Gyula a gyulai gyula.*

Van négy – talán szokatlan, de értelmes – mondatunk, amelyek közül az első kettőnek ugyanolyan a szerkezete, az utolsó háromnak pedig megegyezik a jelentése, de eltérő a felszíni szerkezete. Nézzük meg, hogy lehet a négy mondatot adatbázis-szerűen, táblázatban elrendezni. Ehhez előzetesen fel kell tárni a mondatok szemantikai-szerkezeti hasonlóságait, ennek megfelelően ki kell alakítani egy sémát, amit a táblázat fejlécébe írt szemantikai címkékkel fejezhetünk ki, majd ezt követően le kell tisztítani a szavakról a paradigmatisz tengelyen fölösleges, szintagmatikus jeleket, illetve – ha szükséges – a szintagmatikus sorrendet felül kell bírálni, át kell alakítani, végül a mondatok összetevőit be kell írni a táblázat megfelelő helyeire. Valahogy így:

| családnév | keresztnev | foglalkozás | lakhely |
|-----------|------------|-------------|---------|
|-----------|------------|-------------|---------|

| | | | |
|---|------------------------|--------------------|--------------------|
| A Kovács család | István- fia | foglalkozása tanár | , Monoron lakik |
| A Kovács család István fia Monoron lakik, foglalkozása tanár. | | | |
| , a Gyula családból | Gyula | gyula | , lakhelye Gyula |
| Gyula gyula, a Gyula családból, lakhelye Gyula. | | | |
| A Gyulaék | Gyula- fia | foglalkozása gyula | , Gyulán lakik |
| A Gyula család Gyula fia Gyulán lakik, foglalkozása gyula. | | | |
| Gyula | Gyula | gyula | a-gyula |
| Gyula Gyula a gyulai gyula. | | | |

Ha elvégezzük a fenti egyszerűsítéseket a szintagmatikus dimenzióban, és a táblázat fejlécébe írt szemantikus jelzés szerint tesszük be a sorok és oszlopok megfelelő helyére a mondatok megtisztított, lecsupaszított elemeit, akkor nem pusztán egy táblázatot kapunk, hanem egy – a függőleges, paradigmikus dimenzióban – szemantikailag elrendezett táblázatot, ahol az oszlopokba csak a fejléc jelentésének megfelelő értékek kerülhetnek (keresztnevek a 'keresztnev' oszlopba, településnevek a 'lakhely' oszlopba stb.). Ez az oszlopba rendezési művelet voltaképp az osztályozási relációnak felel meg.²⁴

Ha táblázatokba rendezzük a szövegeinket, akkor a saussure-i láthatatlan dimenzióban kifejezhető asszociatív kapcsolatokat leszűkítjük az osztályozási relációkra. Egy tábla szemantikailag kötött, az oszlopokba csak azonos minőségű elemek kerülhetnek. Azért értelmesebb Saussure eredeti 'asszociatív' jelzője helyett itt a 'paradigmatikus' minősítést használni, mert utóbbinak szűkebb a jelentése, és közelebb van az osztályozási művelethez. A tábla az osztályokban (típusokban, predikátumokban, univerzálékban, általános fogalmakban) kifejezhető – a nyelvhasználatunkban mindig benne rejlő – tudás megragadásának technikája.

Ugyanezt az értelmezést kapjuk vissza akkor, ha elolvassuk, hogyan határozza meg Chris Date, az egyik legismertebb adatbázis tankönyv szerzője az adatbázisok alapját jelentő relációt.²⁵

„Minden relációnak van egy fejléce és egy törzse. A fejléc oszlopnév-típusnév párok készlete, míg a törzs a fejléchez illeszkedő sorok halmaza. Adott reláció fejléce egy predikátumnak tekinthető, és a törzs minden egyes sora egy-egy olyan igaz kijelentést hordoz, melyhez úgy jutunk, hogy megfelelő típusú értéket írunk be a predikátum helyettesítő szimbólumai (place holders) vagy paraméterei helyett.” (Date 2000)

Az adatbázis valamely előzetesen rögzített séma szerint értelmezhető, szemantikusan osztályozási, taxonomikus relációk mentén rendezett, szintaktikai jelektől megtisztított, szintagmatikusan nem-rendezett,²⁶ táblázatokba írt, strukturált, digitális szöveg.

Azáltal, hogy az adatbázisokba nem kell (nem szabad) betenni a szintagmatikus jelek nagy részét, vagyis a szavakról le kell csupaszítani a szövegbe tartozáshoz szükséges, de szemantikailag értéktelen szintaktikai jeleket, illetve az adatokat mindig az adatbázis sémájának megfelelően kell elhelyezni a táblázatok megfelelő helyeire, az adatbázisok nagyon hasonlítanak a szóbeliség sematikus, formuláris,

²⁴ Jurij Šrejder gondolatát idézve: „A tárgyak osztályozása a megismerés leghagyományosabb módszere; az osztályozás eredményeként az ismeretek osztályozási táblázatok formájában jelennek meg.” (Šrejder, 2001)

²⁵ Az adatbázisok meghatározására idézhetjük még Hugh Darwen lényegretörően rövid definícióját, ami a tanulmánya címe is egyben: *What a Database Really Is: Predicates and Propositions.* (Darwen 1998)

²⁶ Itt nincs hely annak kifejtésére, miszerint az adatbázisban nem lényeges se az oszlopok, se a sorok sorrendje, viszont mindkét dimenzióban rengeteg lehetőség van az egységek közti rendezésre.

mellérendelő jellegére (Ong 2010). Az adatbázis olyan szöveg, amelyben nem számít a stílus, nincs stilisztikai szempont, viszont gépiesen pontos ismétlődő szerkezetek vannak.

- * * *

Az adatbázis definíciójában kiemelt szerepe van az osztályozási relációnak. Ez rendezi el az információt a sémával definiált fogalmak mentén úgy, hogy a séma adott pontjába (osztályába) kell írni a szemantikailag oda tartozó értékeket. Ettől a relációtól meg kell különböztetni azokat a relációkat, amelyeket akár a sémában, akár a rekordokban veszünk fel. Nézzünk meg pár példát!

| leftconcept | relation | rightconcept |
|-------------|------------------------------------|---------------------------------|
| kovács | generikus alárendeltje | foglalkozás |
| pék | generikus alárendeltje | foglalkozás |
| Kovács | példánya | családnév |
| Szentendre | individuális partitív alárendeltje | Pest megye |
| megye | általános partitív alárendeltje | ország |
| megye | ország | általános partitív alárendeltje |

Ebben a sémában (táblázatban, adatbázisban) csak annyi ontológiai elköteleződés van, hogy vannak fogalmak, amelyeket össze lehet kapcsolni egy relációval. A két összekapcsolt fogalmat jelezzük a sémában a 'leftconcept' és a 'rightconcept' argumentummal, a kapcsolat minőségét, tartalmát pedig a 'relation' címkével jelölt sémahelyre írhatjuk be. Az adatbázis akkor (és azáltal) jön létre, amikor a rekordokba úgy vesszük fel az értékeket, hogy ügyelünk arra, hogy a séma által definiált helyekre írjuk be azokat. A fenti táblázatban szemantikailag inkonzisztens az a rekord, amely a 'relation' argumentum értékei közé felveszi az 'ország' fogalmat (megye | ország | partitív alárendeltje). A séma argumentumainak pontos értelmezésétől függ az, hogy további két rekord konzisztens-e vagy sem. Ha a 'leftconcept' és 'rightconcept' argumentumok alatt általános fogalmakat értünk, akkor helytelen a rekordok közé felvenni azokat, amelyek individuális fogalmakat kapcsolnak össze, tehát ekkor inkonzisztens rekord az, amely a 'Szentendre' és a 'Kovács' individuális fogalmakat kapcsolja össze valami más individuális fogalommal (tehát rossz a 'Kovács | példánya | családnév' és a 'Szentendre | individuális partitív alárendeltje | Pest megye' rekord). Ha a sémát úgy értelmezzük, hogy a 'leftconcept' és 'rightconcept' argumentumokba megengedjük felvenni az individuális fogalmakat, akkor a két rekord megfelelő.

Definiálhatnánk olyan sémát is, amelyben konkrét relációt adunk meg az argumentumok között.

| instance_of_name | name concept |
|------------------|--------------|
| Budapest | településnév |
| Petőfi | családnév |
| Kovács | családnév |
| Szentendre | településnév |
| város | településnév |

Ebben az esetben olyan sémát definiáltunk, amelyben a tábla magával a példánya relációval akar összekötni dolgokat, ezért a 'példánya' relációra nem érdemes argumentumot (oszlopot) felvenni. Ebben az esetben a bal oldali oszlopba valamilyen névfogalom példányait kell beírni, a jobb oldali oszlopban pedig azt rögzítjük, hogy milyen névtípusról (névosztályról) van szó az adott rekordban. Ebben a szemantikai szerkezetbe nem illeszkedik az a rekord, amely a bal oldalon egy általános fogalmat

(‘megye’) vesz fel, hiszen szemantikailag nem helyes azt állítani, hogy a ‘város’ a ‘településnév’ fogalom példánya lenne (ha csak nem lenne egy ‘Város’ nevű településünk, de akkor azt nagybetűvel kellene beírni a rekordba).

A fenti két példa rávilágít arra, hogy a relációkat többféle módon is értelmezhetjük az adatbázisokban. Egyfelől ott van az adatbázisok definitív minőségét meghatározó tipizálási reláció, amely „csak” annyit vár el, hogy az argumentumhelyekre szemantikailag megfelelő értékeket írjunk. Ezentúl felvehetünk a sémába is relációfogalmakat, ahogy ezt tettük az első példában. Ilyen esetben a rekordokon belül a ‘relation’ argumentumhelyre csak relációfogalmakat lehet írni. Ez utóbbiak viszont már a harmadik relációértelmezést jelentik az adatbázisok világában, amikor is azt írjuk le a világból, hogy az adatbázisban milyen relációfogalmakat használunk.

- * * *

Az adatbázis fogalmi meghatározásában nem elsődleges szempont, hogy milyen módon oldják meg az adatok struktúrába (osztályba) tartozásának jelzését. Többféle technika alkalmazható. Lehet táblázatba rendezni az adatbázis komponenseit, lehet gráfszerű megoldást választani, lehet a sémainformációt és a típusok elkülönítését szövegbe ágyazott metaadatokkal, címkékkel jelezni (XML-állományok).

3.2.1.1.5 Tudásszervezési rendszer

Ha az adatbázis fogalmának meghatározásához igénye vettük a nyelv saussure-i értelemben vett paradigmikus dimenzióját, illetve az e tengely mentén értelmezhető osztályozási relációt és e reláció mentén összekapcsolt nyelvi elemeket, akkor felmerül a kérdés, hogy ebben az értelmezési keretben mit mondhatunk a tudásszervezési rendszerek fogalmáról. Mik azok a tudásszervezési rendszerek, és mi a kapcsolatuk az adatbázisokkal?

A tudásszervezési rendszer definícióját, a legfontosabb típusait, illetve a definiálásukhoz szükséges relációtípusokat a kutatás egy korábbi dokumentumában hosszabban kifejtettük, ezért itt nem térünk ki bővebben a definíciós kérdésekre. Az alábbi tudásszervezési rendszereket mutattuk be:

- terminuslista
- osztályozási rendszer
- tezaurusz
- formális ontológia
- folkszonómia

Korábban alaposan elemeztük, formális definíciókat is adtunk az egyes típusokra vonatkozóan, ezért ezekkel a kérdésekkel itt már nem foglalkozunk. Azt azonban fontosnak tartjuk újra és nagyon hangsúlyosan jelezni, hogy a tudásszervezési rendszereket nem elősorban a tárgyszókészletük által lefedett tudásterülettel, vagyis nem elősorban a szemantikai tartalmukkal, hanem a bennük alkalmazott relációkkal, valamint az építésük és alkalmazásuk során érvényesített normákkal jellemezhetjük.

- * * *

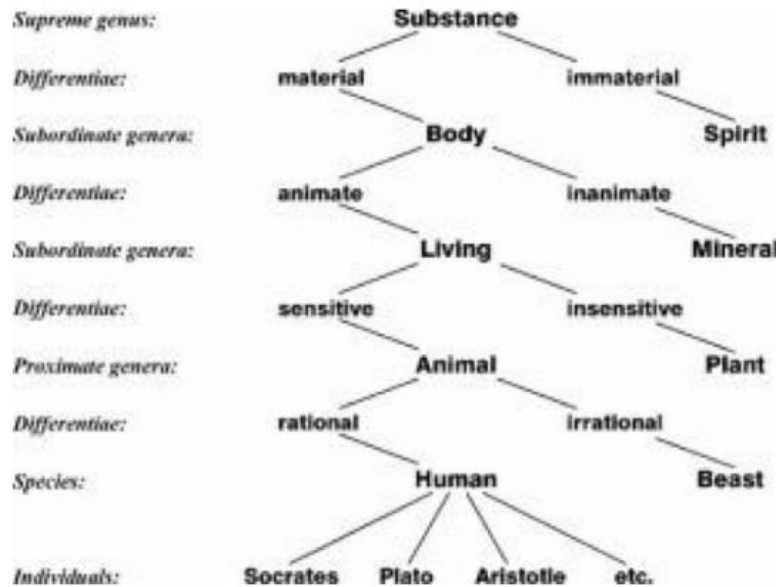
Fontosnak tartunk azonban egy rövid történeti kitekintést a tudásszervezési rendszerek kialakulásáról.

3.2.1.1.5.1 A tudás struktúrába szervezésének történelmi előzményei

A jelenleg használt tudásszervezési rendszerek felépítésével, logikájával kapcsolatos kérdéseket régóta vizsgálják – igaz, sokáig nem ezt a címkét használták a problématerület leírására. A tudásszervezési rendszerek fogalmának pontosabb értelmezéséhez, a velük végezhető, illetve a velük nem végezhető műveletek bemutatásához érdemesnek tűnik egy rövid történeti kitekintést tenni arra, hogy hogyan jutottunk el a mai értelemben vett tudásszervezési rendszerekhez.

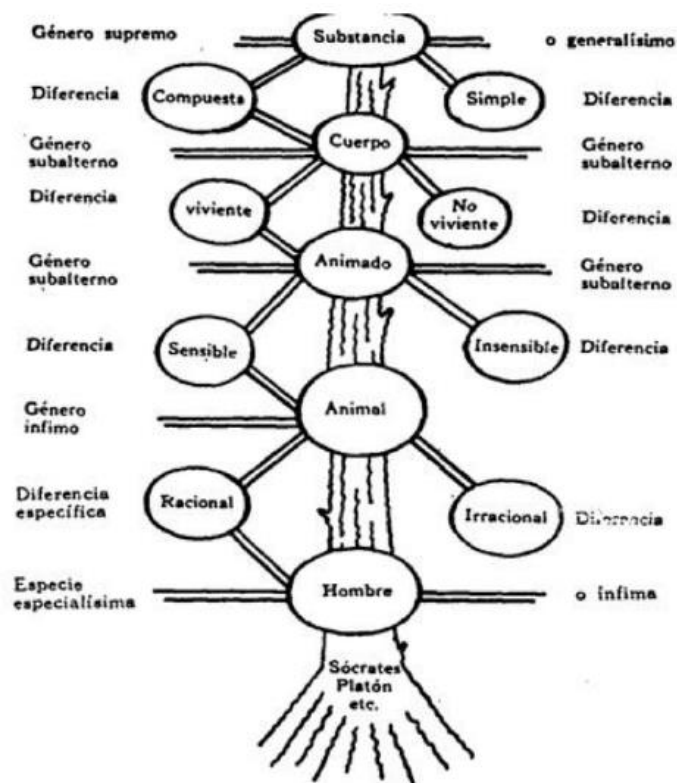
Arisztotelész kategóriatana egyfajta osztályozási rendszernek tekinthető, amelyben a legmagasabb szintű kategóriák egymáshoz való viszonyát, az osztályozási hierarchiába rendelés logikáját rögzítette a generikus alárendeltje reláció segítségével (akkor is, ha ő még nem ezt a megnevezést használta). Arisztotelész szerint az egymás alá- és fölérendelt fogalmak (a ‘genus’ és a ‘species’) közti

különbségeket a „differentia specifica” segítségével adhatjuk meg. A közös tulajdonságok összekapcsolják, a különbözőek elkülönítik a fogalmakat. A fő reláció itt a ’generikus alárendeltje’ reláció, mert biztosítja azt, hogy a tudásunk ilyen elrendezésével teheszünk olyan állításokat, amelyeket már nem kell újra és újra kifejtenuünk, amikor egy általánosabb fogalom alá felvesszünk egy speciálisabbat. Ilyenkor használhatjuk azokat az állításokat, amely a magasabbrendű fogalmakra már kijelentettünk. Arisztotelész kategóriáit mutató ábrára hivatkozva: ha rögzítjük, hogy az ’állat’ (’animal’) ’érző’ (’sensitive’) ’lény’ (’living’), majd kijelentjük, hogy az ’ember’ (’human’) ’racionális’ (’rational’) ’állat’ (’animal’), akkor mondhatjuk azt is, hogy minden ’ember’ ’érző’ ’lény’ is egyben. Erre a minőségre mondjuk azt, hogy a tulajdonságok öröklődnek a generikus alárendeltje reláció mentén.



Arisztotelész osztályozási rendszere

Ezt az arisztotelészi logikát vitte tovább a később híressé vált és Porhipirius fájaként ismert ábra, amin szintén egy osztályozási hierarchia elemei vannak megjelenítve.

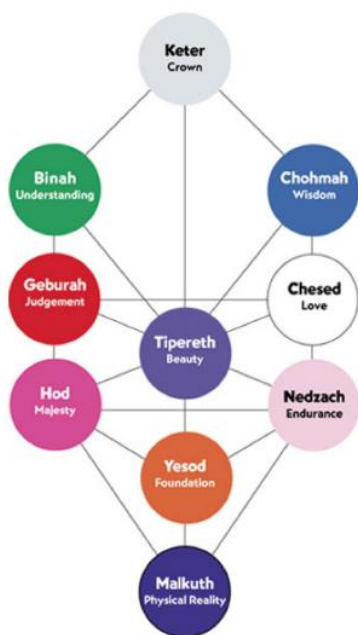


Substancia (género supremo o generalísimo)

Términos intermediarios: {
 géneros y especies sub-
 ordinados: subalternos {
 Cuerpo
 Cuerpo animado
 Animal
 Animal racional
 Hombre (Especie infima especialísima)
 Sócrates, Platón, etc. (Individuos)

Porphyrius fája

Ez a két példa azt mutatja be, hogy a világban létező dolgokat milyen csúcshintű kategóriák segítségével lehet leírni, és ezeket a főkategóriákat milyen kapcsolatok kötik össze.



a Kabbala főkategóriái

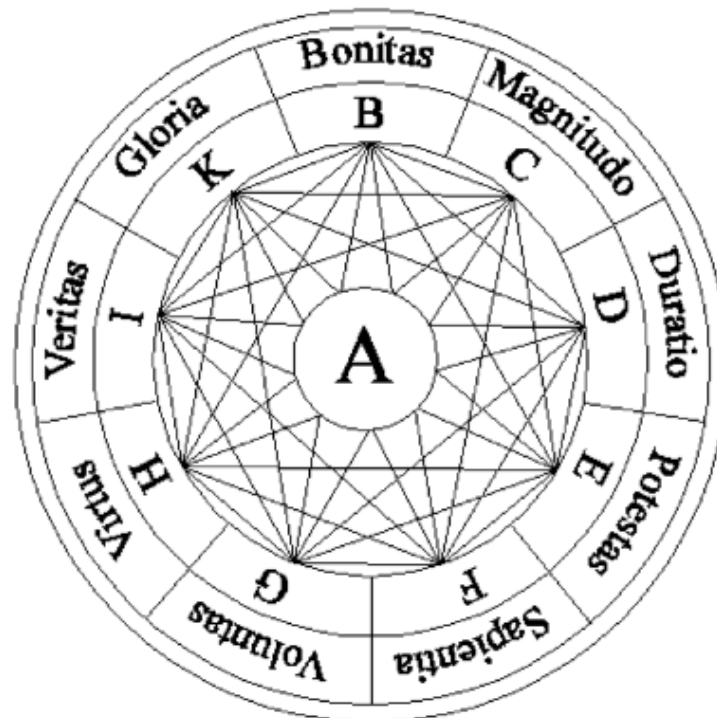
A Széfer Jecirá (Sefer Yetzirah) valamikor a 2-6. században keletkezett könyv, amely a teremtés, a megformálás, az alkotás kérdéseivel foglalkozik. Az 1600 szavas könyv szerint a világ Isten neve betűinek permutációiból keletkezett, tágabb értelemben a világ az ábécé 22 betűjéből és az első 10 szám által jelképezett 10 szefirotból áll. Az elmélet alapján készíthető el a 22 betű kombinációját ábrázoló ‘231 kapu’, amely Isten 231 nevét adja. E felfogás szerint a nyelvnek ontológiai, kozmogóniai értéke van, ami azt is jelenti egyben, hogy a felső szintű fogalmak (szavak) segítségével minden leírható úgy, hogy a fogalmakat (szavakat) egy hierarchikus rendszerben építjük fel – kiindulva a csúcshintű fogalmakból (szavakból). A Kabbala mint egy világleíró séma is erre a rendszere támaszkodik. A lenti táblázatban a főkategóriák jelentésének rövid leírása található.

| | | | | | |
|---------|-------------|-----------|-----|------------------|---|
| Keter | Gloria | Keter | 榮光 | Crown | This represents the first stirrings of the Will within the Godhead, the primal impulse that precedes thought but is necessary for action. It is also called Ayin/Nothingness, because it was out of the infinite void that the Almighty first created. This is the highest, most inaccessible state of the Godhead. |
| Hokmá | Sapientia | Chohmah | 知惠 | Wisdom | This represents the first impulse to create, the flash of intuition or inspiration that precedes conscious thought. It is considered to be a male aspect of the Divinity. |
| Bina | Veritas | Binah | 真理 | Understanding | This represents the point at which Divine inspiration begins to take on definite form. It signifies analytic, distinguishing thought, rather than contemplative intuition. The Binah is the uppermost female element of the Sefirot, and the womb from which the lower seven Sefirot were born. |
| Gedullá | Bonitas | Chesed | 善性 | Love | This represents the generous, benevolent side of God, the quality of unconditional Divine Love. It is connected to the masculine intuition of Chohmah, and serves as a counterpart of the potential destructiveness of Geburah/Judgement. |
| Gebuná | Potestas | Geburah | 能力 | Judgement | This is the aspect of God manifested as a wrathful Divinity of awful punishment. It is considered to be feminine, like the analytic thought of Binah, and serves as a counterbalance for Chesed, so that the world is not so overwhelmed by God's love that it is reabsorbed into the Divine. |
| Tiferet | Virtus | Tiphereth | 力 | Beauty | Also translated as "glory," this is the balancing force between Chesed and Geburah, and is in fact considered to be their offspring. This force unites the upper nine Sefirot. |
| Necá | Eternitas | Nedzach | 永遠性 | Endurance | This represents God's active grace and benevolence in the world—it is a more earthly manifestation of Chesed. |
| Hód | Splendor | Hod | 光輝 | Majesty | This represents the manner in which the judgment of the Deity is dispensed on earth—it is a more earthly manifestation of Geburah. |
| Jeszód | Fundamentum | Yesod | 基盤 | Foundation | This is the channel by which Tiphereth connects to (or impregnates) Shechina or Malkuth, the path through which Divine Creativity and Fertility are visited upon all creation. |
| Malkút | | Malkuth | | Physical Reality | Also referred to as Shechina, this is the culmination and synthesis of all the attributes of God, the quality that links the Eternal Sovereign to the "real" world. It is the Divine Presence, God's immanent and female aspect, the way in which humans experience the Divine. |

a Kabbala főkategóriának leírása

Ramón Llull, katalán filozófus 1275 körül publikálta ‘Ars magna’ című könyvét, amelyben a kombinatorikát használta világmagyarázatra és minden igazságok feltárására. Az Ars magna 1. ábrája a

világ 9 abszolút princípiumát ('principia absoluta') ábrázolja (a kördiagramon ábrázolt kategóriák különböző nyelvű leírását mutatja az ábra alatt látható táblázat).



| fig. A-szobjektum | fig. A-predikátum | fig-T | QUESTIONS | SUBJECTS | VIRTUES | VICES | |
|-------------------|-------------------------|-------------------|---------------|-----------------------|--------------------|------------|-------------|
| B | jóság | jó | különbözőség | Vajon létezik-e? | Isten | justice | avarice |
| C | nagyság | nagy | egyezőség | Mi az? | angyal | prudence | gluttony |
| D | örökkévalóság/tartósság | örökkévaló/tartós | ellentétesség | Miből van? | menny | fortitude | lust |
| E | hatalom | hatalmas | kezdet | Miért van? | ember | temperance | pride |
| F | bölcsesség | bölcs | közép | Mekkora? | képzelőerő | faith | accidie |
| G | akarat | akaratos | vég | Milyen? | érzékelés | hope | envy |
| H | erény | erényes | nagyobbság | Mikor van? | vegetatív funkciók | charity | ire |
| I | igazság | igazságos | egyenlőség | Hol van? | elemiség | patience | lying |
| K | dicsőség | dicső | kisebbség | Hogy van?- Mivel van? | eszköziség | pity | inconstancy |

Ramon Llull 9 abszolút princípiuma

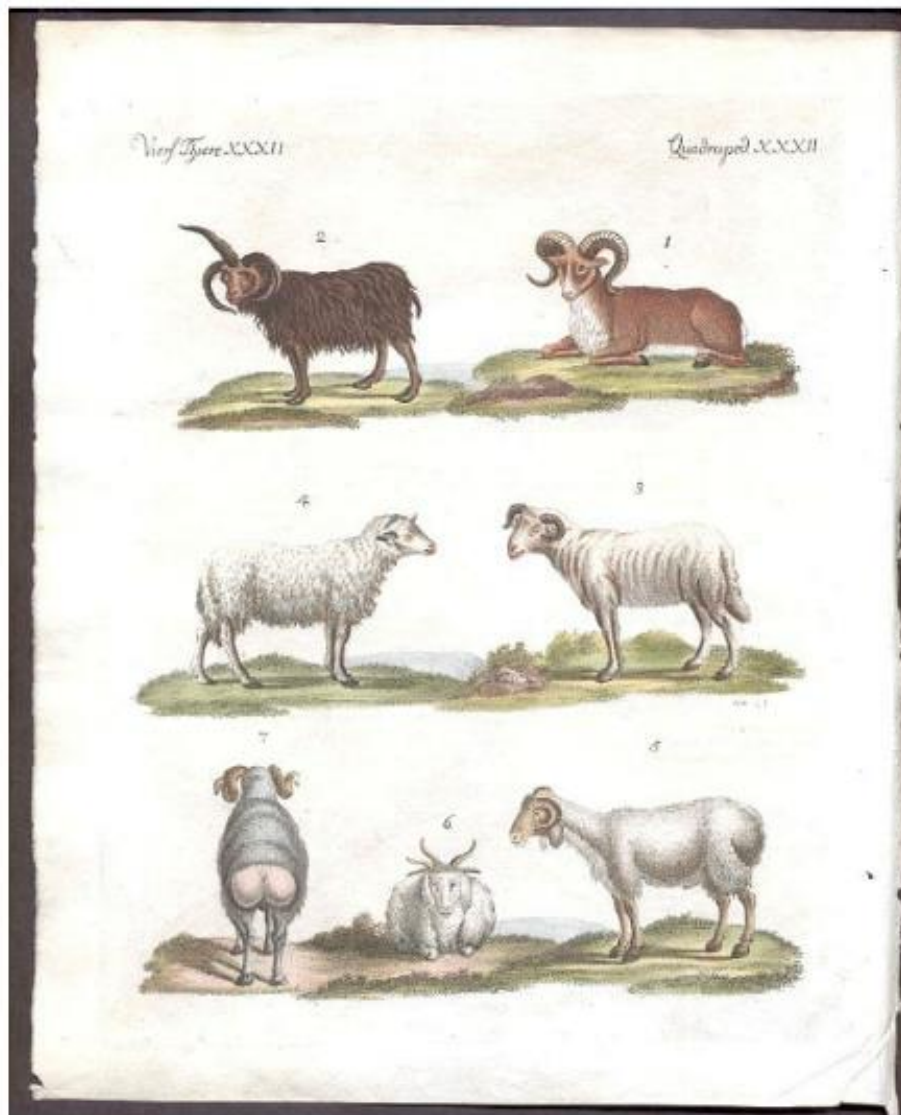
Nagy előrelépést jelentett a tudásszervezési rendszerek irányába az, amikor Georg Philipp Harsdörffer 1651-ben publikálta az 'Ötrétegű gondolkodógyűrű a német nyelvre' (Harsdörffer 1651) című művét. Harsdörffer – betűket, szótagokat és igeikötőket tartalmazó – forgatható korongjaival létre lehetett hozni a német nyelv szinte valamennyi értelmes (és számos jelentés nélküli) szavát.



Harsdörffer gyűrűje

Amikor a könyvtári világ arra a pontra jutott a XIX. században, hogy a könyvek tartalmi leírására felállították a decimális osztályozási rendszert, majd kicsit később a Roget megcsinálta az első tezauruszt, akkor már a mai értelemben vett tudásszervezési rendszerekről lehetett beszélni. Az osztályozási rendszer és a tezaurusz (mint tudásszervezési rendszerek) leírását korábban már elvégeztük, itt csak annyit kell hangsúlyoznunk, hogy nagyjából egy időben jelent meg a mai napig két legfontosabb tudásszervezési rendszer a világban.

Az adatbázist a szöveg fogalma alá soroltuk, vagyis nyelvi rendszernek tekintettük, illetve grafonyelvi technikának is tartottuk, ami azt jelenti, hogy írásnyelvi rendszerről van szó. Ezeket a definitív minőségeket továbbra is elvárjuk mind az adatbázisok, mind a tudásszervezési rendszerek esetében, még akkor is, ha – látszólag – fel lehet mutatni ellenpéldákat. Nézzük meg a következő két ábrát.



képi paradigma: képes állatrendszertan

Ezen az ábrán azonos fajba tartozó állatok képeit láthatjuk, és felmerül a kérdés, nem lehetne-e ezt az ábrát egy primitív képi, de nem nyelvi adatbázisnak minősíteni, hiszen láthatóan ugyanabba a típusba (osztályba) tartozó állatok vannak egy helyen összegyűjtve. Ez a megoldás nagyon hasonlít arra, ahogy egy adatbázistáblában felsoroljuk ezen állatok neveit, és „csak” annyi hiányzik az adatbázis fent definiált fogalmához képest, hogy itt nincs írásbeli információ, tehát nem is lehet grafonyelvi technikáról beszélni.

Fel lehetne tenni ugyanezt a kérdést a következő kép alapján, ahol egy áruházi katalógus képekkel dúsított oldalát láthatjuk.

Men's Sweater Coats, \$2.15.

No. 347547
Men's Sweater Coat, \$2.15
This is a new design in a heavy, warm, and comfortable sweater coat. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, \$2.15. If by mail, postage extra, add 10 cents.



Boys' Fancy Sweaters, 89c.

No. 347547
Boys' Fancy Sweater, 89c
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 89c. If by mail, postage extra, add 10 cents.



Boys' Fancy Knit Waxed Sweaters, \$1.59.

No. 347547
Boys' Fancy Knit Waxed Sweater, \$1.59
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, \$1.59. If by mail, postage extra, add 10 cents.



Children's Fancy Knit Sweaters, 80c.

No. 347547
Children's Fancy Knit Sweater, 80c
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 80c. If by mail, postage extra, add 10 cents.



BOYS' AND YOUTHS' SWEATERS.

For Boys 7 to 14 Years Old. Always Ready to Ship. Order by Mail.

Boys' Half Mock Sweaters, 79c.

No. 347547
Boys' Half Mock Sweater, 79c
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 79c. If by mail, postage extra, add 10 cents.



Boys' Fancy Knit Sweaters, 85 Cents.

No. 347547
Boys' Fancy Knit Sweater, 85 Cents
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 85 Cents. If by mail, postage extra, add 10 cents.



Boys' Australian Wool V Neck Sweaters, \$1.59.

No. 347547
Boys' Australian Wool V Neck Sweater, \$1.59
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, \$1.59. If by mail, postage extra, add 10 cents.



Special Quality Fine All Waxed Sweaters, 80 Cents.

No. 347547
Special Quality Fine All Waxed Sweater, 80 Cents
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 80 Cents. If by mail, postage extra, add 10 cents.



Boys' Fine Waxed Sweaters, 95 Cents.

No. 347547
Boys' Fine Waxed Sweater, 95 Cents
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 95 Cents. If by mail, postage extra, add 10 cents.



Boys' Alternating Stripe Sweaters, 75 Cents.

No. 347547
Boys' Alternating Stripe Sweater, 75 Cents
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 75 Cents. If by mail, postage extra, add 10 cents.



CHILDREN'S SWEATERS.

For Children 2 to 6 Years of Age. Order by Mail.

Children's Fancy Knit Sweaters, 80 Cents.

No. 347547
Children's Fancy Knit Sweater, 80 Cents
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 80 Cents. If by mail, postage extra, add 10 cents.



Children's Sugar Brown Style Golf Coats, 80 Cents.

No. 347547
Children's Sugar Brown Style Golf Coat, 80 Cents
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 80 Cents. If by mail, postage extra, add 10 cents.



Boys' Australian Wool Sweaters, \$1.59.

No. 347547
Boys' Australian Wool Sweater, \$1.59
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, \$1.59. If by mail, postage extra, add 10 cents.



Boys' Fancy Knit Wool Sweaters, 98 Cents.

No. 347547
Boys' Fancy Knit Wool Sweater, 98 Cents
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 98 Cents. If by mail, postage extra, add 10 cents.



Children's Plain Wool Sweaters, 49c.

No. 347547
Children's Plain Wool Sweater, 49c
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 49c. If by mail, postage extra, add 10 cents.




Ladies' All Wool Sweaters, 95c.

No. 347547
Ladies' All Wool Sweater, 95c
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 95c. If by mail, postage extra, add 10 cents.



Boys' Australian Wool Sweaters, \$1.59.

No. 347547
Boys' Australian Wool Sweater, \$1.59
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, \$1.59. If by mail, postage extra, add 10 cents.



Boys' Fancy Herringbone Sweaters, 98 Cents.

No. 347547
Boys' Fancy Herringbone Sweater, 98 Cents
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 98 Cents. If by mail, postage extra, add 10 cents.



Children's Alternating Stripe Sweaters, 75 Cents.

No. 347547
Children's Alternating Stripe Sweater, 75 Cents
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, 75 Cents. If by mail, postage extra, add 10 cents.



Ladies' All Wool Golf Sweaters, \$1.35.

No. 347547
Ladies' All Wool Golf Sweater, \$1.35
This is a new design in a heavy, warm, and comfortable sweater. It features a high collar, a buttoned placket, and a kangaroo pocket. The material is a soft, wool-blend fabric. It is available in various colors and sizes. Price, \$1.35. If by mail, postage extra, add 10 cents.



SOMETHING FREE TO YOU UNDER OUR PROFIT SHARING PLAN. You can now get something for nothing. SEE PAGES 1 AND 4.

képi paradigma: képes áruházi katalógus

Ha eltekintünk a képen olvasható szövegtől, akkor mondhatjuk azt, hogy egy egyszerű képi adatbázist láthatunk, ahol csak egyedi pulóverek vannak összekapcsolva egymással egy „pulóver-osztályon” belül. Eddig valóban erős a hasonlóság a – feltételezett – képi „adatbázis” és a – tényleges – nyelvi adatbázis között. A dolgok azonban gyorsan elkezdenek „bonyolódni”, amint többet szeretnénk kihozni a rögzített információból. A képen szövegszerűen látható, hogy kétféle „alosztályt” elkülönítettek egymástól: a gyerekek és a fiatal fiúk pulóvereit. Ezt az elkülönítést nagyon nehéz lenne csak vizuális jelek, jelzések segítségével megtenni. Bizonyos jegyek esetében még talán mehetne a dolog (mondjuk a fiúk és lányok

esetében), de sok esetben elég reménytelennek tűnő vállalkozás lenne (mondjuk, mint ahogy a mostani példában is van, korosztályi alapon). A képen például a jobb szélén, alul két női alakon látszik, és vélhetően azokon már női pulóverek vannak, amit érdemes lenne lekülöníteni a többi pulóvertípustól, de erre csak képi eszközök révén kisebb az esélyünk. Arról nem is beszélve, hogy nagyon sok tudásunkat, nagyon sok fogalmat nem is lehetne jól reprezentálni képek segítségével. A természeti fajtaikat még csak-csak meg tudjuk jeleníteni, de a dolgok tulajdonságait, a köztük levő kapcsolatokat már nem vagy csak nagy nehézségek árán. Hogyan fejeznénk ki kép segítségével a rokoni kapcsolatokat, vagy azt, hogy valami szebb, ízesebb, értékesebb, drágább stb. A nyelv segítségével mindezeket egyértelműen jelezni tudjuk, képekkel (csak képekkel) nem.

A konklúzióink tehát az kell legyen, hogy a képi adatbázisokról csak egy nagyon limitált értelemben lenne érdemes beszélni, és az ilyen adatrendszereknek kevesebb praktikus haszna lenne, ezért a továbbiakban nem kell foglalkoznunk velük.

3.2.1.1.5.2 Tudásszervezési rendszer és adatbázis

A korábban adott elméleti, formális leírás, valamint az előző fejezetben bemutatott történeti áttekintés után mondanunk kell valamit a tudásszervezési rendszer és az adatbázis fogalmainak egymáshoz való viszonyáról.

Könnyen belátható, hogy a tudásszervezési rendszerek kielégítik azokat a kritériumokat, amelyekkel az adatbázist definiáltuk, ezért azt kell mondanunk azt, hogy a tudásszervezési rendszerek az adatbázisok egy különös típusát jelentik. Miközben felépítjük a tudásszervezési rendszerekben a rájuk jellemző relációk segítségével úgy jellemezzük

3.2.1.2 Dokumentum

A szó elszáll, az írás megmarad – mondják sokszor az írásbeliség „védelmében”. Ezt az előnyt kár lenne tagadni. Az írás jobb emlékeztetőtechnika, mint a beszéd. De amint elkezdjük kihasználni az írás adta előnyt, abban a pillanatban új problémával kell szembesülnünk. Ha mindent leírunk, hogy emlékezni tudjunk, akkor valahogy azt is biztosítani kell, hogy megtaláljuk éppen azt az írásos anyagot, amire adott pillanatban szükségünk van. Az írás univerzumában szükségünk van olyan technikákra, amelyek eligazítanak bennünket. „Behajózunk” az általunk teremtett, egyre bővülő, már-már végtelennek tűnő terekbe, és nagyon hamar pont olyan elveszettnek érezzük magunkat, mint a tengeri hajósok, akik bármerre néztek, mindenütt csak vizet láttak. És hogy haladni tudjanak, navigációs technológiákat teremtettek maguknak. A virtuális világban is pont erre van szükség.

Az aszinkron kommunikáció lényege a tárolás és közvetítés. Ha az üzeneteinket időben eltolva akarjuk közvetíteni, akkor azokat valamilyen hordozón vagy médiumon kell rögzítenünk és tárolnunk, azaz dokumentumokat kell képeznünk, amelyek segítségével már "áthidalhatjuk" az idő csatornáját, vagyis a kommunikációs folyamaton belül elválaszthatjuk egymástól az adás és vétel aktusait.

A dokumentum tehát **tárolja** az üzenetet. És ha tárolja, akkor az így kezelt információ elkezd **halmozódni**. Amint létrehozuk a dokumentumainkat, abban a pillanatban felmerül az az elvárás, hogy a – különböző szempontok szerint összetartozó – dokumentumokat gyűjteményekbe rendezzük, vagyis **archívumokban** tároljuk, és az archívumokon keresztül tegyük mások (sokak, a közönség) számára elérhetővé, újrahasznosíthatóvá a dokumentumainkban tárolt tudást.

A rögzítés eredménye mindig valamilyen **dokumentum**, amely a **hordozó** mint közvetítő (médium), és a rögzített információ mint **tartalom** együttese. Ugyanazt a tartalmat rögzíthetjük többfajta hordozóra (a *Star Wars* negyedik részét mozifilmre, videokazettára, DVD-lemezre, hálózati szerverekre stb.), mindig más dokumentumot kapunk.

Különböző dokumentumtípusokat definiálhatunk aszerint, hogy milyen információtípust milyen hordozóra, médiumra rögzítünk, illetve hogy az adott dokumentumot miként tesszük a közönség tagjai számára elérhetővé. A hagyományos világban a könyv, az újság, a levél szövegeit azon az alapon különíthetjük el a bakelit- és CD-lemezeken, vagy gramofonhangereken tárolt zeneszámoktól, hogy eltérő információtípusba tartoznak (írás, illetve zene). Ugyanakkor ugyanazt a filmet megnézhetjük különböző helyzetekben, különböző hordozókat használva – és ilyenkor eltérő dokumentumnak minősítjük a moziban vetített, a tévében sugárzott, a videokazettán vagy DVD-n megnézett, a hálózatról letöltött vagy épp onnan streamelt mozgóképes tartalmakat.

Viszont ha egyszer elkezdünk dokumentumokat létrehozni (termelni), akkor valahogyan biztosítani kell azt is, hogy a dokumentumok tartalmát újra és újra be lehessen fogadni, vagyis a dokumentumok tárolását és visszakereshetőségét is meg kell oldani, amit a dokumentumokat leíró információval biztosíthatunk. A dokumentum(on tárolt információ) megkeresését segítő információ már másodrendű fogalom, hiszen az információról hordoz információt, ezért nevezik **metainformációnak**. A dokumentumokra rögzített tartalmak elhelyezését, megőrzését, visszakereshetőségét biztosító technikákat, megoldásokat, intézményeket legáltalánosabb szinten archívumként definiálhatjuk.

A hálózati kommunikáció meghatározó vonása az a tény, hogy lassan mindenféle dokumentumhoz a hálózaton keresztül (és csak azon keresztül) juthatunk hozzá. A hálózaton keresztül minden összeér, lassan tényleg egyben lesz az egész, minden ott áll előttünk: filmek, zeneszámok, könyvek, cikkek, hírek, képek, hosszú és rövid szövegek, adatbázisok, és persze mindezeket mindig valamilyen weboldalon keresztül érhetjük el. A **kulturális sokszínűségnek és gazdagságnak** ez a megtapasztalása (megtapasztalhatósága) jelenti a hálózati kommunikáció előnyét, erejét. Jelen van azonban e jelenség árnyoldala is, amelyet az **információs túlterhelés** (information overload) fogalmával ragadhatunk meg. A hálózaton keresztül elérhető információk mennyisége nyilvánvalóan messze meghaladja az emberi feldolgozókapacitás határait. Természetesen senki sem törekszik az összes elérhető információt megszerzésére; az elméletileg elérhető tartalmak döntő része érdektelen a hálózati polgárok számára. Tehát ebben a helyzetben felmerülő kérdés, illetve feladat így fogalmazható meg: hogyan lehetne biztosítani azt, hogy csak a számunkra érdekes, fontos, vagyis **releváns** információt, dokumentumokat lássuk és érthessük el, és sose a teljes tartalomkínálattal kelljen szembesülnünk.

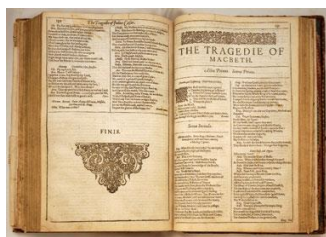
Ehhez a dokumentumok fontossági rendezésére (és ezáltal természetes szűrésére) van szükség, vagyis **relevanciakezelésre**, amelynek révén a számunkra fontosabb dokumentumok kiszűrhetők, és ezáltal könnyebben, gyorsabban elérhetővé tehetők.

Akár figyelembe vesszük a relevanciakezelés kérdését, akár eltekintünk tőle, a hálózati kommunikáció világában mindenképpen felmerül a navigáció problémája, hiszen a weben keresztül elérhető tartalmakhoz, dokumentumokhoz valahogyan hozzá kell férnünk, ez pedig annyit jelent, hogy valahogyan el kell igazodnunk a dokumentumok virtuális terében. Tehát a navigáció alapkérdése az, hogy a felhasználó miként találja meg a hálózaton keresztül elérhető gyűjteményekben, archívumokban, dokumentumokban a számára szükséges és érdekes dokumentumokat, releváns információkat

A hálózati kommunikáció kulcskérdése az, hogy a felhasználók mi módon érik el a hálózaton keresztül az információt. A válaszokhoz elemeznünk kell az **információelérés** (information retrieval) különféle módjait. Ezt a kommunikációs helyzetet úgy érdemes felfogni, hogy a felhasználók a számukra elméletileg rendelkezésre álló teljes **információs egység-folyam** (information item stream) „végtelen” és befogadhatatlan kínálatából mindig valamilyen kisebb, az általuk még befogadható mennyiségű, számukra (vagy valaki más számára) releváns egységekből álló részt (részfolyamot) vesznek ki, kapnak meg és fogyasztanak. A kérdés az, hogy ez a „válogatási” folyamat, ez a **szűrés** hogyan zajlik.

A válaszok keresése során nagyon fontos, hogy a leírások, elemzések, magyarázatok közben használt fogalmainkat minél pontosabban explikáljuk, azaz a rögzített és tiszta értelmezésekkel némileg csökkentjük azt a fogalomhasználati zűrzavart, amely ezen a területen régóta tapasztalható. Az olyan fogalmak, mint **keresés, szűrés, böngészés, szkennelés, eligazodás, navigáció, rendezés, struktúra** stb. különböző szerzőknél, különböző kontextusokban más és más jelentést kapnak, ez pedig nagyon megnehezíti és sokszor pongyolává teszi az információelérés, a navigáció kérdéseiről szóló tudományos-szakmai diskurzust.

A fogalmi tisztázáshoz meg kell vizsgálnunk, hogy a felhasználókat, az információ- vagy tartalomfogyasztókat hogyan minősíthetjük az információhoz, a dokumentumokhoz való hozzáférésük alapján, valamiféle felhasználói aktivitás-passzivitás tengelyhez viszonyítva. Látnunk kell, hogy a felhasználó hogyan kapja meg a „napi információadagját”, és mit tesz ezért ő, illetve mit tesznek ezért mások. Amíg a felhasználói tevékenység minőségét, belső struktúráját, valamint külső meghatározottságait nem látjuk tisztán, addig nem érthetjük meg igazán az információelérés egészének problémakörét. El kell határolnunk egymástól azokat a különböző **információs tereket** is, amelyekkel a felhasználók szembesülnek és amelyekben barangolnak, mert csak így érthetjük meg igazán a különféle információhasználati esetekben rejlő, eltérő lehetőségeket (illetve meghatározottságokat).



A **formai metaadatok** a dokumentumot létrehozásával, kiadásával, terjesztésével kapcsolatos adatokat fogják össze. Ide tartoznak a szerzők, alkotók, kiadók, forgalmazók stb. nevei, a kiadás, megjelentetés, bemutatás ideje, helyszíne, a dokumentumok műszaki, technikai paraméterei, mérete stb. Ezek az adatok függetlenek a dokumentum tartalmától, a tartalom ismerete nélkül is meg lehet adni őket, és éppen ezért – ez nagyon fontos a szabadszavas keresés szempontjából – ezek nem nyerhetők ki sehogy a dokumentum tartalmából gépi úton. A formai metaadatok (nagy részét) “önkéntes” emberi hozzárendelés révén

kapcsolhatjuk a dokumentumokhoz.



A **strukturális metaadatok** írják le az adott dokumentum részei közti kapcsolatokat. Ezek azért fontosak, mert általuk lehet értelmezni, kezelni a részek egészen belüli helyét, szerepét, illetve segítségükkel lehet felépíteni a részekből álló összetett dokumentumokat. Strukturális metaadat például egy filmen belül felhangzó zenemű, egy multimédia jellegű alkotás valamely komponense, vagy a legtöbb webes dokumentum, amely egyaránt tartalmaz szöveget, képet, esetleg beágyazott videót, hangállományt. A html-nyelvben írt webes dokumentumok strukturális metaadatai azok a szövegbe ágyazható vezérlő címkék (tag-ek), amelyek révén az összetett html-dokumentum egészen belül elhelyezhetünk

képeket, címsorokat, hivatkozásokat (, <h1>, <a href> stb.). Ide tartozik még a könyvek fejezetekre osztása, a zeneszámok, filmek belső egységeinek jelzése, de a filmek képi és hangsávjainak összekapcsolásához vagy a transzkriptált beszéd szavainak időbélyeggel való ellátásához, a szavak videófolyamon belüli pozíciójának jelzéséhez is strukturális metaadatokat kell igénybe venni.



Az **adminisztratív metaadatok** segítségével a dokumentumok használatával, fogyasztásával kapcsolatos információkat tudjuk gyűjteni, tárolni. A hagyományos archívumokban ilyen adat a raktári jelzet, amely segít megtalálni egy könyvet a könyvtárban, egy filmtekercset a filmtárban vagy a kölcsönzési adat (ki, mikor, mit kölcsönzött az archívumból). Az adminisztratív metaadatok körébe sorolhatjuk a jogi információkat, a dokumentumok használatára vonatkozó **jogosultsági adatokat** is, bár ezek kezelésére nem annyira az archívumok, mint inkább a jogkezelő intézmények világában került sor. A hálózat világában a digitális dokumentumok elérésével, fogyasztásával kapcsolatos webes

felhasználói interakciókat (letöltések, meghallgatások, megtekintések, olvasások, futtatások, hozzászólások, értékelések, lájkolások, kommentek, szavazások, vásárlások stb. számát) nyomon követő, mindenre kiterjedően és nagyon pontosan gyűjthető, használati, értékelési, **forgalmi adatok** is idetartoznak. Utóbbiak új minőséget, új szintet hoznak el a dokumentumok és fogyasztóik közti viszony leírásában, értelmezésében, a hálózati szolgáltatások újszerűségéhez és sikerességéhez nagyban hozzájárulnak az effajta fogyasztási, értékelési adatok.

A mennyiségeket tekintve mind a tartalmi, mind a forgalmi, mind a strukturális metaadatok nagyságrendje eltér az adminisztratív metaadatokéhoz képest, hiszen amíg az előzőek arányosak az – audiovizuális – dokumentumok számával (ami a százezres, milliós nagyságrendben mozog egy-egy szolgáltató esetében, de ha az összes audiovizuális dokumentum mennyiségét keressük, akkor is a milliárdos számosság lehet a felső becslés, az is csak akkor, ha a házi gyártású homevideókat is figyelembe vesszük). Ezzel szemben az audiovizuális dokumentumok használata, fogyasztása során keletkező forgalmi adatok a felhasználók, nézők, még pontosabban a látogatások, az olvasások, nézések, értékelések számosságával arányosak, amelyek nagyságrendje a csillagos ég – egyrészt azért, mert a fogyasztható dokumentumokhoz képest sokkal több a fogyasztó, másrészt azért, mert a felhasználók ugyanazt – vagy csak nagyon lassan bővülő – kínálatot időben folyamatosan újrafogyasztják. A forgalmi

adatok nagyságrendjét a dokumentumok és a felhasználók közt kapcsolatot teremtő felhasználások határozzák meg, és ebben a kapcsolatrendszerben egy kombinatorikai robbanás tapasztalható.

A különböző metaadatoknak más a funkciója, más a forrása, más technológiát kell a kezelésükre fordítani. A strukturális metaadat funkciója a tartalom konzisztens, strukturált kezelésének, szolgáltatásának biztosítása, a tartalmi és formai metaadatok elsődleges célja és értelme az, hogy a felhasználókat segítsék a megfelelő tartalmak megtalálásában (a felhasználók közé érve itt magukat a tartalomtulajdonosokat, tartalomszolgáltatókat is, akiknek szükségük van egy saját "leltárkönyvre", nyilvántartási rendszerre). A strukturális, formai és tartalmi metaadatokat, valamint az adminisztratív metaadatok egy részét a "tulajdonos", a szolgáltató rendeli a dokumentumokhoz – a fogyasztáshoz képest előzetesen. A forgalmi metaadatok ezzel szemben mindig a "fogyasztások után", a felhasználók tevékenysége nyomán keletkeznek. Abban az értelemben nincs céljuk, ahogy a formai és tartalmi metaadatoknak van, tehát nem mondhatjuk, hogy a tartalmak megtalálhatósága érdekében jönnének létre, mégis fel lehet őket arra használni, hogy a feldolgozásuk révén megfelelő felhasználói profilokat lehessen építeni és ezek alapján támogatni a felhasználók mozgását, eligazodását a digitális térben. A kollaboratív szűrési technológiák, az adatbányászat a fogyasztói ízléscsoportok beazonosítására, szegmentálására, ezek fogyasztói magatartásának feltérképezésére, és mindezek alapján a beazonosított felhasználói csoportok testére, ízlésére szabott fogyasztási kínálat kialakítására alkalmasak.

A szemantikus navigáció olyan térbeli navigáció, amikor jelentést hordozó írásjeleket helyezünk a képernyőre abból a célból, hogy azok segítsék a felhasználók eligazodását valamilyen információs tartományban, amely egyaránt lehet metaadat- vagy tartalominformációs tartomány. Ezeket a vizuális nyelvi jeleket a képernyő által behatárolt térben egymáshoz viszonyítva helyezzük el, és – a nyelvi írásjelek önmagában vett jelentései mellett – az ilyen térbeli viszonyoknak is lehet önálló jelentést, navigációs potenciált tulajdonítani.

Kiindulásként érdemes röviden összefoglalnunk mindazt, amit a navigáció problémájának általános jellemzésekor leírtunk. Ha az *információt* tárolni akarjuk, akkor azt (vagyis a *tartalmat*) valamilyen hordozóra rögzítve *dokumentumot* (könyvet, képet, videót, hanglemezt stb.) hozunk létre. A tárolási tevékenység egyik kiemelt célja a tartalom későbbi befogadásának biztosítása. Ezt a célt a dokumentumok *archívumokba* rendezésével érhetjük el. A tartalom eléréséhez meg kell találnunk az archívumon belül a keresett dokumentumot, és e célból kiegészítő információt, *metaadatot* kell a dokumentumokhoz rendelnünk, melynek elsődleges célja az információelérés támogatása, a dokumentum visszakereshetőségének biztosítása az archívumon belül.

A tárolás és információelérés gyakorlati problémáira a legkiérleltebb választ az évszázados könyvtári hagyományban találhatjuk meg. Nem véletlen, hogy a web létrejöttétől kezdve sokak gondolták úgy, hogy érdemes hasznosítani ezen értékes tudást és tapasztalatot. Az ilyen kezdeményezések azonban kevés sikerrel jártak, és erre magyarázatot kellene találnunk.

A válaszuk kibontásához szükségünk lesz a metaadattípusok fogalmaira. Korábban elkülönítettük egymástól a *formai*, a *tartalmi* és az *adminisztratív-használati* metaadatokat. A hagyományos világban a navigáció támogatására e három adattípusból csak az első kettőt, vagyis a formai és a tartalmi metainformációt használták. Az adminisztratív, használati, forgalmi adatok gyűjtésére, kezelésére nem álltak rendelkezésre a megfelelő technológiák. Látni fogjuk, hogy a számítógépes, hálózati kommunikáció pont ezen a téren hozott forradalmi változást. A hagyományos világban azonban ezzel az adattípussal nem foglalkoztak érdemben (legalábbis a navigációhoz kapcsolódva).

A „megmaradó” két metaadattípusba tartozó formai és tartalmi leíró adatokat más és más módokon kezelték – jelentős mértékben azért, mert másfajta adatmodellt kellett alkalmazni a háttérben. A dokumentumok *formai metaadatait* csak azáltal lehet megfelelően menedzselni (akár analóg, akár digitális formában!), ha pontosan és egyértelműen minősítjük az egyes metaadattípusokat, metaadatelemeket. Ez egyenes arányban áll a *formai metaadatrendszer* strukturáltsági fokával (tehát bonyolultságával). Ahány formai adatelemet akarunk kezelni, annyiféle entitást és majdnem ugyanannyi relációtípust kell definiálnunk, hogy aztán azok konkrét értékeit a dokumentumokhoz rendelhessük. Persze a formai elemek bonyolultsága nem feltétlenül jelent meg minden felhasználó előtt, mivel a formai adatelemek közül csak elég kevés bizonyult olyannak, amelyet navigációs szempontból valóban sokan használtak. A művek alkotói (írója, rendezője, zeneszerzője, operatőre, előadója stb.), a művek címei, esetleg az alkotás, a publikálás ideje – talán ezek voltak a leggyakrabban használt formai

metaadatok. Ettől persze a teljes formai adatrendszer bonyolultsága megmaradt (amelyet a különböző archívumi szakmák képviselői kezeltek saját céhszerű praxisukon belül).

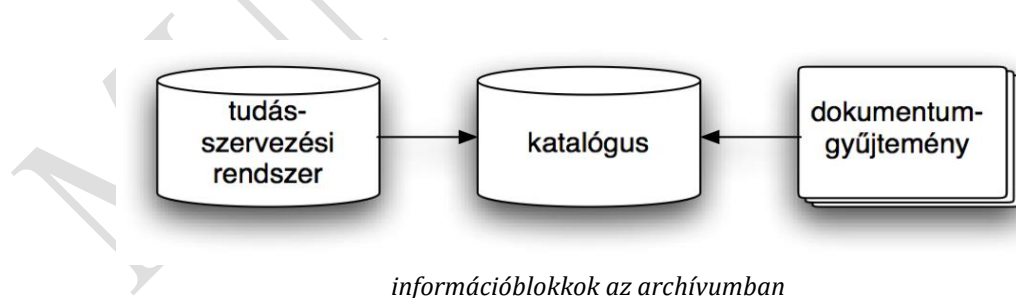
A tartalmi leíró tevékenységgel más a helyzet. Többféle lehetőség van a *tartalmi metaadatok* könyvekhez rendelésére. A legelterjedtebb megoldás az, hogy a dokumentumok tartalmát valamilyen előre kidolgozott készletből válogatott kifejezésekkel, kulcsszavakkal, tárgyszavakkal jellemezzük. Amikor ezt tesszük, akkor a szóban forgó terminuskészletet már érdemes önálló elemként elkülönítenünk az archívumi modellünkben, amelynek megnevezésére használhatjuk a nemzetközileg elfogadott kifejezést: *tudásszervezési rendszer* (Knowledge Organization System, KOS).

A tudásszervezési rendszerek a tartalmi feltáró munka során a dokumentumokhoz rendelhető tárgyszavakból és az ezek között tételezett relációkból állnak. Érdemes feltenni a kérdést, hogy miért van szükségünk a relációk értelmezésére. Legalább két fontos funkcióra már előzetesen rámutathatunk, noha ezeket csak a későbbiekben fejtjük ki bővebben.

Egyrészt a tárgyszó-hozzárendelés célja és értelme a dokumentumok minél pontosabb tartalmi leírása. A természetes nyelv rugalmasságának „ára” a szavaink, kifejezéseink többértelműsége. A pontos tartalmi leíráshoz arra van szükségünk, hogy a nyelv egyébként sokértelmű szavait, kifejezéseit egyértelmű jelentés mellett tudjuk a dokumentumokhoz rendelni. Ez az egyértelműsítés azáltal teremthető meg (legalábbis a hagyományos archívumi gyakorlatban), hogy a tárgyszavaknak megmutatjuk a fogalmi környezetét (vagyis azt, hogy más fogalmakkal milyen kapcsolatban állnak). Ezért a tárgyszavazást végző személyeknek a tárgyszókészletet annak teljes struktúrájával együtt kell látniuk (a tárgyszavak közötti relációk ezen felül abban is segítséget adhatnak, hogy támogatják a rendszeren belüli navigációt a tárgyszókeresési munka során). Másrészt pedig a relációk másik fontos „funkciója” az, hogy a segítségükkel alternatív tárgyszavakat lehet megtalálni. Ilyen esetekben a tartalmi leírás pontosításához lehet hozzájárulni.²⁹

Addig jutottunk tehát, hogy megállapítsuk: a metaadat-hozzárendelési munkát – a szemantikai egyértelműsítés érdekében – kétféle módon is támogatta a könyvtári gyakorlat. Egyrészt a formai adatelemekre egy sokdimenziós (sok relációból álló), nem túl összefüggő és minden dimenzióban nagyon lapos, de pontosan rögzített struktúrát teremtett, másrészt a tartalmi leírás egyértelműsítésére a tudásszervezési rendszerek elemeit vette használatba. A továbbiakban ez utóbbi mozzanatra koncentrálnunk.

A tartalmi leíró tevékenység során a dokumentumokat a tudásszervezési rendszer elemeivel jellemezzük úgy, hogy összekötjük az éppen elemezett dokumentumot a KOS-rendszer kiválasztott elemeivel. Ha az összekapcsolás tényét rögzítő adatot a többi információtól elkülönítve kezeljük, akkor létrehozuk (fenntartjuk) a katalógus „intézményét”. Azt mondhatjuk tehát, hogy a tartalmi leíró tevékenység a katalógus céduláinak, rekordjainak írását jelenti. Az archívum általános modelljében három nagyobb információs blokk van, amelyeket az alábbi ábrával szemléltethetünk:



3.2.1.4 Gyűjtemény, információs tartomány

Az információt dokumentumokban rögzíthetjük és tárolhatjuk. A dokumentumokat dokumentumgyűjteményekbe, archívumokba gyűjthetjük, rendezhetjük, és tárolhatjuk (a bennük található információval együtt). Ebben az értelemben archívumnak kell minősítenünk minden olyan gyűjteményt, amelyhez kapcsolódóan emberek dokumentumokat, illetve a dokumentumokban tárolt információkat szeretnének megkeresni, megtalálni. tehát ide kell sorolnunk a könyvtárak, a zenetárak,

²⁹ A tudásszervezési rendszereknek létezik egy harmadik funkciója is: a logikai következtetések végrehajtásának támogatása, de ezt a funkciót jelen tanulmányban nem érdemes tárgyalnunk.

a képtárak, a fotótárak, a filmtárak, a levéltárak, vagyis minden „hagyományos archívum” mellett a könyveket, videókat, CD-eket, DVD-eket, filmeket, fotókat, képeket stb. árusító boltokat, ezek raktárait, de a családi háztartások könyvszekrényeit, CD- vagy DVD-polcait is. Az archívum tehát egy kereshető dokumentumgyűjtemény. Kimondhatjuk még azt a tételt is, hogy az ilyen keresés elsődleges célja nem a dokumentum megtalálása, hanem – és lényegét tekintve mindig – egy dokumentumban található valamilyen információt keresünk. Ezt az információt nevezhetjük *tartalomnak* (tartalominformációnak). A tartalom annyiféle lehet, ahányféle információtypust elkülöníthetünk egymástól. Az egyik dimenzió mentén a tartalom lehet auditív, vizuális vagy audiovizuális, a másik felosztás mentén lehet nyelvi vagy nem-nyelvi, a nyelvi információk típusa pedig lehet beszéd, írás, hipertext vagy adatbázis). Az archívumban való keresés alapvető célja a kívánt tartalom elérése (majd a befogadása). A tartalomban (azaz a tartalmat tároló dokumentumban) is lehet keresni, természetesen mindig a tartalom típusának megfelelő módon. Ezt nevezhetjük elsődleges keresésnek. Amikor egy könyvet gyorsan átlapozunk egy fejezetcím után kutatva, vagy amikor egy atlaszban egy bizonyos ország térképét keressük, vagy amikor egymás után behallgatunk egy CD-n a számokba, hogy megtaláljunk egy korábban már hallott zeneszámot, akkor az információtypushoz igazodó, keresési (mintaillesztési) képességünket hasznosítjuk: karaktersorozat, képmintázat, hangmintázat, ritmusképletet keresünk a dokumentum egészén belül. Ez a képesség (tehát a szöveg-, kép, valamint hangmintázat beazonosításán alapuló keresés) azonban már nem elég, ha nagyon nagy számú dokumentumot tárolunk a gyűjteményünkben.³⁰ Ahhoz, hogy ebben az archívumi keresésben valóban sikeresek legyünk, szükségünk van egy kiegészítő információra, amelynek segítségével gyorsan és könnyen megtalálhatjuk a kívánt tartalmat a gyűjteményen belül. Ezt az információtypust nevezzük *metainformációnak*. Ennek hiányában sok dokumentum megtalálhatatlanná válna, ha bekerülnének egy gyűjteménybe. Metainformációk nélkül egy képet, zeneszámot, filmrészletet, de akár egy szöveget, könyvet, vagy cikket is csak úgy találhatnánk meg, ha bemennék a gyűjteményt tároló helyiségbe (a „raktárterembe”), és egyenként kézbe vennék, megnéznék, meghallgatnák a dokumentumokat. De ha hozzárendelnék a dokumentumokhoz az őket jellemző adatokat (metaadatokat), akkor új lehetőség nyílna a keresésre, amely a nagyon durva lineáris hozzáférés helyett másfajta elérést is lehetővé tene. Ezekből a metaadatokból ugyanis létrehozhatunk egy másik, önálló információs teret, amelybe csak akkor lépünk be, ha valamit keresni akarunk. Az itt folytatott keresés tehát másodlagos, amelynek célja valamelyik dokumentum tartalmának megtalálása – a dokumentumok metaadatterében folytatott keresésen keresztül. A metaadatter lényegében – ahogy a neve is mutatja – írás-, pontosabban adatbázis-alapú. A *metaadat* írásjelek segítségével, strukturált módon rögzített nyelvi információ. Elképzelhető nem adatbázis-alapú metainformáció is, de ezek jelentősége, elterjedtsége nem tekinthető komolynak.³¹

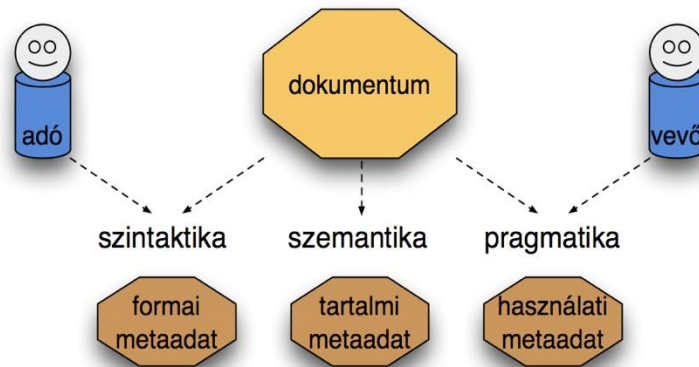
A metaadatterben zajló keresés következő fontos kérdése az, hogy milyen típusú metaadatot kezelhetünk. A hagyományos archívumok praxisában három típust különítettek el egymástól: a *formai*, a *tartalmi* és az *adminisztratív-használati metaadatokat*. A *formai metaadatok* körébe tartozik minden olyan információ, amely a dokumentumot annak tartalmától függetlenül jellemzi. Ide sorolhatjuk a dokumentum címét, alkotóját, kiadóját, a megjelenítés helyét, idejét, a dokumentum fizikai jellemzőit, nagyságát, hosszát, hordozóját stb. Ugyanakkor a *tartalmi metaadatokat* csak a tartalom ismeretében lehet a dokumentumokhoz rendelni. Célja az, hogy ezen – a dokumentum tartalmánál jóval kisebb mennyiségű – leíró információ ismeretében is képet lehessen alkotni, mi lehet a teljes dokumentum tartalma. Az *adminisztratív metaadatok* a dokumentum használatával kapcsolatosak. Ilyen jellegűek például azok az információk, amelyek az írják le, hogy az archívumon belül hol találhatóak meg a dokumentumok, de az is ide tartozik, hogy hányszor vették ki a könyvtárból a könyvet, vagy hányszor hallgatták meg a lemezt a zenetárban stb. Ez utóbbi információs tartomány az elektronikus, de főleg a digitális kommunikáció során nyeri el igazi jelentőségét.

Csak az érdekesség kedvéért jegyezzük meg, hogy bár nem teljeskörű és nem is teljesen megalapozott az összefüggés, de azért felismerhető egy, a háromfajta metaadat és a shannoni kommunikációs modell

³⁰ A gépek majd segíteni tudnak ebben is, sőt, számukra nem jelent igazi gondot a feldolgozandó anyag mennyisége sem. Erről még lesz szó a későbbiekben.

³¹ A képarchívumok dokumentumaihoz például „bélyegképeket” is rendelhetünk, amelyek halmazában a képminta szerinti keresés segítheti az „eredeti” képek megtalálását, de az effajta lehetőségek – legalábbis a számítógépes keresés megjelenéséig – nem voltak a gyakorlatban kihasználhatók. Maradt (és vélhetőleg marad is) a metaadat-keresés.

három szintje között fennálló laza párhuzam. Egyrészt a formai metaadat a shannoni modellben foglalt adó és üzenet (tartalom) közötti, a tartalmi metaadat az üzenet (tartalom) és a valóság közötti, az adminisztratív-használati metaadat pedig az üzenet (tartalom) és a vevő közötti kapcsolatot jellemzi, másrészt pedig a formai metaadat megfeleltethető a szintaktikai, a tartalmi metaadat a szemantikai, míg az adminisztratív-használati metaadat a pragmatikai szintnek. Ezt az összefüggést mutatja az alábbi ábra:



a metaadatok és a shannoni modell összefüggései

Összefoglalásképpen: az archivumi keresés többféle információs térben valósulhat meg, és mindig nagyon pontosan látnunk kell, hogy milyen típusú információs térről, illetve keresésről beszélünk éppen. A dokumentumtérben a dokumentumok tartalmaiban, a metaadattérben a dokumentumok metaadataiban kereshetünk. Utóbbi keresés háromfajta lehet a háromféle metaadattípushoz igazodva. A keresési tartományok (és kereséstípusok) tehát az alábbiak lehetnek:

- tartalom
 - állókép
 - mozgókép
 - hang
 - beszéd
 - írás, hipertext
 - adatbázis
 - összetett dokumentum (multimédia)
- metaadat
 - formai metaadat
 - tartalmi metaadat
 - adminisztratív-használati metaadat

Tehát a különböző információs tartományokban lehetővé válik az adott információtípusra jellemző mintázat szerint zajló, mintaillesztéses keresés. Ezt azért nagyon fontos hangsúlyozni, mert amíg a hagyományos világban gyakorlatilag csak a metaadat-keresésre volt lehetőség,³² addig a hálózati kommunikáció – ennek továbbfejlesztése mellett – megteremtette a szöveges tartalomban (pontosabban az írásban) való keresés lehetőségét a szabad szavas keresés segítségével, majd vélhetőleg fokozatosan lehetővé fogja tenni a hang- és képmintázat alapján történő keresést is az audiovizuális archívumok tartalmaiban.

³² Az ember mindenféle tartalomtípus szerint tud mintaillesztéses keresést végrehajtani, de csak kis tételben, archivumi méretekben már nem. A számítógép e tekintetben új minőséget jelent.

A befektetett munkának köszönhetően idővel egyre jobbak lesznek azok a keresési technológiák, amelyek révén a nyelvfüggetlen audiovizuális térben is lehet keresni. A beszédfelismerés, a hangfelismerés, a dallamfelismerés technológiái előtt nagy jövő áll. A beszédfelismerő technológiák azt teszik lehetővé, hogy auditív nyelvi (beszédszöveg) dokumentumokban is ugyanúgy tudjunk keresni, mint a vizuális nyelvi (írasszöveg) dokumentumokban úgy, hogy a beszédet konvertáljuk digitális – írott – szöveggé, és az így kapott szövegtörzsekben már a hagyományos keresőtechnológiákat vehetjük igénybe.

A „tiszta” vizuális (nyelvfüggetlen) dokumentumokban való keresés is régóta megjelent már, de az eddig eltelt időben még nem igazolódott be, hogy az ilyen típusú keresés valóban annyira hasznos lenne legalább is a weben. Az első képfelismerő technológiák messze megelőzték a webes szolgáltatásokat, gondoljunk csak az ujjlenyomat- vagy arcfelismerő rendszereket. A weben a Google képkereső már elég régóta lehetővé teszi a megadott képi mintázat szerint keresést, de egyelőre nem látszik ezek igazi haszna (és egyelőre a képességeik sem annyira meggyőzők).

Akárhogy is van, a névtér projekt számára nem érdekesek a nyelvi tartományon kívüli keresés lehetőségei, ezért ezekkel nem foglalkozunk a továbbiakban, mert kivezetnek a projekt skópjából.

3.2.1.5 Információelérés és metaadat

Minden médiaszolgáltató rendelkezik valamilyen dokumentumgyűjteménnyel, tartalomkínálattal, amit valamilyen technológia révén felkínál az érdeklődők számára. A lineáris broadcast szolgáltató műsorfolyamba fűzi a tartalmait, az interaktív broadband szolgáltató vagy egy keresőszolgáltatáson, vagy egy szűrőszolgáltatáson keresztül teszi megtalálhatóvá, elérhetővé a tartalmait. A keresőszolgáltatás azon alapul, hogy a felhasználó megadja azokat az információkat, amelyek alapján a keresőmotor visszaadja a feltételeknek megfelelő találatokat. Ilyenkor a felhasználó – elvileg – a teljes dokumentumgyűjteményben keres. Példa lehet erre a YouTube keresője, amikor a felhasználó által megadott keresőfeltételhez illeszkedő találatokat ad vissza a teljes videoállományból. A szűrőszolgáltatás ezzel szemben azon az elven működik, hogy a szolgáltató előre megadja azokat a feltételeket, amelyek mentén ő maga megszűri saját tartalomkínálatát, és a felhasználó már a teljeshez képest egy erősen csökkentett kínálatot kap induláskor. Egy klasszikus portál kínálata lehet a példa erre. A szolgáltató által előre megadott szűrési feltételeken természetesen lehet módosítani, változtatni akkor és azzal, ha a szűrési feltételek közé beveszik a korábbi fogyasztási, forgalmi adatok alapján készített profilozási adatokat. Ez persze igaz a klasszikus keresőszolgáltatásra is, hiszen oda is be lehet kötni – rárakódó szolgáltatásként – valamilyen kollaboratív filtering szolgáltatást, amire az Amazon könyvtárház „customers who bought this item also bought” szolgáltatását hozhatjuk fel példaként.

Az előzetes elvárásokhoz képest kicsit meglepő, hogy az interaktív broadband szolgáltatások világában is felbukkan a program jellegű, tehát rögzített lineáris rendben, egymás után befogadható tartalomkínálat. A YouTube “autoplay”-opciója (a videók egymás utáni, automatikus lejátszása) vagy a Facebook személyes üzenőfalára kitett tartalmak – részben idő szerinti, részben profilozási adatok alapján történő – lineáris elrendezése mutat mintát erre.

A felhasználók digitális aktivitásáról begyűjtött adatok (digitális lábnyomok) adják a profilképzés alapját. Ezt igazából csak az interaktív szolgáltatásokon keresztül lehet begyűjteni. A profilképzés jó a felhasználónak, mert jobb ajánlatokat kap a szolgáltatótól, de jó a szolgáltatónak is, hiszen a korábbiakhoz képest jóval mélyebb ismeretet biztosít a felhasználói igényekről, viselkedésekről. Fel kell azonban ismerni, hogy minden profilképzési technológiát használó szolgáltatás tekinthető egyben egyfajta rávezető szolgáltatásnak is, amelynek célja a felhasználók további fogyasztásának ösztönzése. A profilozási technológiák, az ajánlórendszerek lényegében a forgalmi metaadatokból építkeznek. Ezek a felhasználók aktivitásából erednek, és nagyon sok esetben valamilyen felhasználó oldali értékelő megnyilvánulásoknak tekinthetők (tetszés, szavazás, vásárlás, elolvasás stb.). Ezen értékelési szempontok beemelése biztosítja azt sok esetben, hogy a társas navigáció technológiáihoz szükséges adatok rendelkezésre álljanak.

Az audiovizuális tartalmakban való keresés új útját jelenti a beszédtechnológiák révén transzkriptált szövegekben való szabadszavas keresés. Az ilyen szolgáltatás mindazon előnyöket biztosíthatja, amelyekkel a hagyományos szabadszavas keresők rendelkeznek. Előzetes feldolgozás révén indexelni lehet a tartalmak teljes szövegállományát, és ez alapján lehet keresni a teljes videoszöveg-környezetben. Ráadásul ez a technológia arra is alkalmassá tehető, hogy a vizuális képfolyamra a hallható beszéd

alapján feliratokat lehessen kitenni. Ha pedig vállalható minőségi szintre jut a gépi fordítás, akkor a transzkriptált szöveg alapján más nyelveken is lehet feliratokat készíteni. Ma még nem tudni, mikor lesz valóság ebből a lehetőségből, de azt tudni lehet, hogy ennek a feladatnak a megoldását csak nagyon kevés szolgáltatótól remélhetjük, és az ilyen szolgáltatás a tartalominfrastruktúra egyik fontos komponense lesz a jövőben.

A keresőszolgáltatás másik fontos rétegét, minőségét jelentik azok a komponensek, amelyek az audiovizuális tartalmak formai és tartalmi metaadatai alapján teszik kereshetővé a dokumentumgyűjteményt. Ehhez arra van szükség, hogy ezek a metaadatok egységes rendszerben rendelkezésre álljanak. Ilyen audiovizuális keresőszolgáltatások vannak már. Magyarországon ilyen a NAVA, a filmes világban ilyen az IMDB (vagy idehaza a nyilvánosan még nem elérhető MozgóKépTár). Ezek a szolgáltatások alapvetően a mozgóképek rövid tartalmi leírásaiban, illetve a filmográfiai adatokban keresnek. Részlegesen ide sorolhatjuk még a programújság szolgáltatókat (mint a port.hu), amelyek a szolgáltatásuk melléktermékeként nyújtanak filmes keresőszolgáltatást (ha már a tévé- és moziprogramok feldolgozása során hozzájutnak a filmes metaadatokhoz). Meg kell itt említeni, hogy talán ez a szolgáltatási réteg, ez a keresőszolgáltatási lehetőség kívánna meg a leginkább azt, hogy a tartalombirtokosok, tartalomszolgáltatók vagy más, rávezető szolgáltatásokat végző szereplők együttműködjenek egymással, hiszen itt közös érdek lehet a metaadatok egységes, közös kezelése, karbantartása.

A metaadatok kollaboratív kezeléséhez hozzátartozik közös névterek fenntartása is, és ezek között különösen fontos lehet az audiovizuális dokumentumok (elsősorban a műsorszámok és a filmek) egységes, központi azonosítása (nemzeti névterek működtetése szoros kooperációban a nemzetközi névterekkel).

A rávezető szolgáltatások következő rétegében lehet elképzelni a forgalmi, fogyasztási, értékelési, tehát a felhasználó-oldali adatok elemzésére támaszkodó profilozási és az erre építhető ajánlási, szűrési szolgáltatások rendszerét. Újra hangsúlyozni kell itt, hogy az ilyen profilképzési szolgáltatások annál sikeresebbek lehetnek, minél több adatot tudnak összegyűjteni és elemezni, és e "kényszer" miatt arra kell számolni, hogy ezt a lehetőséget csak kevés infrastruktúra-szolgáltató fogja tudni felépíteni és a piac számára kiajánlani.

Az egységes, közös infrastruktúrahasználat abból a szempontból is előnyös lehet a tartalomszolgáltatók számára, hogy az infrastruktúraszolgáltató minden használati, forgalmi adatot összegyűjthet, és ezek alapján olyan összetett, integrált statisztikai kimutatásokat, forgalmi riportokat, felhasználói profilokat, kampánytervezési segédleteket stb. készíthet, amire egyébként nem lenne módja a tartalomszolgáltatónak.

Az audiovizuális tartalomszolgáltatások bármelyik, korábban bemutatott keresőszolgáltatási lehetőséget hasznosíthatják. Látni kell azonban, hogy ezekhez a szolgáltatásokhoz is komoly technológiai tudás szükséges. Infrastruktúra kell, de más infrastruktúra kell a szabadszavas tartalomkereséshez, a szabadszavas keresés kiterjesztéseként tekinthető transzkriptált kereséshez, a metaadatokban való kereséshez és a forgalmi adatok elemzésén alapuló profilozáshoz. Ezek mind komoly technológiai tudást, infrastrukturális alapokat igényelnek, és ha a különböző keresési logikák és képességek közti szinergiákat is ki akarjuk használni, akkor ezeket közös infrastruktúrán keresztül lehet igazán megvalósítani.

* * *

Az információelérés, a navigáció problémájának tárgyalását az információ, illetve az információs tartomány fogalmainak bemutatásával, tipizálásával kell kezdenünk (az itt röviden kifejtett gondolatokat bővebben tárgyalja [Syi 2007]).

3.3 Információelérés

Jelen kontextusban olyan információról beszélünk, amely valamilyen hordozón rögzítve van, ezért dokumentumnak minősíthető. Ebből a feltételből következően áll elő az a helyzet, hogy a dokumentumokban rögzített információ halmozódni kezd, és az így egyre bővülő nagyságú információs

terekben egyre komolyabb feladattá válik, hogy megtaláljuk azt az információt, azt a dokumentumot, amely a felhasználók számára relevánsnak bizonyulhat.

Ez a helyzet úgy jellemezhető, hogy a – valamilyen szempont szerint felhalmozott – dokumentumokból **információs tartományokat** képzünk, és a **felhasználókat** megpróbáljuk segíteni abban, hogy elérjék azokat a **releváns dokumentumokat**, amelyek megfelelnek az **információs igényeinek**.

Jelen kontextusban a nyelvi és vizuális információt tartalmazó dokumentumokra szűkítjük a vizsgálódás fókuszát, de bizonyos esetekben megengedjük, hogy kitekintsünk más információ típusokhoz köthető dokumentumtípusok tartományaira is.

3.3.1.1 Információs igény

Az eddigiekben használtuk az információelérés és a keresés fogalmait, de nem igazán tisztáztuk, mit értünk e fogalmak alatt, és mit gondolunk az egymáshoz való viszonyukról. A következőkben ezért bemutatjuk, hogyan lehet és érdemes az információelérés fogalmához kapcsolódó további alapfogalmakat definiálni.

Pontosabb lenne az információelérés helyett az **információhasználat** általánosabb fogalmát elemezni, de ezt az általánosítást mégsem tesszük meg. Az általánosítás értelme az lenne, hogy a modellünkbe így megragadhatnánk a kétirányú kommunikáció jelenségét, vagyis azokat az eseteket is le tudnánk írni, amikor a felhasználók lényegi információt küldenek el valamely címzett számára, azaz nem pusztán információbefogadóként, hanem információkibocsátóként is működnek. Ilyenkor teljessé, azaz kétirányúvá válik a kommunikációs folyamat, amelyet a **tranzakció** fogalmával ragadhatnánk meg. De ezt a kérdéskört – bármennyire fontos – most nem elemezzük, mivel a navigáció jelenségére fókuszálunk. Természetesen a későbbiekben, ahol szükséges, utalni fogunk a felhasználói aktivitás jelenségére, valamint annak fontosságára, de ehhez nincs szükségünk arra, hogy ezt a fogalmat is felvegyük a navigáció jelenségét leíró fogalmi modellünkbe. A teljes kép kialakítása érdekében természetesen fel kell tüntetnünk akkor, amikor azt vizsgáljuk, milyen információhasználati lehetőségeink, igényeink vannak a hálózati kommunikáció során.

E kérdés megválaszolásának érdekében sokan azt vizsgálják, hogy milyen információs igények kielégítésére irányul az aktuális számítógép-használat. Az információs igények elemzésekor többen (például [Broder 2002], [Hearst et al. 2002]) az alábbi használati típusokat különítették el egymástól:

- **navigációs**
az aktuális eszközhasználat közvetlen célja valamely dokumentum, szolgáltatás, weboldal, gyűjtemény elérése, megtalálása
- **informálódási**
az aktuális eszközhasználat közvetlen célja valamely információ elérése, megtalálása az elérhető weboldalakon, dokumentumokon, szolgáltatásokon belül
- **tranzakciós**
az aktuális eszközhasználat közvetlen célja valamely weben keresztül végezhető tevékenység végrehajtása (vásárlás, szerepjáték, virtuális valóságban való tevékenység, dokumentum szerkesztése, nyomtatása, levélírás, szavazás, egy szolgáltatás által biztosított eszközkészlet valamely elemének használata stb.)

Sajnos az **információs igény** általánosabb fogalma és annak egyik típusa, az *informálódási igény* két terminusa túl közel van egymáshoz, amely könnyen félreértésekhez, pongyolasághoz vezethet. Ezen csak úgy tudunk segíteni, hogy a későbbiekben igyekszünk egyértelműen jelezni, melyik fogalomról van éppen szó.

Ezen klasszifikáció másik nagy problémája az, hogy gyakran előállhatnak olyan esetek, amikor nem lehet pontosan elhatárolni az informálódási és navigációs igényeket egymástól. Ennek az az alapvető oka, hogy a dokumentum fogalma lényegéből fakadóan homályos, azaz nem tudjuk egzakt módon meghatározni, mit lehet dokumentumnak, mit lehet a dokumentum egy részének, illetve mit lehet dokumentumok halmazának tekinteni. Ha egy kétalakos képet kettévágunk, akkor két – teljes értékű – képet (dokumentumot) kapunk, ha egy hosszabb novellát önálló könyvként jelentetünk meg, akkor dokumentum, ha egy antológia részeként, akkor dokumentumrész lesz. Ebből a bizonytalanságból fakad, hogy ugyanaz az információelérési törekvés hol navigációs igényként (valamely dokumentum

megtalálása egy gyűjteményen belül), hol pedig informálódási igényként (valamely konkrét információ megtalálása egy dokumentumon, szolgáltatáson belül) jelenik meg. Mivel ez a többértelműség a fogalmi szint meghatározatlanságából fakad (amelye lényegéből adódóan feloldhatatlan), ezért ezzel a pongyolással együtt kell élnünk.

Mivel fentebb jeleztük, hogy a **tranzakciós igényeket** (illetve műveleteket) itt és most érdemben nem tárgyaljuk, marad az információs és navigációs igény (illetve művelet) fogalma. Ez a két fogalom már korábban is szerepelt, amikor a tartalomtartományban, illetve a metaadattérben való keresés különbségéről írtunk. Az **informálódási igényünket** akkor elégíthetjük ki, ha már valamely dokumentum tartalmában keresünk, vagyis amikor már a tartalmat fogadjuk be, míg a *navigációs igény* akkor (és addig) jelentkezik, amikor (és ameddig) meg akarjuk találni a számunkra fontos, érdekes dokumentumokat a gyűjteményen belül. Azt írtuk, hogy mindig a tartalom megtalálása és befogadása a cél, és a tartalomban való keresés elsődleges szerepű a metaadatok közti kereséshez képest. A fenti tipizálással tehát nem léptünk még nagyot előre, csak új terminusokat találtunk a modellbe felvett fogalmainkra. Mivel a tranzakció fogalmával most nem foglalkozunk, ezért az információhasználat fogalma helyett elegendő az információelérés fogalmát elemeznünk – és ezennel érkeztünk el arra a pontra, mikor már a *keresés* fogalmát pontosabban meg kell ragadnunk.

3.3.1.2 Relevanciakezelés

Az információs igények kielégítése akkor lehet sikeres, ha az igény kielégítése során a felhasználók úgy érzik, hogy a folyamat végén releváns információt kapnak, amikor valamilyen információs szolgáltatást igénybe vesznek. Ezért van szükségünk arra, hogy röviden kitérjünk arra a kérdésre, mit jelent a relevancia fogalma, és milyen megoldásokkal lehet biztosítani a felhasználó számára releváns dokumentum, releváns információ megtalálását.

A relevancia fogalma nyilvánvalóan értékvonatkozású, mindig kifejez valamilyen értékelő viszonyt. Az információelérés elméletében a kezdetektől fogva használják, és mindig valamilyen keresőszolgáltatás minőségének mértékét fejezik ki vele. Van azonban más értelmezése is a fogalomnak, amikor nem mértékként hivatkozunk rá, hanem egy dokumentum, egy információ a felhasználó számára vett fontosságát fejezzük ki vele. Ekkor érték kategóriaként használjuk a fogalmat. A következő fejezetekben ezt a két értelmezést mutatjuk be.

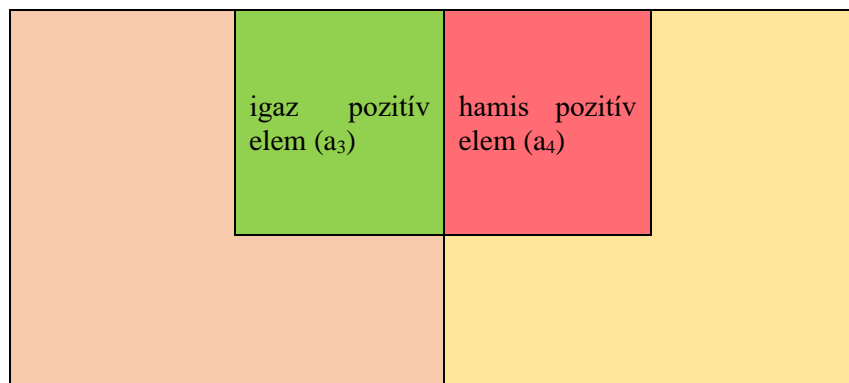
3.3.1.2.1 Relevancia mint mértékkategória

Az információelérési technológiák világában régóta használt fogalompár a szolgáltatások **relevanciájára** vonatkozóan a **pontosság** (precision) és **felidézés** (recall) kategóriája. Azt feltételezve, hogy tudjuk az információs tartomány minden eleméről, hogy releváns-e vagy sem az adott információs igény esetén, illetve tudjuk azt is, hogy a keresés eredménye, a találati halmaz mekkora hányadot képvisel a két – releváns, illetve nem-releváns – részhalmazban, akkor a kétfajta relevanciamérési fogalmat értelmét az alábbi ábrán szemléltethetjük. Az ábrán feltüntetett kifejezések értelme:

- pozitív elem az, amit a rendszer megtalál (vagyis a találat), ami lehet
 - igaz pozitív elem (mert releváns találat) vagy
 - hamis pozitív elem (mert nem-releváns találat).
- negatív elem az, amit a rendszer nem talál meg (nem-találat), ami lehet
 - hamis negatív elem (mert releváns elem nem-találata) vagy
 - igaz negatív elem (mert irreleváns elem nem-találata)

A fogalmak egymáshoz való viszonya tehát a következő:

| releváns elem | irreleváns elem |
|------------------------------|-----------------------------|
| hamis negatív elem (a_1) | igaz negatív elem (a_2) |



A fenti értelmezések alapján a két mértéket már könnyen definiálhatjuk (a színekkel is elkülönített területeket a_i szimbólumokkal jelölve):

- pontosság = $a_3/(a_3+a_4)$
- felidőzés = $a_3/(a_3+a_1)$

A pontosság azt fejezi ki, hogy a találati halmazon belül milyen aránya van a megfelelő találatoknak a találatokhoz egészéhez képest. Minél magasabb ez az érték egy adott szolgáltatásra vonatkozóan, annál biztosabbak lehetünk abban, hogy egy adott találat releváns lesz. Ebből azonban nem tudhatjuk azt, hogy a találati halmazunk milyen viszonyban áll a teljes információs tartományhoz képest. Lehet, hogy a kapott eredményünk pontos, tehát nincs vagy alig van benne irreleváns találat, de ettől még könnyen elképzelhető az is, hogy a találati halmaz a teljes tartomány kicsiny részét fedi le, tehát lehetnek még releváns találatok, amiket nem kaptunk meg a szolgáltatástól. Ennek a jelzésére alkalmas a felidőzés kategóriája.

A relevanciamértékek elméletében természetesen további fogalmakat is használnak, de itt nem szükséges ezek kifejtése. Fontos viszont, hogy beemeljünk egy másik szempontot a relevancia modelljének vizsgálatába.

Amikor a relevanciafogalmakat kidolgozták, használatba vették, evidenciaként kezelték, hogy egyértelműen el lehet dönteni azt, hogy az információs tartományon belül mi számít relevánsnak, mi nem. Van, amikor ez a feltételezés helytálló, de van, amikor nem. Különösen nincs ez így a weben található dokumentumok esetében. Márpedig ekkor is nyilván ugyanolyan erős az elvárás a releváns találatok iránt a kereséseket végrehajtó felhasználókban.

3.3.1.2.2 Relevancia mint érték kategória

A navigáció mellett az informálódási igények jelentik az információs igények másik nagy csoportját. Ha eltekintünk a korábban jelzett bizonytalansági tényezőtől (ti. az informálódási és a navigációs igények között húzódó határ elmosódottságától), akkor azt mondhatjuk, hogy a felhasználó a navigációs lépések után végül elérkezik egy olyan szolgáltatási pontra (képernyőoldalra), ahol már a neki fontos tartalom egésze (vagy annak egy része) jelenik meg. Ekkor itt, ezen a ponton (képernyőoldalon) már neki kell megtalálnia a számára leginkább fontos információt (tartalmat vagy tartalomrészét).

A szolgáltatás irányából, a szolgáltató nézőpontjából nézve ez a probléma a felhasználó számára már könnyen és egyértelműen kezelhető. Ez pedig a megfelelő **információs architektúra** kialakításának kérdését jelenti: Milyennek kell lennie a képernyőoldaloknak ahhoz, hogy az átlagos felhasználó könnyen és gyorsan megtalálja rajta a számára fontos információt? Az információs architektúra tervezés a site-ok, a képernyőoldalak belső struktúrájával foglalkozik.

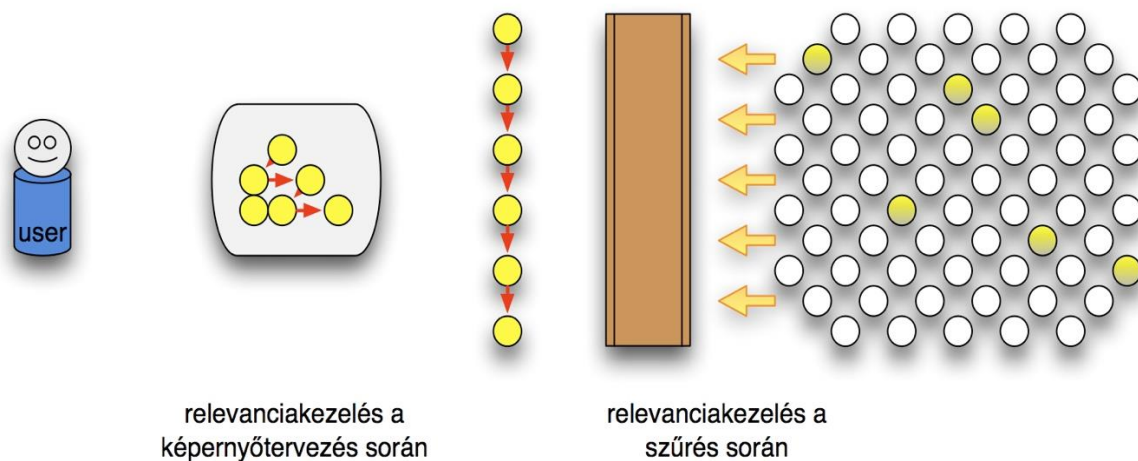
Ebben a kérdésben kiemelt fontossággal bír a **relevancia** fogalma. Ám ez a fogalom sajnos megint csak többértelműséggel terhelt, két szempontból is.

Egyrészt a relevancia fogalma az **érték** általánosabb fogalmára vezethető vissza, utóbbi kategóriában viszont van egy szubjektív mozzanat, amely a társadalmi kapcsolatok minden szintjén lényegi többértelműséget teremt (tehát ez a többértelműség nemcsak a hálózati kommunikáció világában érvényes). Hogy kinek, milyen információ, milyen dokumentum releváns, az az adott egyén konkrét –

tehát időtől, helyzettől függően változó módon létező, ható – értékkelköteleződésén múlik. Más lehet releváns két személynek ugyanabban a helyzetben, illetve ugyanannak a személynek két különböző helyzetben. Ez a „bizonytalanság” az értékfogalmak lényegi, inherens tulajdonsága.

Másrészt a relevancia fogalmával kapcsolatos bizonytalanság abból is fakad, hogy a relevancia kérdését mind az informálódási, mind a navigációs igények kielégítésének tárgyalásakor lehet (és kell) használni. A felhasználó a számára releváns információt, tartalmat szeretné mihamarabb megkapni, és közömbös számára, hogy ehhez külön-külön hány navigációs és hány informálódási lépést kell megtennie. A releváns tartalom mielőbbi elérése mellett persze az is fontos cél, hogy a felhasználó minél könnyebben, minél kevesebb lépésben férjen hozzá az őt érdeklő tartalmakhoz.

Tehát a dokumentumok kiválogatása – mind a szinkron, mind az aszinkron – szűrés során ugyanúgy a relevancia problémájához tartozik, mint az információs architektúra tervezésének összes kérdése.



a relevancia fogalmának kétfajta értelmezése

Mindenesetre amikor az informálódási igények kielégítéséről van szó – vagyis amikor már a releváns dokumentum tartalma van megjelenítve –, akkor a relevancia kérdése azt jelenti, hogy mit hova kell tenni az adott képernyőoldalon ahhoz, hogy a felhasználó a neki fontos információt minél könnyebben és gyorsabban megtalálhassa. A újonnan megjelenő oldalakat a felhasználók minden esetben szkennelik, „letapogatják”, és ezen befogadási folyamat során fontos, hogy mit, hol és hogyan érzékelnek, észlelhetnek. Ennek megtervezése sokkal inkább az interfésztervezés körébe tartozik, így itt nem kell mélyebben elmerülnünk ebben a témában, de annyit azért érdemes megjegyeznünk, hogy egy adott tartalom megjelenítéséhez érkező fontos kérdéssé válik a megjelenítendő metaadatokban, illetve tartalomban a szortolás, rendezés művelete. Ugyanis az előzetesen már megszürt anyagot akkor is rendezni kell, amikor már a leszűkített kínálatot kell (lehet) a felhasználók számára felkínálni, mivel még ilyenkor is kérdés, hogy az adott méretű képernyőn milyen rendben, milyen sorrendben kell az információt, tartalmat, metaadatokat megjeleníteni. Ez a kérdés azért is különösen fontos, mert a teljes képernyőfelület bizonyos részei köztudottan értékesebbek, fontosabbak a képernyő más részeinél, ezt pedig érdemes messzemenően figyelembe venni a képernyőtervezés során.

Csak érdekességképpen említjük, hogy a képernyők mérete és a szolgáltatásokhoz kapcsolódó felhasználói műveletek között olykor összefüggést mutatható ki, ahogy erre felhívta a figyelmet a híres interfésztervező szakember, Jakob Nielsen is [Nielsen 2009a]. Szerinte ugyanis a web kezdeti világában azért volt jellemző a keresésszupremácia (search dominance) jelensége (mint ahogy ma már nem az), mert kicsi volt a képernyő, és emiatt a böngészés során sokszor még nem lehetett megfelelő élményt biztosítani a felhasználók számára. Nielsen szerint ez a vonás köszön vissza napjainkban a mobilkommunikáció területén is.

Első pillanatra talán ennek ellentmond az a tény, hogy a 2000-es évek végének nagy sikerterméke, az iPhone viszont pont azért nem támogatja a keresés műveletét, mert az eszközt kényelmetlen keresésre használni. A kereséshez ugyanis a felhasználónak írnia kellene, amihez viszont más térkihasználás lenne

szükséges, ezt pedig nem akarták felvállalni az iPhone fejlesztésekor, mivel az iPhone filozófiájaként azt fogalmazták meg, hogy „Mindig megkapod a legfrissebb információt, neked csak böngészned kell!”

3.3.1.2.3 Relevanciafogalmak értékelése

Az előző néhány fejezetben áttekintettük a különböző relevanciaértelmezéseket, és a későbbiekben még visszatérünk a relevanciakezelés kérdésére a különböző típusú navigációs lehetőségek bemutatása során. Az eddigi elemzések összegzéseként, valamint a később következő fejtegetéseket megelőlegezve rögzíthetjük, hogy a relevancia fogalmának különböző jelentéseit az alábbi értelmezések mentén különíthetjük el egymástól.

- térbeli közelség relevanciája
- nyelvi (szintaktikai és szemantikai) közelség relevanciája
- értékjelzések relevanciája
- interfész relevanciája

A **térbeli közelség** relevanciája a valós fizikai térben (leggyakrabban a földrajzi térben) való mozgás, fizikai-térbeli navigáció során számít. Erről bővebben szólnunk még a térbeli navigáció fogalmának tárgyalásakor. A **nyelvi közelség** relevanciája a dokumentumokban, információs tartományokban (információs terekben) értelmezhető, amivel azt fejezhetjük ki, hogy bizonyos nyelvi egységek milyen közelségben vannak egymástól szintaktikai vagy szemantikai jellemzőik alapján. Az értékjelzések relevanciája a felhasználók interakciói során kifejezett értékjelzések alapján rendel valamilyen relevanciamutatót az információs tartomány elemi egységeihez. Az eddig felsorolt három relevanciafelfogáshoz képest némiképp más szempont szerint, más értelmezés mentén definiálhatjuk az **interfész relevancia** fogalmát. Akkor beszélhetünk erről, amikor a felhasználó interfészeken keresztül megjelenített információ befogadhatóságát ítéljük meg. Ezt a relevanciafogalmat a továbbiakban nem fogjuk elemezni, míg a többi értelmezésre még kitérünk.

3.3.1.3 Információs igények kielégítése és a felhasználó aktivitás

Az információelérési vagy információhozáférési tevékenységet nagyon gyakran összemossuk a keresés fogalmával, ami helytelen, ezért ezt a pongyolaságot most korrigálni kell.

Először is szeretnénk leszögezni, hogy ebben a fejezetben csak a navigáció problémájára fókuszálunk, tehát eltekintünk az információs igények teljes problémakörének tárgyalásától. Ha ezután azt a kérdést tesszük fel magunknak, hogy a felhasználók milyen módon juthatnak hozzá a gyűjteményekben tárolt tartalmakhoz, vagyis milyen navigációs műveleteket végezhetnek, akkor kiderül, hogy a keresés csak egy a kínálkozó lehetőségek közül. Azt kell figyelnünk, hogy a felhasználónak milyen tevékenységet kell elvégeznie azért, hogy hozzáférhessen valamilyen konkrét információfolyamhoz (egybefűzött tartalomcsomaghoz). Az alábbi három műveletet, három „logikát” különíthetjük el egymástól:

- **keresőlogika**
a felhasználó a teljes elérhető archivumi tartományból érheti el a kívánt tartalmakat úgy, hogy ő adja meg az általa választott/használt nyelven a keresőfeltételeket – ezért nem lehet biztosan felkészülni rá, nem lehet tudni a keresés előtt, milyen és mekkora találati listát kaphat a keresőszemély;
- **böngészőlogika**
a felhasználó szabadon barangolhat az előzetesen kidolgozott és számára – a saját döntéseitől részben függő módon – felkínált navigációs/döntési térben, amely a szolgáltató és/vagy a közösség által választott nyelven, terminusokkal a teljes archivumi halmaz valahogyan megszürt részéből kínálja fel választást a felhasználó számára (ide tartozik a linkgyűjtemény, a menü, az ajánlórendszer is);
- **műsorkövető logika** (szériális logika)
az elérhető tartalmak lineáris sorba vannak fűzve, a fogyasztó ezt az egybefűgő tartalomkínálatot fogadhatja kötött időrendben, nincs valós választási lehetősége.

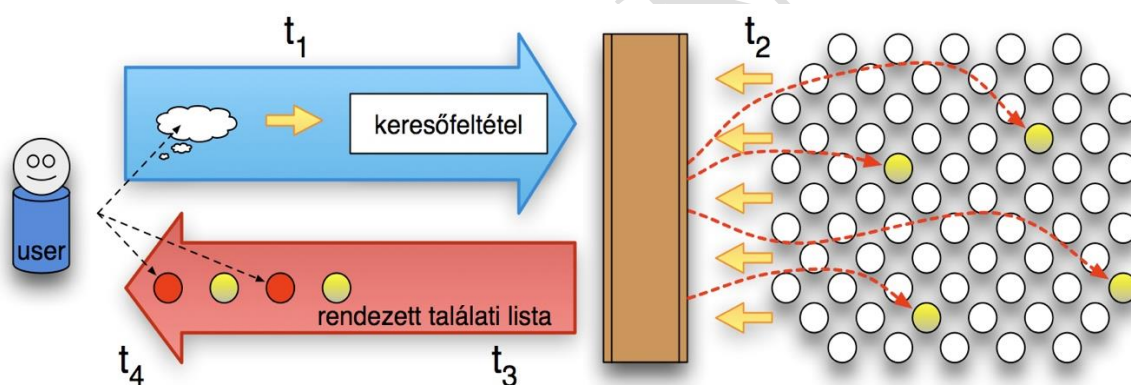
Nézzük meg kicsit alaposabban, mit is jelentenek az egyes műveletek! Mindhárom esetben abból az alaphelyzetből kell kiindulnunk, hogy a felhasználók akkora dokumentumgyűjteménnyel állnak szemben, amelynek használatához már szükségük van navigációs támogatásra.

3.3.1.3.1 Keresés

Amikor **keresünk** (searching), akkor mi adjuk meg a keresőfeltételt, amelynek alapján valamilyen választ kapunk az archívumszolgáltatótól. Egyelőre szűkítsük le a keresést a nyelvi, tehát szöveges keresésre. Ebben az esetben azt mondhatjuk, hogy mi magunk biztosítjuk a keresőkifejezést, mégpedig a „saját nyelvünkön”, ez pedig két szempontból is fontos. Egyrészt így a felhasználói irányból is nyelvi üzenetek formálódnak és mozognak a szolgáltató felé, akinek ezt értelmeznie kell, másrészt pedig szükség van egy nyelvi input eszközre, amellyel a felhasználó közvetlenül átadhatja a nyelvi üzenetét a kommunikációs eszköz (közvetve a szolgáltatója) számára. Ez ma még valamilyen billentyűzetet jelent, a valamilyen távoli jövőben pedig bizonyos körben és használati helyzetekben nyilván el fog terjedni a hangalapú „keresésifeltétel-beviteli lehetőség”.

Amikor a rendszer a keresőfeltétel alapján a feltételhez illeszkedő találatokat ad, akkor *leszűkíti* a teljes dokumentumtér kínálatát, tehát mondhatjuk, hogy abban a pillanatban *szűrést* hajt végre az összes dokumentum halmazán. A kereséshez **szinkron szűrés** kapcsolódik, amely nem előzetesen, hanem a kereséssel egyidőben, annak eredményeként valósul meg.

A keresőlogika érvényesülése egyben azt is jelenti, hogy a felhasználó feltételezhetően tudja, mit keres. De ez sokszor mégsem jellemzi a felhasználókat, mert a keresési tudatosság, illetve a keresési motiváltság könnyen hiányozhat az emberekből. A keresés helyzetét az alábbi ábrával szemléltethetjük:



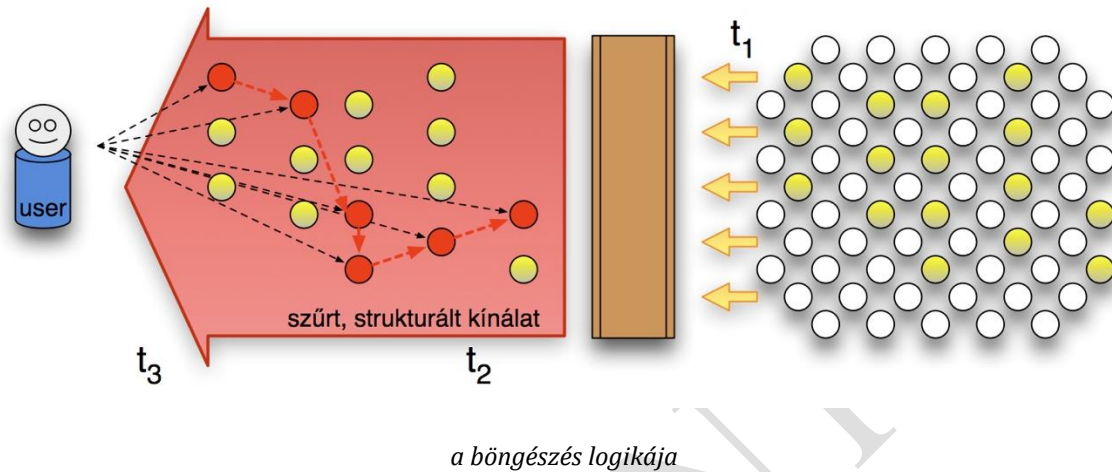
a keresés logikája

A keresés értékelése szempontjából nagyon fontos mozzanat, hogy a keresés műveletében a felhasználó aktív tevékenységet végez, amennyiben ő maga adja meg a keresőfeltételt. Ez a teljes folyamat első lépése (t_1). Ezután a keresőszolgáltató mintaillesztéssel megkeresi a feltételeknek megfelelő dokumentumokat (t_2), majd a találati listát valamilyen módon csoportosítva és sorba rakva felkínálja a felhasználó számára (t_3), amelyből a felhasználó végül kiválaszthatja, hogy melyik dokumentumot akarja megtekinteni (t_4). Ez az utolsó mozzanat felhasználói választást, döntést, interaktivitást kíván.

3.3.1.3.2 Böngészés

Ezzel szemben a **böngészés** (browsing) során lépésről lépésre haladunk, és a szolgáltató által felkínált navigációs információk, nyelvi kifejezések, bélyegképek, ugrópontok közül választunk. Ekkor csak a „szolgáltató nyelvén” történik aktív nyelvi kommunikáció („nyelvi adás”), míg a felhasználó oldalán csak passzív nyelvi befogadás van. A felhasználó böngésző magatartása kétféle lehet attól függően, hogy milyen „egységek között mozog”. Lehetséges, hogy egy ideig csak metaadatok között böngészget – ez azt jelenti, hogy az adott dokumentumgyűjtemény elemeinek megtalálásához leginkább illeszkedő metaadatokat keresi, majd miután megtalálta ezeket, hozzájut a legjobbnak tűnő dokumentumok

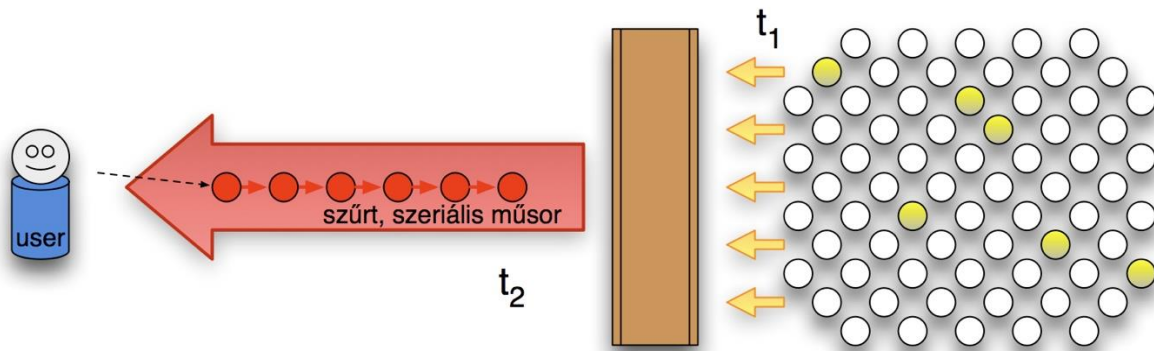
halmazához. A böngészés ilyenkor a metainformációs térben zajlik (ami egyben szemantikus keresést is jelent). A másik böngészési lehetőség során a felkínált dokumentumok (pontosabban azok reprezentánsai) között barangolhat a felhasználó. Állóképeket reprezentáló bélyegképek, videókat reprezentáló fázisképek, zeneszámokat reprezentáló metaadatokat, kisképek lehetnek különböző szempontok szerint csoportosítva, és a felhasználó ezek között válogathat a továbblépéshez. A böngészési tevékenység szerkezetét mutatja a következő ábra:



A böngészés műveletéből hiányzik a keresésben még megtalálható aktív, kezdeményező lépés. Ekkor nem a felhasználó adja az első inputot, amire a rendszer válaszol a dokumentumtér szinkron szűrésével, hanem az interakciós kapcsolatot megelőzi a szolgáltatás előzetes, *aszinkron szűrése* (t_1). Az szinkron szűrés eredményeként előálló metaadat-halmazt a szolgáltatás valamilyen bejárható struktúrába rendezve kínálja fel a felhasználó számára (t_2), és a felhasználó ebből az előzetesen megszürt és valamilyen módon strukturált kínálatból válogathat magának (t_3). A felhasználó interaktivitása ebben az apró lépésekből álló választássorozatban valósul meg.

3.3.1.3.3 Műsorkövetés

Ha a **műsorkövető** (vagy program) logika érvényesül, akkor a felhasználó passzív szerepben van. Miután kiválasztott egy „belépési pontot” (egy csatornát, egy lejátszási listát), azután egymás után kapja a dokumentumokat. Ilyenkor már nincs igazán választási lehetősége, a programrend szerint történik/történhet a dokumentumok befogadása. A programlogikát "tisztán" csak a két időbeli információtypust (az auditív vagy a mozgóképi tartalmakat) közvetítő médiumokon keresztül lehetséges érvényesíteni. Az audiovizuális dokumentumok ilyen egymást követő sorozatát műsornak nevezzük, és a műsorszolgáltatás szerkezetét így ábrázolhatjuk:



a műsorkövetés logikája

Ebben a felhasználó-szolgáltató kapcsolatban is előzetes aszinkron szűrés valósul meg, amikor a programtervezés során a műsorrendbe illesztik a kiválasztott dokumentumokat (t_1). Ebben a viszonyrendszerben a felhasználó számára csak annyi választási lehetőség marad, hogy eldöntse, mikor kezd el és mikor hagyja abba a műsor fogyasztását (t_2). Ez a logika a tipikus broadcast tévénézés és rádiózás sajátja, bár felbukkan a hálózati kommunikáció világában is (például a YouTube játszási listái). A hagyományos broadcast modellben csak annyi változik idővel, hogy a csatornák számának emelkedésével a felhasználónak lehetősége nyílik a programajánlatok (csatornák) közötti választásra. Ekkor megjelenik távirányítóval történő szörfözés, barangolás lehetősége, ám ez a csatornaválasztási (vagy belépési pont választási) lehetőség mindegyik navigációs logika esetében érvényesülhet, hiszen a keresést vagy a böngészést kínáló szolgáltatások között is ugyanúgy lehet választani, mint a tévé- vagy rádiócsatornák között.

3.3.1.4 Információs igények kielégítése és az információs térstruktúra

A felhasználók navigációs lehetőségeit más – és számunkra fontos – szempont szerint tipizálta Dourish és Chalmers [Dourish & Chalmers 1994]. Háromfajta navigációtípust különítettek el egymástól:

- **térbeli navigáció** (spatial navigation)
a felhasználó térbeli viszonyok mentén mozog egyik helyről a másikra: fentebb, lentebb, kívül, belül, jobbra, balra, közelebb, távolabb, mellette;
- **szemantikus navigáció** (semantic navigation)
a felhasználó szemantikus viszonyok mentén mozog egyik helyről a másikra: nagyobb, gyorsabb, szebb;
- **társas navigáció** vagy közösségi navigáció (social navigation)
a felhasználó társas viszonyok mentén mozog egyik helyről a másikra: kedveltebb, gyakoribban vagy gyérebben látogatott.

Minden navigáció alapja és mintája a **térbeli navigáció**. Ez nem meglepő. Már a navigáció fogalmának első, eredeti értelmezése is a végtelen térben való eligazodás igényét, lehetőségeit és technikáit fejezte ki. Említettük már a bevezetőben: amikor az első hajósok olyan messze bent jártak a tengeren, hogy már minden irányban elvesztettek minden tájékozódási lehetőséget, azaz bármerre néztek, csak tengert láttak, akkor elemi erővel tört fel bennük a tájékozódás, a „végtelen térben” való eligazodás igénye. Ez az élmény formálta a navigáció minden korabeli és később megjelenő technikáját.

Ugyanez a **végtelenségérzet**, valamint az ebből fakadó **elveszettségérzet** jellemzi a hatalmas archívumokban információt kereső ember hozzáállását is, csak ebben a helyzetben nem a valóságos, hanem egy képzelt/képzett, egy virtuális/információs térben kell a navigáció feladatát megoldania. A hálózati kommunikáció nem minőségileg, hanem csak mennyiségileg változtat ezen az alapélményen.

A térbeli navigáció szabályszerűségeinek folyamatos kutatása, elemzése azért is különösen fontos, mert minden másféle navigációt csak a képernyőkön keresztül valósíthatunk meg, és ilyenkor csak a térbeli

metaforát tudjuk használni. A navigációhoz szükséges információkat a képernyő képzett terében valóságos térbeli viszonyok mentén jeleníthetjük meg.

Az interakciótervezés, az információs architektúra tervezés, a navigációs struktúra tervezés mind, mind csak a térbeli navigáció megszokott formáira, megoldásaira támaszkodhatnak.

A **szemantikus navigáció** a képernyőn megjelenített, strukturált formai és tartalmi metaadatok közötti mozgást, eligazodást teszi lehetővé. Azok a navigációs pontok, amelyek a továbblépési lehetőségeket reprezentálják, egyrészt önmagukban is valamilyen szemantikai jelentéssel rendelkeznek, másrészt valamilyen szemantikai kapcsolat mentén vannak összekötve egymással. A felhasználónak tehát ebben a szemantikailag értelmezhető és szemantikailag tagolt virtuális-nyelvi térben (semantic space) kell mozognia, hogy a végén hozzáférhessen az őt érdeklő dokumentumok halmazához.

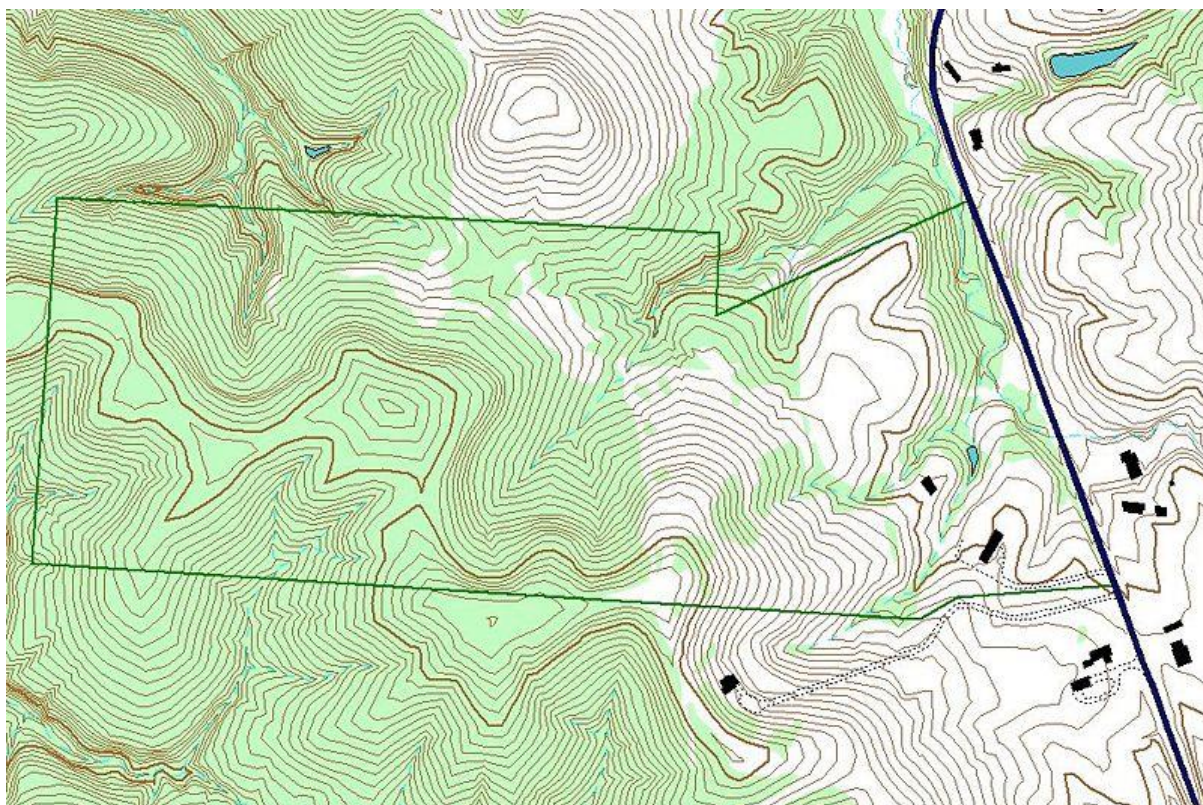
A **társas navigáció** alapvetően a használati információkra, illetve a felhasználók egymás közötti viszonyát reprezentáló adatokra támaszkodik. Az ilyen navigációt biztosító rendszer a továbblépési, eligazodási lehetőségeket annak alapján számolja ki és jeleníti meg, hogy talál-e valamilyen kapcsolatot bizonyos felhasználók között, valamint van-e valami hasonlóság a korábbi dokumentumhasználati adatokban a felhasználók vagy a dokumentumok jellemzői között. Ez a megközelítés, ez a gyakorlat tehát a társas kapcsolatokon, a közös használati mintákon alapul. A társas navigáció olyan szimbólumokat, gyakorlatokat, szabályokat, szabályosságokat alakít ki, amelyek révén az „egyszerű” (jelentés és értelem nélküli) térből (space) társadalmi tér (social space), vagy más néven: helyszín (place) formálódik ki.

A továbbiakban e három navigációtípus sajátosságait elemezzük.

3.3.1.4.1 Térbeli navigáció

A térbeli navigációval kezdjük, mert ennek tárgyalása itt röviden elvégezhető. Bár ez az alapja az összes navigációs tevékenységnek, számunkra most mégsem érdekes a térbeli navigáció egészének problémaköre, mivel mindez nem igazán releváns az információs terekben történő navigáció szempontjából. A térbeli navigáció esetében ugyanis mindig igazodni kell a valóságos térbeliség sajátosságaiból fakadó kényszerekhez. Ez így van a virtuális 2D-s vagy 3D-s terek esetében is. Ahhoz, hogy a képernyőn látható valóságos vagy virtuális világot térbelinek érezzük, arra van szükség, hogy a geometriai primitívekre, alapformákra vonatkozó invariánsok közül (helyzet, hossz, osztóviszony, párhuzamosság, kettős viszony, szomszédosság) közül legalább egy (de inkább több) érvényesüljön (Ankerl 1991). Ennek hiányában nem érezhetünk valódi térbeliségélményt. És ezek az invariánsokra vonatkozó elvárások, kényszerek hiányozhatnak, hiányoznak a szimbolikus, információs terekből.

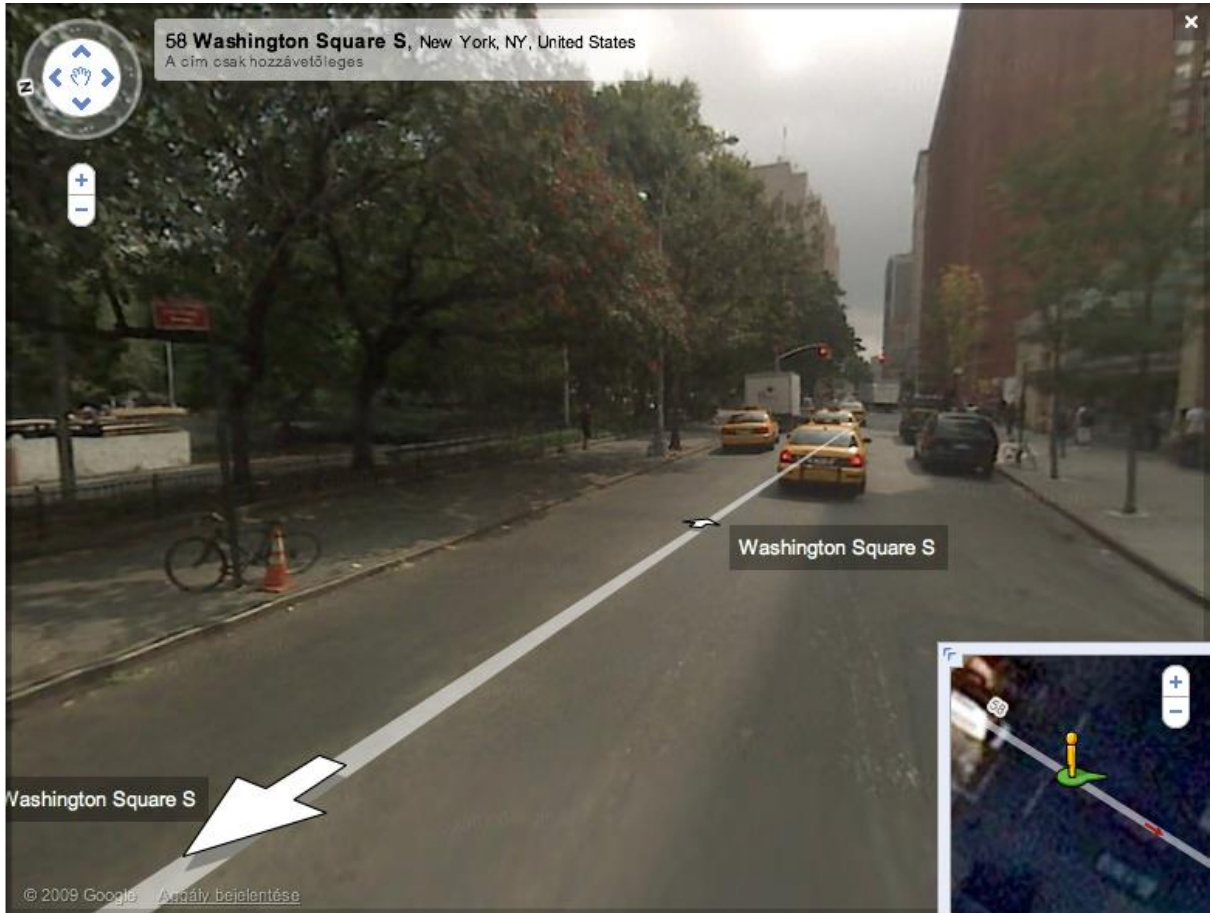
A valós fizikai tér digitális, virtuális reprezentációjában, a virtuális földrajzi térben történő navigáció a térinformatika, valamint a mérnöki tervezés körébe tartozik, de ezek a szakterületek nem fontosak e tanulmány témájának szempontjából. Sokféle módon lehet a földrajzi teret reprezentálni. Ezekből mutatunk be négy példát:



ArcView: térinformatikai térbeli navigáció



Google Maps: utcaterkép szerinti térbeli navigáció

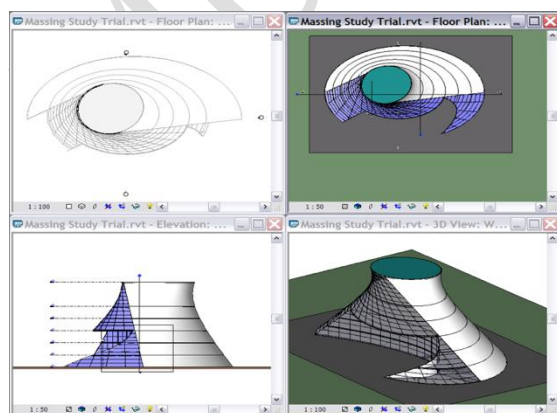


Google Maps: utcafénykép szerinti térbeli navigáció



Google Maps: műholdas felvétel szerinti térbeli navigáció

A valós földrajzi térben való eligazodás térinformatikai technikái mellett léteznek másfajta térrepresentációs célok és megoldások. A mérnöki tervezést (gépek, házak stb. tervezését) segítő informatikai eszközök „elszakadnak” a földrajzi tértől abban az értelemben, hogy a fizikai teret relatívnak tételezik, s az általuk létrehozott tereket bárhova el lehetne tolni a valós vagy virtuális földrajzi térben, de a fent említett invarianciaviszonyokhoz ezekben a világokban is igazodni kell.



AutoCad formatervezés: 3D-navigáció



AutoCad háztervezés: 3D-navigáció

A mérnöki térkezelői rendszerek mellett léteznek még a virtuális világok, a 3D-s játékok, amelyekben a térben való mozgás és eligazodás élménye megegyezik a valóságos térmozgás élményével – és itt is ugyanazok a kényszerek hatnak, mint a többi tényleges térbeli navigációs helyzetben.

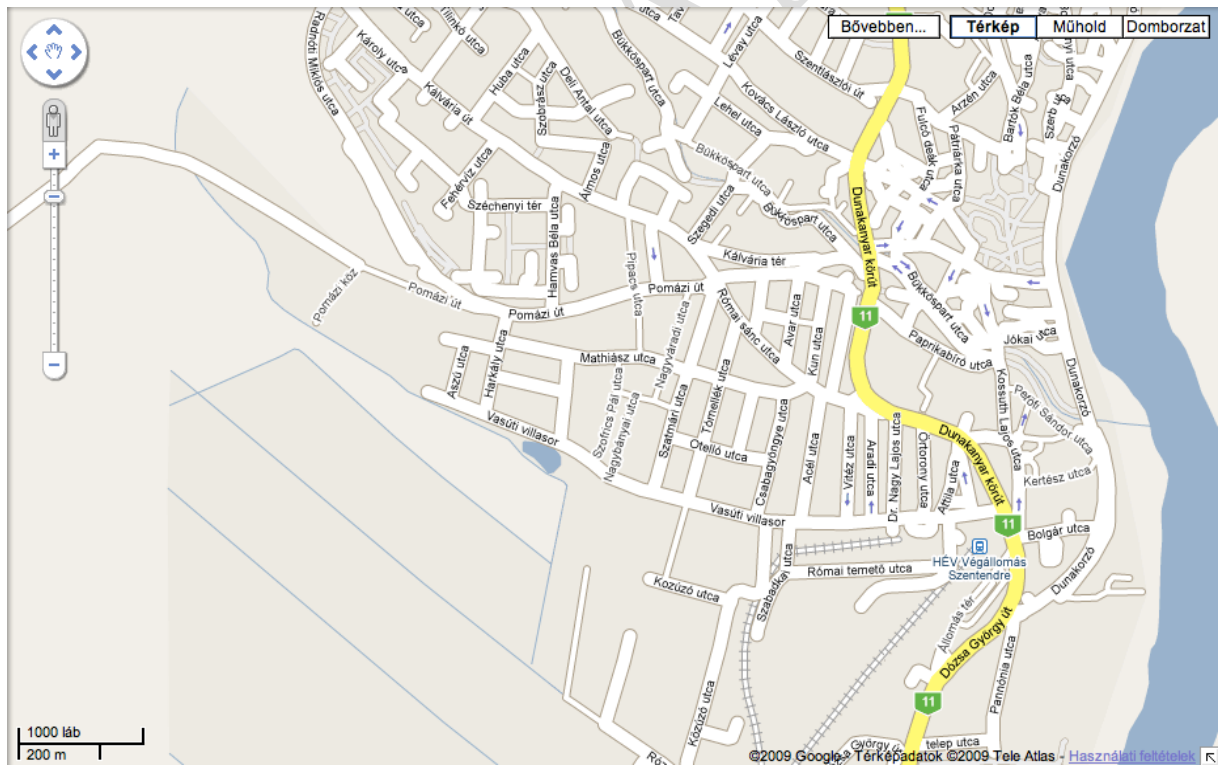


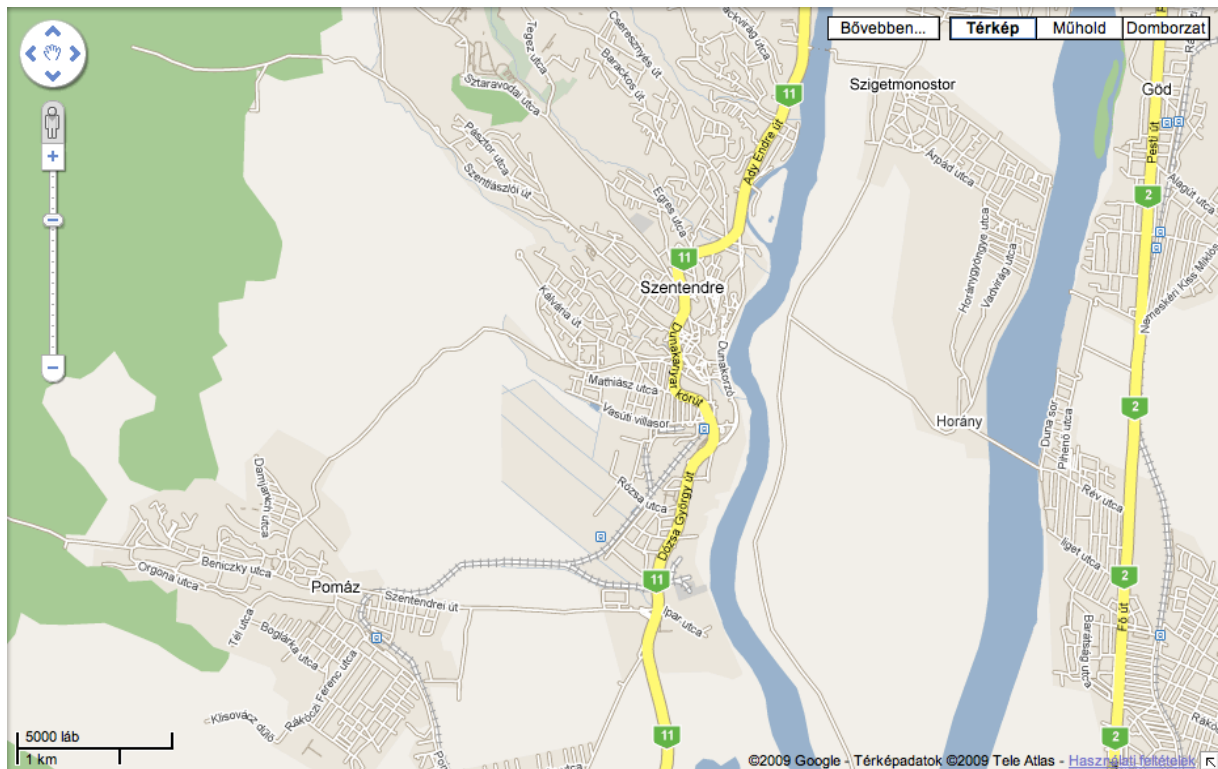
Secondlife virtuális világ: 3D-navigáció



Doom virtuális játék: 3D-navigáció

A képernyőn megvalósuló térbeli navigáció egyik fontos eszköze az ún. *ZUI-elv* (ZUI – Zooming User Interface). Mivel a képernyőn sosem fér el a kezelni kívánt – valóságos vagy virtuális – tér teljes egésze, ezért meg kellett találni a módját, hogy a tér egyik részletének megjelenítése után hogyan lehet a tér egy másik részletét is megmutatni a képernyőn. A térképészeti praxisban mindig is evidencia volt, hogy a földrajzi teret különböző léptékben, felbontási mélységben lehet reprezentálni, így adta magát, hogy a digitális térkezelésben is különböző felbontású térképekkel, változó léptékű reprezentációval kell/lehet dolgozni. A digitális információkezelés ehhez azt az új lehetőséget adta, hogy a különböző felbontások között sokkal könnyebben lehet váltani a *zooming* (léptékváltás, közelítés/távolítás) technikával. Alább három térkép látható, amelyek eltérő felbontással ugyanazon középpont körül mutatják a földrajzi tér – a felbontás által meghatározott módon – befogható részét:





három térképrészlet a zoomolás eredményeként

A zooming elvet később ki lehetett terjeszteni mindenfajta térkezelésre. Természetesen amikor a fizikai tér invarianciakényszereire nem igazodó szimbolikus-információs tereket hoztak létre, akkor is kézenfekvő volt ennek a megoldásnak az alkalmazása. A gyakorlatban ilyen esetekben beszéltek, beszélnek ZUI-ról. Ennek problémáival azonban nem itt, a navigáció jelenségének tárgyalásánál, hanem a felhasználói felületek elemzése során kell foglalkozni.

Korábban jeleztük már, most újra hangsúlyozzuk: habár jelen tanulmányban nem tárgyaljuk a térbeli navigáció kérdését, ettől még ez a navigációtípus a legfontosabb valamennyi közül, hiszen ez szolgál minden más eligazodási tevékenység alapjául (legalábbis a képernyőn keresztül megvalósuló, vizuális navigáció esetében). A vizuális navigáció mindenképpen valamilyen térben létező, ott elhelyezett objektumok, jelek segítségével valósul meg. Ebben az értelemben minden – vizuális – navigáció egyben térbeli is. (ebből következően kissé pongyola a térbeli navigáció fogalma, de a tanulmányunk kontextusában ez nem annyira zavaró, mivel a szűkebb értelemben vett térbeli navigáció kérdésével itt nem foglalkozunk.)

A következőkben tárgyalandó – szemantikus és társas – navigáció ugyanúgy térbeli navigáció, csak ezekben az esetekben a képernyőben elhelyezett navigációs objektumokra vonatkozóan nem kell érvényesítenünk az „igazi” térbeli navigáció esetében szükségszerűen elvárt invarianciakényszereket.

3.3.1.4.1.1 Térérzékeny keresés lehetőségei

A térbeli navigáció digitális térben kétféle módon is megvalósulhat. Egyfelől az okos eszközök a földrajzi térben való pontos helyzetük érzékelésével segítséget képesek adni az őket használó emberek számára. A navigációs szolgáltatások példájára hivatkozhatunk itt. Ekkor a szó szoros értelmében vett térbeli navigációról beszélhetünk. Másfelől azonban a térbeli navigációt lehet úgy is értelmezni, ami már nem a valóságos földrajzi térben való eligazodást segíti, hanem a földrajzi térhez köthető információs tér adott szegmensében való keresést.

A dokumentumok szimbolikus terében a dokumentumokat valamilyen módon jellemző információs térben értelmes lehet a földrajzi térhez valahogyan kapcsolódó adatokban való ún. „térérzékeny keresés”. Az ilyen keresésekben rejlő lehetőségeket a projekt egy másik dokumentumában fejtjük ki bővebben.

3.3.1.4.2 Szemantikus navigáció

A szemantikus keresés nem a földrajzi térben, hanem a szimbolikus térben, valamilyen információs tartományban való eligazodást segíti. Ennek első szintjét a szabadszavas keresés jelenti, ami annyira lapos szemantikai képességgel rendelkezik, hogy vitatni lehetne, vajon tényleg szemantikai keresésről van-e szó, de az egyszerűség kedvéért itt elfogadhatjuk azt, hogy már ekkor is van valamilyen szemantikai rétege a keresésnek. A szabadszavas keresés rövid tárgyalása után sorba vesszük a szabadszavas keresés szemantikai gyengeségét meghaladni képes technológiákat, az adatbázis-alapú, a kérdés-válasz alapú és a KOS-alapú kereséseket.

3.3.1.4.2.1 Szabadszavas keresés

Hatalmas előrelépést, minőségi váltást jelentett, amikor a számítógépek segítségével az írásos dokumentumok tartalmában, a szövegben keresni lehetett. A számítógép gyorsaságban és pontosságban sok nagyságrendben verte, veri az embert ezen a téren. Mivel a digitális térben egyre nagyobb arányban kerültek, kerülnek fel az írásos dokumentumok, amit a keresőmotorok feldolgoznak, kereshetővé tesznek az invertált indexek segítségével, sokak számára úgy tűnt és úgy tűnik, hogy nincs is másra szükség, mint hogy minden szöveges dokumentumot digitalizáljunk, és betegyük a már eleve digitális formában létrejött szövegek mellé, és a keresőmotorok gyorsan és pontosan navigálhatóvá teszik ezt az információs tartományt. Ezt az ideát érhetjük tetten azokban az elképzelésekben, amelyek szerint mindenre ott a Google, ami mindent megtalál, nincs is szükség semmi másra ('googlism').

Nem vitatva, hogy a szabadszavas keresés – a korábbi állapotokhoz képest – valóban hihetetlen mértékben kibővíti az emberi keresés lehetőségeit, látnunk kell azt is, hogy a szabadszavas keresésnek nagyon erős korlátai, komoly gyengeségei is vannak. A következő fejezetben bemutatunk párat ezekből a hiányosságokból.

3.3.1.4.2.1.1 A szabadszavas keresés szemantikus vaksága

Rögzítsük le újra a kiinduló álláspontunkat: a web egyik forradalmi újdonságát (és erejét) a *szabadszavas keresés* megjelenése jelentette, amelynek sikere sok embert annak kimondására sarkallt, hogy talán nincs is szükség másfajta keresési lehetőségre. A szabadszavas keresés (főleg a relevanciakezelés különböző megoldásaival együtt) valóban nagyon jó eredmények érhetőek el, és korábban elérhetetlennek tűnő lehetőségeket kínál számunkra, de mégis tudnunk kell, hogy a technológiának nagyon komoly és kiküszöbölhetetlen hiányosságai vannak.

A szabadszavas keresés legfontosabb problémája nyilván az, hogy csak szöveges dokumentumok esetében működőképes, vagyis az audiovizuális dokumentumok esetében nem használható.³³ De még „tisztá” szöveg esetében is komoly gondok adódhatnak vele. Mivel a szabadszavas keresőkben nem tudjuk minősíteni a keresett kulcsszavakat, ezért nincs mód a metaadatok minősített keresésére (a szöveg strukturálatlanságával nem képes megbirkózni ez a technológia, emiatt a metaadatok kezelésére egyáltalán nem képes).

A szabadszavas keresés során nem tudunk a szerző vagy a dokumentum címe szerint keresni. Ha a dokumentumban „valódi” címként szerepel egy karaktersorozat, míg egy másikban csak hivatkoznak ugyanerre a címre, akkor mindkettőt visszkapjuk – függetlenül a kifejezés „metainformációs státusától”.

Mondhatnánk persze azt, hogy a tartalmi feltárás területén viszont hatalmas előnyt jelent az a lehetőség, hogy – legalábbis első körben úgy tűnik – nem kell elvégezni a tárgyszavazás fáradságos munkáját, mivel az automatikus gépi indexelés elvégzi azt az ember helyett. Ez vitathatatlan tény, és ez valóban komoly előnyt biztosít, de azért ennek a megoldásnak is vannak hátulütői. A gépi indexelés ugyanis nem tud mit kezdeni a nyelv többértelműségével, amelyet viszont az ember igenis képes kezelni a feltáró munka során. Nézzünk meg ezekből néhány példát!

A gépi keresés nem tudja elkülöníteni a *homoním* jelentéseket, akár a köznévi, akár a tulajdonnévi alakok közötti homonimákról van szó:

macska (állat) – vasmacska (eszköz)

³³ Audiovizuális dokumentumon képi és/vagy hangyi információkat tartalmazó dokumentumot értünk.

Cica ('Révész Cica József klarinétos') – cica (macska)

A keresést támogatásának egyik fontos eszköze a szavak, kifejezések közötti *szinonimitás*, amellyel az automatikus indexelés semmit sem tud kezdeni. A számítógép számára nem nyilvánvaló (nem ismert) az alábbi három terminus (durván) azonos jelentése:

macska - cicus - cica

Nem kell elmerülnünk abban a vitában, amely a szinonima és a *poliszémia* elválasztásáról zajlik a nyelvészet területén – elég csak jeleznünk, hogy a szavak bizonyos jelentései között gyakran nincs teljes átfedés (vagyis nem beszélhetünk szinonimitásról), mégis szoros kapcsolat állapítható meg közöttük. Az alábbi példa nem lehet szinonima, hiszen állatról és emberről van szó, mégis érezzük a két jelentés kapcsolódását egymáshoz.

cica (macska) – cica (csinos, fiatal nő)

Aztán a nyelv megint „továblép” egyet, amikor a pozitív töltetű jelentést átértelmezi, és ironikusan használja a terminust.

cica (csinos, fiatal nő) – fitnesscica (kövér, csúnya nő)³⁴

Az embernek nehezebb észrevenni azt, ami a számítógép számára nyilvánvaló: nagyon sok olyan *akronimát* (betűszót) képzünk, amelyek a köznyelv valamely szavával teljesen megegyeznek.

CICA (Criminal Injuries Compensation Authority) – CICA (Confederation Internationale du Credit Agricole)
– cica (macska)

A keresés során talán nem annyira fontos, de mégis említésre méltó az az emberi képesség (amellyel a gépek egyelőre nem rendelkeznek), hogy azonnal felismerjük bizonyos szavak, kifejezések ellentétes jelentését, vagyis kezelni tudjuk az *antonima* nyelvi jelenségét.

kutya – macska

A nyelv rugalmasságának egyik legfontosabb eszköze a *metonima*, amellyel sokféle értelemben kiterjeszthetjük egy szó jelentését

Háromszintes iskolába jár. (iskola mint épület)

Nyolcosztályos iskolába jár. (iskola mint intézmény, testület)

Tovább bonyolódik a helyzet, amikor a hálózat többnyelvűségét is figyelembe kell vennünk, és a fenti többértelműségek a különböző nyelvek között is létrejöhetnek. Előfordulhat „többnyelvű szinonimitás”, amikor különböző nyelveken fejezzük ki ugyanazt a jelentést, tartalmat:

macska (magyar) - cat (angol) – Katze (német)

³⁴ Az interneten keringenek olyan videók, amelyek konditeremben a futógépet használó kövér nőket írják le a fitnesscica címkével.

A másik irányból tekintve értelmezhetünk „többnyelvű homonimitást” is, amikor ugyanaz a szóalak két nyelven teljesen más jelentéssel bír.

cica (magyarul) – čiča (popsi) csehül

A fenti példák mind azt mutatták, hogy egyetlen karaktersorozat milyen sokféle jelentéssel rendelkezhet, amelyeket a számítógépek – egyelőre még – nem, az emberek viszont képesek megfelelő módon, vagyis szemantikus gazdagságuk teljességében értelmezni. A 'cica' karaktersorozatnak az alábbi különböző értelmezéseit adhatjuk meg (s a felsorolás nyilván nem teljes):

magyar: cica (macska)
angol: CICA (tulajdonnevek)
magyar: cica, fitnesscica, plázacica (nő)
magyar: papírcica (origami cica)
magyar: porcica (kosz)

A számítógép mindezekről semmit sem tud, a 'cica' terminust nem tudja értelmezni, nyugodtan mondhatjuk azt, hogy teljes **szemantikus vakságban** szenved – ma még. A Szemantikus Web Kezdeményezés célja és értelme pont az, hogy a keresőmotor számára szemantikus képességeket fejlesszünk ki.

3.3.1.4.2.1.2 Szabadszavas keresés és relevanciakezelés

A webes keresőmotoroknak azonban van még egy igen komoly problémája a szemantikus vakság mellett. A szabadszavas keresés **teljes-szöveges keresést** (fulltext search) is jelent, amely abból a szempontból előnyösnek és kívánatosnak minősíthető, hogy ezáltal a szöveges dokumentumok teljes szókészlete kereshetővé és elérhetővé válik (ez különösen szórványszavak esetében lehet nagyon hatékony), ugyanakkor a **relevanciakezelés** területén új nehézségek merülnek fel – vagyis pontosabb lenne azt állítani, hogy a relevancia egész problémaköre teljesen új megvilágításba kerül. Amikor a keresőmotorok a szöveges dokumentum összes szavát leindexelik, és az invertált indexet követve a szövegben előforduló bármely szó alapján képesek visszaadni a dokumentumot magát, akkor rögtön felmerül a kérdés, hogy az adott keresőszó vajon jól jellemzi-e az adott dokumentumot, vagyis kellően releváns-e a szó és a dokumentum kapcsolata. A válasz nyilvánvalóan csak az lehet, hogy nem minden szó jellemezheti ugyanolyan mértékben a dokumentumot, tehát szükség lenne egy olyan módszerre, amely a szavakat tartalmazó dokumentumokhoz valamilyen relevanciaértéket rendel, és ennek alapján a fontosabbnak minősített dokumentumokat előbbre lehet rangsorolni a találati listákban. Az első generációs keresőgépek a relevanciaértékeket magában a dokumentumban keresték. Ebben persze önkéntelenül is követték azt a hagyományos, offline gyakorlatot, amely szerint a dokumentumok tárgyszavakkal leírhatók. Ez az évszázados könyvtári hagyomány azt sugallta, hogy a teljes-szöveges indexállományból ki lehet választani néhány (vagyis kevés) releváns tárgyszót a dokumentum tartalmi leírására. A kérdés csak az volt, hogy vajon hogyan lehet megtanítani a számítógépnek azt, amit a könyvtárosok könnyen és jól meg tudtak oldani (ti. a kevés releváns tárgyszó kiválasztását). A keresőmotorok fejlesztői az első időszakban (a web kezdeti időszakában) olyan szempontokat próbáltak meg szem előtt tartani, mint:

hányszor fordul elő a szó a dokumentumban (ha többször, akkor „többet ér”)

hol szerepel a szó a dokumentumban (ha az elején, akkor „többet ér”)

szerepel-e a szó a dokumentum címében, alcímében (ha igen, akkor „többet ér”)

Ezek a kezdeti próbálkozások azonban kevésbé (vagy egyáltalán nem) voltak sikeresek. A Google volt az első keresőmotor, amely valóban működőképes relevanciakezelést valósított meg. A Google

megoldása, a PageRank azonban teljesen más szempont szerint működött, mint az elődei. A dokumentumok relevanciaértékeinek számításakor ugyanis nem a dokumentumok tartalmát (a benne szereplő szavakat) vette figyelembe, hanem a dokumentumokban elhelyezett linkek (más oldalakra mutató utalások) tényét, számát, súlyát. A weboldalak készítőinek szubjektív ítéleteit lehetett ezáltal összegyűjteni és aggregálni valamiféle közösségi fontossági mutatószámmá. A relevanciakezelés nem jelent egyebet, mint valamilyen módon kifejezni, hogy adott dokumentum, adott kontextusban fontos egy személy vagy egy közösség számára.

A könyvtárosok addig más technikát alkalmaztak a relevanciakezelés problémájára. Amikor a könyvtárosok a dokumentumok tartalmi leírását végezték, akkor a rendelkezésükre álló, elméletileg összes lehetséges leíró tárgyszó közül kiválasztották azokat, amelyeket a legfontosabbnak tartottak, és ezeket hozzárendelték a dokumentumokhoz. A leíró tárgyszavaknak ez a kiválasztása, szűrése egy bizonyos szempontból nagyon hasonlított a keresőmotorok teljes-szöveges indexeléséhez. A könyvtárosok az előre rögzített tudásszervezési rendszer elemeiből válogatták ki a legfontosabbnak tartott leíró elemeket, vagyis számukra ugyanúgy rendelkezésre állt egy előzetes szóhalmaz, amelyből aztán választaniuk kellett, mint ahogy a keresőmotorok is minden egyes dokumentumról felállították a dokumentum összes szavából álló szóhalmazt, és ennek elemeihez tudták hasonlítani a későbbi felhasználói keresések során megadott keresőfeltételeket. A kétféle gyakorlat között annyi a különbség, hogy a keresőmotorok nem tudták, nem tudják jól kiválasztani a dokumentumot valóban jellemző, releváns tárgyszavakat (vagy pedig más relevanciakezelő megoldást alkalmaztak, mint a Google).

A probléma megoldására, a hiányzó szűrési, kiválasztási tevékenység elvégzésére szakemberek munkába állítása látszott megfelelőnek. Részben ezért indítottak a web kezdeti időszakában olyan szolgáltatásokat a szabadszavas keresőmotorok (a HotBot, az AltaVista és társaik megjelenésével párhuzamosan), amelyek szemantikai szempontból kívánták meghaladni a keresőmotorok szolgáltatásait. A legismertebb próbálkozás a Yahoo! Directory webkatalógusa volt, amely a weboldalakot egy saját osztályozási rendszer segítségével próbálta meg szemantikailag elrendezni és a felhasználók számára megtalálhatóvá tenni.



- [Arts](#) - - [Humanities](#), [Photography](#), [Architecture](#), ...
- [Business and Economy \[Xtra!\]](#) - - [Directory](#), [Investments](#), [Classifieds](#), ...
- [Computers and Internet \[Xtra!\]](#) - - [Internet](#), [WWW](#), [Software](#), [Multimedia](#), ...
- [Education](#) - - [Universities](#), [K-12](#), [Courses](#), ...
- [Entertainment \[Xtra!\]](#) - - [TV](#), [Movies](#), [Music](#), [Magazines](#), ...
- [Government](#) - - [Politics \[Xtra!\]](#), [Agencies](#), [Law](#), [Military](#), ...
- [Health \[Xtra!\]](#) - - [Medicine](#), [Drugs](#), [Diseases](#), [Fitness](#), ...
- [News \[Xtra!\]](#) - - [World \[Xtra!\]](#), [Daily](#), [Current Events](#), ...
- [Recreation and Sports \[Xtra!\]](#) - - [Sports](#), [Games](#), [Travel](#), [Autos](#), [Outdoors](#), ...
- [Reference](#) - - [Libraries](#), [Dictionaries](#), [Phone Numbers](#), ...
- [Regional](#) - - [Countries](#), [Regions](#), [U.S. States](#), ...
- [Science](#) - - [CS](#), [Biology](#), [Astronomy](#), [Engineering](#), ...
- [Social Science](#) - - [Anthropology](#), [Sociology](#), [Economics](#), ...
- [Society and Culture](#) - - [People](#), [Environment](#), [Religion](#), ...

a Yahoo! címlapja 1996-ban

A szolgáltatás címlapjának nagy részét a Yahoo! Directory tölti be, ami egyértelműen jelzi, hogy ekkor még ez a szolgáltatás állt a Yahoo! portfóliójának központjában – még ha vannak már másfajta szolgáltatásai is a cégnek (például személykereső, városi térképek, tőzsdei adatszolgáltatás). Idővel – a portál bővülő funkcionalitásával párhuzamosan – egyre több és több szolgáltatás linkjét, ikonját tették ki a címlapra azon az áron, hogy a Directory felsőszintű kategóriájának egyre kevesebb hely jutott. 2004. november elsején végül meghozták azt a döntést, hogy ha nem is bezárják a Directory szolgáltatást, de leveszik a Yahoo! címlapjáról – és így közvetve elismerték a szolgáltatás sikertelenségét.

My Personalize **Finance** **Shop** **YAHOO!** **Mail** **Messenger** **HotJobs** ? Help

HotJobs - Have a Happy New Job, Get Started Now, New Jobs Daily, How Much Should You Be Earning?

Search for: on the Web [Advanced Preferences](#)

New! Yahoo! Health - [Fitness Guide for a healthy 2004](#)

Shop [Auctions](#), [Autos](#), [Classifieds](#), [Real Estate](#), [Shopping](#), [Travel](#)

Find [HotJobs](#), [Maps](#), [People Search](#), [Personals](#), [Yellow Pages](#)

Connect [Chat](#), [GeoCities](#), [Greetings](#), [Groups](#), [Mail](#), [Messenger](#), [Mobile](#)

Organize [Addresses](#), [Briefcase](#), [Calendar](#), [My Yahoo!](#), [PayDirect](#), [Photos](#)

Fun [Games](#), [Horoscopes](#), [Kids](#), [Movies](#), [Music](#), [Radio](#), [TV](#)

Info [Finance](#), [Health](#), [News](#), [Sports](#), [Weather](#) [More Yahoo!...](#)

Personal Assistant [Sign In](#)
Get free email with SpamGuard!
Yahoo! Mail - [Sign up now](#)

Yahoo! Gamesource
Reviews: [Final Fantasy XI](#), [Socom II](#), [Max Payne 2](#), [More...](#)

In The News 4:02pm, Thu Jan 1

- Revelers ring in 2004 amid heavy security
- Two more Iranian quake survivors found
- Officials worry about sea-based terror
- Pakistan flight lands in Indian capital
- Apple users threaten to sue over iBook
- Woman has heart attack on jet full of MDs
- NCAA: [Football](#) - [Hoops](#) | [NFL](#) · [NBA](#) · [NHL](#)

[News](#) - [Photos](#) - [Sports](#) - [Stocks](#) - [Weather](#)

Marketplace

- [Canon Powershot Digital ELPH](#)
Greater zoom power, more life-like color reproduction, movie clip recording, and even its own little printer
- [Manage asthma online with MediCompass](#)
- [Sony Digital Cameras](#) - Direct from Sony. Free shipping and 6 month financing
- [100% online degrees](#) - Get your Bachelor's, Master's or Doctoral here
- [SBC Yahoo! DSL](#) - Free activation and installation. **Only \$26.95/mo for a year.**

[Shopping](#) - [Computers](#) - [Electronics](#) - [Travel](#)

Web Site Directory - Sites organized by subject [Suggest your site](#)

Business & Economy
B2B, Finance, Shopping, Jobs...

Computers & Internet
Internet, WWW, Software, Games...

News & Media
Newspapers, TV, Radio...

Entertainment
Movies, Humor, Music...

Regional
Countries, Regions, US States...

Society & Culture
People, Environment, Religion...

Education
College and University, K-12...

Arts & Humanities
Photography, History, Literature...

a Yahoo! címoldala a directory eltünése előtt, 2004

A drámai fordulatot is ennek a szolgáltatásnak a sorsa is mutatja. Noha a Yahoo! idővel a legsikeresebb webes szolgáltatások közé került, de a webkatalógusa egyre inkább háttérbe szorult, míg végül „bezárták”, a könyvtárosaikat, archívoraikat pedig szélnak eresztették (vagy más feladatokra csoportosították át őket).³⁵

A szakszerű és fegyelmezett rend képviselője eltűnt, ám ezzel párhuzamosan megjelent valami más. A web2 jelenségkörbe tartozó szolgáltatások (mint a Flickr, Del.icio.us, Digg, YouTube, Instagram, Facebook stb.) ugyanis olyan metaadat-kezelési módszereket építettek ki, amelyek a munkát – amit addig a szakemberek végeztek el a hagyományos és digitális archívumokban egyaránt –, az új szolgáltatásokban az önként dolgozó felhasználókra bízta. Ezt a megoldást, pontosabban az ilyen rendszereket nevezték el **folkszónómiának**. A szemantikus navigáció jelenségkörén belül az az igazán fontos kérdés, hogy mi okozta a hagyományos archívumi navigációs megoldások bukását és a hálózati

³⁵ A Yahoo! Directory szolgáltatáshoz hasonló utat futott be a Google által felkarolt DMOZ Open Directory projekt is, amely szinte a kezdetektől fogva és elég hosszú ideig elérhető volt a Google kezdőlapjáról, aztán egyszer csak eltűnt onnan.

jelenség, a folkszonómiák sikerét, illetve miként értelmezhetjük az utóbbiakhoz köthető paradigmaváltást.

Elemzésünkben leegyszerűsített gondolatmenetet követünk az eddig felvázolt fogalmi modell összetevőire támaszkodva. A legfontosabb kérdésünk az lesz, hogy a digitális archívumok elterjedésével milyen módon lehet biztosítani a tárolt dokumentumok metaadatokkal történő ellátását, és a dokumentumok visszakereshetőségét. Ehhez a tudásszervezési rendszerek fogalmát és annak különböző típusait lehet majd igénybe venni, de meg kell vizsgálni azt is, hogy mit előnyt és hátrányt jelent, ha a szövegben tárolt tudásunkat strukturáljuk, azaz adatbázisokba rendezzük és felkínáljuk az ezekben való keresés lehetőségét, illetve milyen előnyei és hátrányai vannak a kérdés-válasz (QA) rendszereknek, és milyen lehetőségeket kínálnak az ilyen rendszerek az információkeresés során.

3.3.1.4.2.2 Adatbázis-keresés

A szabadszavas keresőknek a hatalmas előnyeik mellett van néhány jelentős hátrányuk is. Ezek abból a tényből fakadnak, hogy ezek a keresőmotorok – szöveges – dokumentumok gyűjteményeiben keresnek, és az invertált indexek révén találatként dokumentumokat adnak vissza – azzal az „instrukcióval” vagy ígérettel, hogy a keresett kifejezés valahol megtalálható a listázott dokumentumokban, tehát a felhasználónak valószínűleg érdemes elolvasnia a dokumentumokat. A dokumentumok relevanciáját pedig azzal fejezik ki, hogy találati lista elemeit az adott keresőmotor relevanciaképző algoritmus szerint rangsorolják – előbbre téve azokat a dokumentumokat, amelyeket az algoritmus fontosabbnak „gondol”. Ezzel a megoldással, technikával szemben több komoly kritikát lehet megfogalmazni.

Egyrészt az elérhető (kereshető) dokumentumok számosságának növekedésével egyre nagyobb gondot jelent az információátterhelés jelensége, vagyis az a tény, hogy olyan számban kapunk találatokat a keresőmotoroktól, hogy már régóta nyilvánvaló, hogy azok döntő része befogadhatatlan, „használatlan” számunkra. A többszázas nagyságrend olykor-olykor még átnézhető, de a többezres, pláne a többmilliószámú találati oldalak listázása nyilván „tartalmatlan”, diszfunkcionális.

Másrészt nagyon sok esetben problémát jelenthet az a tény, hogy „csak” dokumentumokat kap a felhasználó (mindegy, hogy milyen nagyságrendben). Nem csak olyan információs igény létezik ugyanis, amikor egy keresőkérdésre elegendő lehet egy dokumentumlistát kapni, amely lista elemeiről tudjuk, hogy azokban majd megtalálhatjuk a minket érdeklő információt. Vannak olyan információs igényeink is, amikor konkrét kérdésre konkrét válaszokat szeretnénk kapni.

Harmadrészt a szabadszavas keresés szemantikailag üres, lapos, „buta”. Ennek megint csak az az oka, hogy ezek a keresőmotorok „csak” a dokumentumok – teljes – szövegében keresnek, de nem képesek a típusos keresésre. Márpedig a tudásunk osztályozáson, tipizáláson alapul, és ezen tudásunk szövegbe rendezése mellett létezik másfajta reprezentációs technika is.

A keresők szemantikai gyengeségére vonatkozóan érdemes néhány újabb példát bemutatni. A 'Star Wars 8' keresés esetében nyilván nagyon sok esetben az a helyzet, hogy valaki a film alapadatait szeretné megtudni, de elképzelhető olyan igény is, amely inkább a filmről szóló, egyébként hasonló című könyv adatait szeretné megtalálni. Az is lehet, hogy nem 'Jancsó Miklós'-ra általában, hanem speciális kontextusban, fókuszáltabb kérdésfeltevés mentén vagyunk kíváncsiak. Érdekelhetnének azok (és csak azok) a filmek, amelyeket Jancsó Miklós rendezett. Ilyen esetben irreleváns lehet az életrajza, bármely alapadata, a róla szóló könyv, a vele készült interjú átirata stb. Hogyan pontosíthatjuk az ilyen keresési igényeket? A keresőkérdés ('Jancsó Miklós') megadásán túl azt is elvárjuk a keresőtől, hogy Jancsó Miklósról csak a rendezői szerepek között keressen, ne bárhol, ahol előfordul a neve. Ha viszont a valaki a Jancsó Miklós által írt könyveket szeretné megtalálni, akkor másfajta típusinformációt kellene még igénybe venni ahhoz, hogy a keresőigényt ki tudjuk elégíteni. Ezt az igényt lehet jól kiszolgálni az adatbázis-alapú keresésekkel.

Az adatbázis strukturált szöveg lévén definíció szerint rendelkezik szemantikai képességekkel, mégpedig azzal, hogy a sémainformáció alapján szemantikailag jóval fókuszáltabb keresésre van mód. Az adatbázisban való keresés mindig valamilyen típusos keresést jelent, vagyis a keresőfeltételek megadás nem az egész információs tartományra (tehát nem az egész adatbázisra) vonatkozik, hanem annak a séma alapján jól leszűkíthető – jelentéssel is rendelkező – részére. Ha úgy adunk meg egy keresést, hogy a keresőfeltételt valamilyen fogalmat leíró sémakomponensre vonatkoztatjuk, akkor találatul a fogalom alá tartozó típusfogalmakra vagy előfordulásokra számíthatunk (természetesen csak akkor, ha az adatbázis a sémának megfelelően konzisztensen van felépítve).

Ha egy filmes adatbázisban keresünk, akkor megadhatjuk, hogy a keresőfeltételként megadott 'Jancsó Miklós' kérésre honnan, milyen típusú mezőkből várunk találatokat: ott keressünk, ahol a film rendezője vagy ott, ahol szereplőként látható egy filmben, vagy ott, ahol a róla szóló könyvek címeiben említik meg a nevét. Ez technikailag úgy valósítható meg, ha a keresőfeltétel mellett megadjuk a keresés helyet is (vagyis a keresés típusát).

Nyugodtan mondhatjuk, hogy az adatbázis-keresés adja a szemantikus keresés prototípusát. A szemantikai fókuszáltságnak természetesen ára van, többféle értelemben is. Először is sokkal nehezebb adatbázist, azaz strukturált szöveget építeni a szabad, strukturálatlan szöveghez képest. Minden egyes rekord felvitelek mezőről mezőre (argumentumról argumentumra) döntéseket kell hozni, hogy a beírt kívánt adat szemantikailag megfelel-e a sémában rögzített tervezői elvárásoknak. Ezt a szabad folyású szöveg esetében sosem kell vizsgálni. Másfelől a jól-felépített adatbázisban való keresés is többtinformációt kíván – mondjuk a szabadszavas kereséshez képest. Először is ismerni kell az adatbázisok szabványos keresőnyelveit (például a relációs adatbázisok lekérdezésére szolgáló sql-nyelvet). További „nehézség”, hogy csak akkor tudunk adatbázisban keresni, ha ismerjük a sémáját. Ennek hiányában ott állunk „Szezám barlangja” előtt, és nem tudunk belépni a kapuján.

3.3.1.4.2.3 QA-keresés

A **Question Answering (QA)**, magyarul a kérdés-válasz rendszerek fogalmát többféleképpen is értelmezik, de a sokféle felfogásban az közös, hogy az ilyen rendszerek a felhasználók által feltett kérdésekre nem – akármilyen hosszúságú – dokumentumokat, hanem rövid válaszokat (kifejezéseket, mondatokat) adnak vissza. A legszemléletesebb példát a Google-től hozhatjuk. A következőkben bemutatunk oldalpárokat, amelyek a Microsoft által fejlesztett Bing keresőmotor, illetve a Google-kereső találati oldalait mutatja ugyanazon kérdések esetén. A páros összehasonlítás célja és értelme az lesz, hogy látni fogjuk, hogy a Google bizonyos keresőkifejezések esetén már meghaladja a szimpla keresőmotor logikát, és a találati oldalát megosztva QA-eredményeket is szolgáltat.

Az első példa legyen az, amikor 'Jancsó Miklós' a keresőkérdés, amikor is a neves magyar filmrendezőről szeretnénk információhoz jutni. Ebben az esetben a Bing és a Google-kereső az alábbi oldalakat adja vissza.

The screenshot shows a Bing search results page for the query "jancsó miklós". The search bar at the top contains the text "jancsó miklós". Below the search bar are navigation tabs for "All", "Images", "Videos", "News", and "My saves". The main results area displays 29,600 results. The top result is "Jancsó Miklós (filmrendező) – Wikipédia" with a link to the Hungarian Wikipedia page. Below it is another Wikipedia entry for "Miklós Jancsó" in English. Further down is a result from "port.hu" titled "Jancsó Miklós - port.hu". At the bottom, there is an "Images of jancsó miklós" section with a grid of five images and a link to "Üdvözet - jancso.szote.u-szeged.hu".

keresés 'Jancsó Miklós'-ra a Bing-keresőben

A Bing a klasszikus megoldást hozza: egymás után listáz oldalakat, amelyekre „azt állítja”, hogy tartalmazzák a keresőkifejezést, a jobb oldalon pedig a korábbi keresések alapján összeállított, további keresési opciókat ajánl (related search). A Bing felmutat pár képet is, de ezzel most nem érdemes foglalkoznunk.

A Google is ugyanazt kínálja, mint a Bing – a találati oldal bal sávjában: kattintható linkek vannak felsorolva, és a felhasználó úgy juthat a releváns információhoz, ha egymás után elugrik a felkínált oldalakra.

The screenshot shows a Google search for "jancsó miklós". The search bar is at the top left. Below it are navigation tabs: "Összes", "Képek", "Videók", "Hírek", "Térkép", "Egyebek", "Beállítások", "Eszközközök". The search results on the left list several entries:

- Jancsó Miklós (filmrendező) – Wikipédia**: [https://hu.wikipedia.org/wiki/Jancsó_Miklós_\(filmrendező\)](https://hu.wikipedia.org/wiki/Jancsó_Miklós_(filmrendező)) - Jancsó Miklós (Vác, 1921. szeptember 27. – Budapest, 2014. január 31.) kétszeres Kossuth-díjas és Balázs Béla-díjas magyar filmrendező, forgatókönyvíró, ...
- Jancsó Miklós (egyértelműsítő lap) – Wikipédia**: [https://hu.wikipedia.org/wiki/Jancsó_Miklós_\(egyértelműsítő_lap\)](https://hu.wikipedia.org/wiki/Jancsó_Miklós_(egyértelműsítő_lap)) - Jancsó Miklós (1868–1930) orvos, belgyógyász; Jancsó Miklós (1903–1966) ... az összes cikk, amelyben előfordul az a szöveg, hogy Jancsó Miklós (vagy ...
- Jancsó Miklós (operatőr) – Wikipédia**: [https://hu.wikipedia.org/wiki/Jancsó_Miklós_\(operatőr\)](https://hu.wikipedia.org/wiki/Jancsó_Miklós_(operatőr)) - Szülei Jancsó Miklós filmrendező és Wowersznyi Katalin. Apai nagyapja Jancsó Sándor (1887-?), aki a székesfehérvári Hadjárva Intézet igazgatója volt.
- Jancsó Miklós - PORT.hu**: <https://port.hu/adattlap/szemely/jancso-miklos/person-793> - És akkor Jancsó Miklós megújította a magyar filmet! - Szegénylegények 1965-ot írunk, és a Cannes-i Filmfesztiválon is versenyző Szegénylegények... PORT.hu ...
- Meghalt Jancsó Miklós - Origo**: www.origo.hu/filmklub/blog/.../20140131-meghalt-jancso-miklos-filmrendezo.html - 2014. jan. 31. - Kilencvenkét éves korában elhunyt Jancsó Miklós Kossuth-díjas filmrendező, a Szegénylegények, a Csillagosok, katonák és a ...
- Jancsó Miklós fia is szerepelt A Pál utcai fiúkban - Nyika így néz ki ma ...**: <https://femina.hu> - Hazai sztár - 2017. okt. 1. - Jancsó Miklós 28 éves volt, amikor 1949. december 21-én feleségül vette Wowersznyi Katalint, aki két gyermeknek adott életet: Nyika 1952-ben ...

On the right side, there is a knowledge panel for "Jancsó Miklós" (Filmrendező). It includes a grid of photos, a "További képek" link, and biographical information:

- Jancsó Miklós** (Filmrendező)
- Jancsó Miklós kétszeres Kossuth-díjas és Balázs Béla-díjas magyar filmrendező, forgatókönyvíró, érdemes és kiváló művész. Wikipédia
- Születési dátum:** 1921. szeptember 27., Vác
- Meghalt:** 2014. január 31., Budapest
- Házastárs:** Csákány Zsuzsa (házas. 1981–2014.), TOVÁBBIAK
- Gyermekek:** Jancsó Nyika, Jancsó Dávid, Jancsó Katalin, Jancsó Zoltán
- Könyvek:** Vizi privát publicche virtú, Vitam et sanguinem, TOVÁBBIAK

Below the knowledge panel is a "Filmek" section with a grid of movie posters, including "Szegényl...", "Csillagosok, katonák és a...", "Még kér a...", "Magánbű...", and "Csend és...".

keresés 'Jancsó Miklós'-ra a Google-keresőben

A találati oldal jobb részén azonban valami más látható. Nem dokumentumok felsorolását kapjuk, hanem Jancsó Miklósról szóló közvetlen adatokat. Ez annyit jelent, hogy nem kell tovább ugranunk, már az első oldalon releváns adatokat kapunk a keresett személyről. A Google itt már nem egy dokumentumlistát, hanem a feltett kérdésre azonnali válaszokat kínál. Mi a különbség? Amíg a keresőmotor a dokumentumok listájával csak a keresett információ helyét adja meg, addig a QA-szolgáltatás magát a keresett információt.

De nem csak erről van szó! Ha megnézzük a Google-kereső dokumentumlistáját, a harmadik helyen 'Jancsó Miklós (operatőr)' című Wikipédia-oldalt látjuk. Ez az oldal nyilvánvalóan nem a filmrendezőről szól, hanem valaki másról (egyébként a rendező fiáról, aki operatőr, de ez most mindegy). Ez a mozzanat egyben jelzi azt, hogy a keresőmotorok szemantikailag gyengék, hiszen azt a világtudást, hogy ti. ebben az esetben két embernek ugyanaz a neve, nem birtokolja, és így – az eredeti kereső szándékhoz képest – irreleváns, „hamis” találatokat ad vissza.

De térjünk vissza a QA-rendszerek értékeléséhez. A Google-kereső jobb oldalán látható információs blokk felépítéséhez arra van szükség, hogy a Google felismerje (feltételezze), hogy itt QA-igényt kell (lehet) kielégítenie. Ezután tudnia kell, hogy a lehetséges 'Jancsó Miklós' nevű emberek közül vajon melyikre irányulhat a kereső szándék. Ezt nyilván a korábbi keresések története alapján dönti el. Ha sokkal többen kattintottak olyan találati linkekre, amely a rendező Jancsó Miklóssal volt kapcsolatos, akkor azt feltételezi, hogy vélhetőleg most is erről lehet szó, ezért végzi el ezt a szelektálást. E döntés

alapján pedig előkeresi azokat a – rendezőre vonatkozó – adatokat (születési, halálozási adatok, foglalkozás, képek, munkásság, életút stb.), amelyeket szöveges dokumentumokból, adatbázisokból szedett ki korábban.

A következő páros oldalakon ugyanezt a kettősséget láthatjuk. A 'Star Wars 8' keresésre a Bing találati listát és kapcsolódó keresési lehetőségeket ad vissza, míg a Google a QA-blokkban felsorolja a film alapadatait.

The screenshot shows a Bing search results page for the query "star wars 8". The search bar at the top contains the text "star wars 8" and a magnifying glass icon. To the right of the search bar, there are links for "Sign in" and a menu icon. Below the search bar, there are tabs for "All", "Images", "Videos", "News", and "My saves". The "All" tab is selected. The search results are displayed in a grid format. The first result is "Star Wars: The Last Jedi - Wikipedia" with a link to the Wikipedia page. The second result is "StarWars.com | The Official Star Wars Website" with a link to the official website. The third result is "Star Wars 8 – Az utolsó Jedik | Filmek" with a link to a Hungarian film website. The fourth result is "News about Star Wars 8" with a link to Bing News. A small image of Han Solo is shown next to the news result. To the right of the search results, there is a "Related searches" section with several suggestions: "star wars 8 rumors and spoilers", "star wars episode 8 spoilers", "star wars episode 8 plot rumors", "star wars 8 news spoilers", "star wars episode 8 leaks", "star wars 8 plot theories", "star wars episode 8 latest news", and "star wars episode 9 plot".

keresés 'star wars 8'- ra a Bing-keresőben

Google search results for "star wars 8". The search bar shows "star wars 8" and the results include:

- Star Wars: Az utolsó Jedik - PORT.hu**
 - URL: <https://port.hu/adattap/film/mozi/star-wars-az-utolso...star-wars-the.../movie-178442>
 - Értékelés: 6,2/10 - 492 szavazat
 - PORT.hu 2018. március 8.: Sorra kerülnek elő a Star Wars 8 jobbnál jobb, kivágott jelenetei Luke Hant gyászolta, Pasma máshogy halt meg Az utolsó Jedikben ...
- Star Wars!! VIII.-RÉSZ - Videó**
 - URL: <https://videa.hu/videoek/.../star-wars-viii.-rezs-amscl-DsBJSK11hsY19t...>
 - 2017. dec. 18.
 - A(z) "Star Wars!! VIII.-RÉSZ !! Am,sci-fi(szink)!!" című videót "Szabó István 1" nevű felhasználó töltötte ...
- Csillagok háborúja VIII: Az utolsó Jedik – Wikipédia**
 - URL: https://hu.wikipedia.org/wiki/Csillagok_haboruja_VIII:_Az_utolso_Jedik
 - Poe nem ért egyet Holdo titkolózó vezetői stílusával, ezért Finn-nei, BB-8-el A többi Star Wars filmben feltűnő Warwick Davis most Wodbin szerepét játssza.
 - Szereplők · Gyártás és forgatás · Bemutató · Fogadtatás
- Amit tudni érdemes a Star Wars 8 előtt, de nincs benne a filmekben**
 - URL: <https://www.pcguru.hu/hirek/amt-tudni-erdemes-a-star-wars-8-elott-de.../43546>
 - Szeretnéd tudni, hogyan alakult meg az Első Rend? Miért hozta létre Leia az Ellenállást? Mit csinált Luke az endori csata után? Itt megleled a választ.
- hvg.hu - Star Wars 8**
 - URL: https://hvg.hu/cimke/Star_Wars_8
 - Míg a világ legtöbb országában egyértelműen kasszasiker az új Star Wars-film, addig Kínában lényegesen kevesebben váltottak rá jegyet. Azért pár milliót innen ...
- star wars 8 | 24.hu**
 - URL: <https://24.hu/tag/star-wars-8/>
 - A Star Wars VIII. rész sokkal inkább a csontok a csont trilógiától, mint az ébredő Erő, Luke Skywalker ...

On the right side, there is a featured video for "Csillagok háborúja VIII: Az utolsó Jedik" with a 74% rating and a description of the film's release date (December 13, 2017) and director (Rian Johnson).

keresés 'star wars 8'-ra a Google-keresőben

Amikor 'Donald Trump', amerikai elnökre keresünk, akkor a Bing listákat ad, a Google listát és QA-szolgáltatást.

Bing search results for "donald trump". The search bar shows "donald trump" and the results include:

- News about Donald Trump**
 - bing.com/news
 - Moon heads to US amid fears for Trump-Kim summit - cnn.com**
 - CNN · 9m
 - Moon will meet with Donald Trump Tuesday, ahead of the US President's planned summit with North Korean leader Kim Jong Un in Singapore next month.
 - AP FACT CHECK: Trump on Russia probe, border; Pruitt ...**
 - Chicago Tribune · 1h
 - President Donald Trump is skimming over the facts involving the investigations into Russian meddling in the 2016 election.
 - US relations under Trump dominate Mexico presidential ...**
 - KSL · 53m
 - Mexico's turbulent relationship with the U.S. government under President Donald Trump dominated the country's second presidential debate Sunday.
- Donald J. Trump - Official Site**
 - URL: <https://www.donaldtrump.com>
 - Help continue our promise to Make America Great Again!
 - Contact · Gallery · Official Trump Store · News · Donate · About
- Donald J. Trump (@realDonaldTrump) | Twitter**
 - URL: <https://twitter.com/realdonaldtrump>
 - The latest Tweets from Donald J. Trump (@realDonaldTrump). 45th President of the United States of America 🇺🇸. Washington, DC
 - Account Status: Verified
- trump donald .org - Donald J Trump**

On the right side, there is a "Related searches" section with links like "trump update today", "email address for president trump", "trump latest breaking news", "trump schedule today", "did donald trump assault women", "trump blog sites", "donald trump and ryan white", and "trumps platform 2016".


keresés 'donald trump'-ra a Bing-keresőben

Google donald trump

Összes Képek Videók Hírek Térkép Egyebek Beállítások Eszközök


Nagyjából 26 600 000 találat (0,62 másodperc)

Vezető hírek




Trump kivizsgálhatja, volt-e beépített FBI-ügynök a választási kampánycsapatában

24.hu
1 órája



Donald Trump nem tudta helyesen leírni a saját felesége nevét

Vélvet
17 órája



Az FBI informátort küldhetett a Trump-kampány tagjaira


444
9 órája

→ További találatok erre: donald trump

Donald Trump – Wikipédia
https://hu.wikipedia.org/wiki/Donald_Trump
 Donald John Trump (New York, 1946. június 14. –) amerikai üzletember, politikus és médiaszemélyiség; 2017 óta az Amerikai Egyesült Államok elnöke.
 Melania Trump · Ivana Trump · Barron Trump · Barack Obama

Donald Trump - Wikipedia
https://en.wikipedia.org/wiki/Donald_Trump
 Donald John Trump (born June 14, 1946) is the 45th and current President of the United States, in office since January 20, 2017. Before entering politics, he was ...

Donald J. Trump (@realDonaldTrump) | Twitter



Donald Trump
 Az Amerikai Egyesült Államok elnöke

Donald John Trump amerikai üzletember, politikus és médiaszemélyiség; 2017 óta az Amerikai Egyesült Államok elnöke. Karrierjét a The Trump Organization vállalat elnökeként építette fel. Wikipédia

Házastárs: Melania Trump (házas. 2005.), Marla Maples (házas. 1993.–1999.), Ivana Trump (házas. 1977.–1992.) Felkapott

Születési dátum: 1946. június 14. (életkor 71 év), Jamaica Hospital Medical Center, New York, Egyesült Államok


Magasság: 1,9 m

Nettó vagyon: 3,1 milliárd USD (2018.) Forbes

Párt: Republikánus Párt

Gyermekek: Ivanka Trump, Donald Trump, Barron Trump, Tiffany Trump, Eric Trump

A felhasználók ezeket keresték még Még 15+



keresés 'donald trump'-ra a Google-keresőben

A 'Balaton' kifejezésre keresve is hasonló különbséget tapasztalhatunk:

Bing budapest

All Images Videos News | My saves


16,200,000 Results

Budapest Travel & Tourism Guide - Tourist information of ...
<https://www.budapest.com>
 Budapest tourist information – detailed info about accommodation, sights, travel, things to do, spas, recreation and more... by BUDAPEST.COM
 Things to Do · City Guide · Budapest Sightseeing Tour · Tourist Information · Travel · Budapest Map

Budapest - Wikipedia
<https://en.wikipedia.org/wiki/Budapest>
 Etymology "Budapest" is the combination of the cities Buda and Pest, which, together with Óbuda, united in 1873. An early documented occurrence of the combined name "Buda-Pest" was in the 1831 book Világ (World/Light) by Count István Széchenyi.
Country: Hungary **Region:** Central Hungary
Area code: 1 **Elevation:** Lowest (Danube) 96 m, Highest (Jáno...
 Liberty Statue · House of Terror · Museum of Fine Arts · Matthias Church

Budapest portál | Budapest Portál
budapest.hu
 Budapest Főváros Önkormányzata számára kiemelten fontos a versenyképes tudással rendelkező hallgatók képzésének és ezáltal a humán erőforrás fejlesztésének támogatása, ezért az előző évek sikerei után idén is meghirdeti a Budapest Ösztöndíj Program pályázatát a fővárosi székhelyen vagy telephelyen működő ...

News about Budapest
bing.com/news



9 hidden places in Budapest worth discovering - KTVZ
 KTVZ · 1d
 Millions of travelers head to Budapest each year ready to be amazed by its stunning skylines, brilliant...

Related searches

- budapest hungary travel warnings
- budapest tourism official site
- is budapest safe for americans
- budapest airport
- what to see in budapest
- top things to do in budapest hungary
- tourist attractions in budapest hungary
- budapest tripadvisor

keresés 'budapest'-re a Bing-keresőben

The screenshot shows a Google search for "budapest". The search bar at the top contains the word "budapest". Below the search bar, there are navigation tabs: "Összes", "Járatok", "Térkép", "Képek", "Hírek", "Egyebek", "Beállítások", and "Eszközök". The search results include a Wikipedia entry for Budapest, a "Budapest portál" (portal) with a link to "budapest.hu", and a "Vezető hírek" (top news) section with three featured articles. On the right side, there is a map of Budapest and a detailed information panel for Budapest, Hungary, including its area, climate, and travel information.

keresés 'budapest'-re a Google-keresőben

Az oldalak összehasonlítása alapján látszik egyfelől a két kereső közti különbség, másfelől a QA-szolgáltatás eltérő minősége. A szimpla keresőmotor szemantikai ürességére már utaltunk: ezek a keresők nem tudják elkülönítve kezelni sem a megadott kifejezés eltérő jelentéseit, sem azok eltérő típusba sorolásait. A hivatkozott példánk azt szemléltette, hogy a kereső nem tudja megkülönböztetni a „két” Jancsó Miklóst egymástól. A keresőmotor csak a karakterazonosságra figyel, de az azonos karakterekkel jelölt különböző dolgokat képtelen elhatárolni egymástól. A QA-rendszer már okosabb, hiszen ki tudja választani a legvalószínűbb megoldást, ami arra utal, hogy valamilyen mértékben már tudja tipizálni a keresőkifejezést (a sok opcióból kiválasztja az egyiket). Természetesen előfordulhat, hogy a legvalószínűbb opció kiemelése nem lesz kielégítő bizonyos felhasználók számára: aki például az operátor Jancsó Miklósról keresett rá, annak nyilván nem hasznos a QA-blokkba kiemelt információ. A Balaton keresés esetében is tudunk ilyen „ellenpéldát” mutatni: a Google QA erre a kérdésre a népszerű magyarországi tó alapadatait gyűjti ki, ami az esetek döntő többségében megfelelhet a keresők elvárásainak, de előfordulhatnak olyan keresési igények is, amelyek a Balaton nevű magyarországi településre irányulnak, és ezek számára irreleváns minden, a tóra vonatkozó adat.

A Google kérdés-válasz rendszere bizonyos esetekben már többszavas, „teljes mondatos” kérdést is felismer, és visszaadja a megfelelő választ (angolul). Nézzünk meg két példát. Ha azt kérdezzük meg, hogy 'ki volt az USA elnöke 1980-ban' ('who was the us president in 1980'), akkor a találati oldal tetején megkapjuk a választ: 'Jimmy Carter'.

Nagyjából 104 000 000 találat (0,58 másodperc)

Az Amerikai Egyesült Államok elnöke (1980.)

Jimmy Carter



Jimmy Carter további adatai

Visszajelzés

United States presidential election, 1980 - Wikipedia, the free ...

https://en.wikipedia.org/.../United_States_presidential_election,_1980 ▼ Oldal lefordítása

The **United States presidential election of 1980** was the 49th quadrennial **presidential election**. ...

Carter's critics saw him as an inept leader who had failed to solve the worsening economic problems at home. His supporters defended the ...

1984 · 1976 · John B. Anderson · Reagan Era

Historical polling for U.S. Presidential elections - Wikipedia, the free ...

https://en.wikipedia.org/.../Historical_polling_for_U.S._Presidential_e... ▼ Oldal lefordítása

Ugrás a(z) **United States presidential election, 1980** részhez - Incumbent **President Jimmy Carter** ... of Americans who saw Carter as a lesser evil ...

United States presidential election of 1980 | United States government ...

<https://www.britannica.com/.../United-States-presidential-election-of-...> ▼ Oldal lefordítása

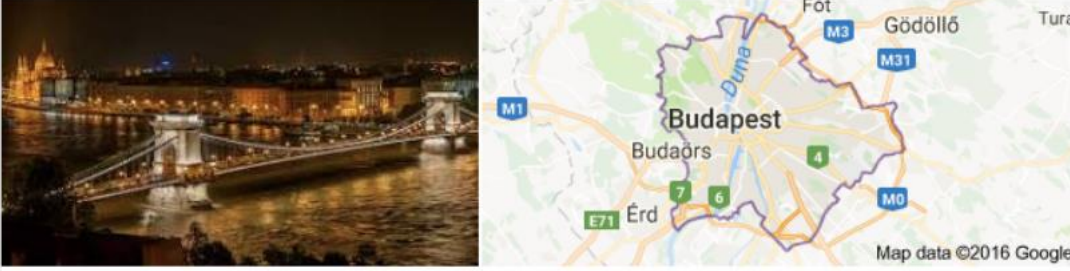
United States presidential election of 1980, American presidential election held on ... Reagan's most serious opposition came from Bush, who won support from ...

kérdés: 'who is the us president in 1980' – válasz: 'Jimmy Carter'

Ha azt a kérdést tesszük fel, hogy 'mi Magyarország fővárosa' ('capital of hungary'), akkor a kereső felelete 'Budapest' lesz.

Nagyjából 73 600 000 találat (0,52 másodperc)

Magyarország / Főváros



Budapest

Közelgő események és érdekességek

Visszajelzés

Budapest - Wikipedia

<https://en.wikipedia.org/wiki/Budapest> ▾ Oldal lefordítása

Budapest is the capital and the largest city of Hungary, and one of the largest cities in the European Union. It is the country's principal political, cultural, ...

Buda Castle · Hungarian Parliament Building · Pest · Budapest (disambiguation)

List of historical capitals of Hungary - Wikipedia, the free encyclopedia

https://en.wikipedia.org/wiki/List_of_historical_capitals_of_Hungary ▾ Oldal lefordítása

The European country of Hungary has had more than one capital city in its history. These were: ... changed the capital title with Esztergom (one of the centres of the Principality of Hungary until Stephen's crowning. ... (History of Hungary from the prehistory to 2000), Pannonica Kiadó, Budapest,

kérdés: 'capital of hungary' – válasz: 'Budapest'

Meg kell még jegyeznünk itt azt, hogy a QA-rendszereknek van egy másik fontos minősége, ami a keresés egy másfajta problémájára kínál valamilyen megoldást. Ezek a rendszerek nemcsak abban térnek el a hagyományos keresőktől, hogy dokumentumok helyett mondatokat adnak vissza, de egy első lépést jelentenek a természetes nyelven zajló, ember és gép közt megvalósuló kommunikáció irányába is. A 'who is the us president in 1980' keresőkérdés már természetes nyelven feltett „kérdőmondat”, és az erre adott (egyébként magyar nyelvű) válasz is – kicsit suta – mondatként jelenik meg: 'Amerikai Egyesült Államok/Elnök (1980.): Jimmy Carter'.

A természetes nyelvű igény nem csak a QA-rendszerek esetében jelentkezik, ugyanolyan erős az elvárás, hogy az adatbázisok lekérdezésekor is tudjunk természetes nyelven kommunikálni a géppel, de ettől még elég távol vagyunk.

3.3.1.4.2.4 KOS-alapú keresés

A tudásszervezési rendszerekben is nyelvi kifejezések szerepelnek. Ebben a minőségükben megegyeznek a szöveggel, adatbázissal. Abban különböznek a szövegtől, hogy a tudásszervezési rendszerekben nem szintagmatikailag teljes mondatok, hanem csak a legkisebb nyelvi egységek, a szavak (esetleg kifejezések) szerepelnek bennük. A tudásszervezési rendszerek a nyelv alkotóelemei, a szavak (esetleg kifejezések) és a köztük felvehető kapcsolatok együttesével mindig valamilyen nyelvi struktúrát alkotnak. Ebben az értelemben a tudásszervezési rendszert lehet egy speciális adatbázisnak

tekinteni, aminek a tudásszervezési rendszer típusát jellemző struktúra modelljét értelmezhetjük az adatbázis sémájaként.

A tudásszervezési rendszerek alkalmazásának célja és értelme valamely ismeretterület tartalmi (és formai) leírása. Ez annyit jelent, hogy első körben a tartalmi és formai leíráshoz használható tárgyszavakat,³⁶ illetve az köztük felállítható relációkat rögzítik (a rendszerbe építik), majd a második körben a tárgyszavakat valamely dokumentumgyűjtemény elemeihez rendelik. Mindez arra jó, hogy a tárgyszavak invertált indexként működve kereshetővé teszik magukat a dokumentumokat. Ez akkor értelmes lehet, amikor a dokumentumokban tartalmában is lehet keresni szabadszavas technika alkalmazásával, mivel a tárgyszavak révén fókuszáltabb keresést lehet végrehajtani. Ez a keresés pontosságát csak akkor növelheti, ha a dokumentumokhoz rendelt tárgyszavak jól reprezentálják a dokumentum tartalmát (és formai jellemzőit). Azon dokumentumtípusok esetében azonban, amelyek nem „nyitottak” a szabadszavas keresés felé, még inkább hasznos lehet a tárgyszó-alapú keresés, hiszen nincs más mód a nyelvi keresésre.

A KOS-hoz kapcsolható keresésnek kétféle értelmet is adhatunk. Egyfelől – ha egyszer elfogadjuk, hogy a KOS-ok is adatbázisoknak tekinthetők, akkor – kereshetünk a KOS tartalmában mint adatbázisban. Ekkor a KOS-keresés a KOS-on belüli keresést jelenti. Ennek eredményeként megjeleníthetjük a KOS-ban tárolt tudást, vagyis azt, hogy a keresett terminusnak, címkének milyen kapcsolatai vannak az adott KOS-ban kezelt relációk mentén. Ekkor a KOS-ban tárolt, világról való tudást mozgósíthatjuk.

A KOS-alapú keresés másik értelme az lehet, hogy a KOS-ban tárolt világtudást arra használjuk, hogy valamilyen más információs tartományban való keresés relevanciáját növeljük. Ehhez szükség van arra, hogy az információs tartomány komponenseihez hozzá legyenek rendelve a dokumentumok valamilyen tulajdonságát leíró KOS-tárgyszavak.

Ebben az értelemben a KOS-alapú keresés lényege az, hogy a KOS-ban tárolt tárgyszavak keresőkifejezésekké válnak, és – jó esetben – segítik a relevánsabb keresést egyfelől, illetve képesek arra is, hogy adott tárgyszó helyett ajánljanak más terminusokat másfelől, ami nyilván akkor hasznos és értelmes, ha az „eredeti” keresőkifejezés esetén kevés találatot kapunk (vagy esetleg semmit sem kapunk). Vizsgáljunk meg ezt az opciót néhány példán keresztül!

szezám barlangja kapcsolódó keresései

[szezám tárulj jelentése](#)

[szezám jelentése](#)

[szezám tárulj eredete](#)

[szezám tárulj angolul](#)

Google-kereső további találati ajánlata a 'szezám barlangja' kifejezésre

³⁶ Technikai értelemben másodlagos, hogy milyen terminusokat használunk a tudásszervezési rendszerek elemeire. Különböző szakmai területeken más és más a preferált terminus. Számunkra a 'tárgyszó', a 'kulcsszó', a 'címké', a 'tag', a 'lexikai egység', a 'taxon' stb. terminusok mind ugyanúgy alkalmazhatók a KOS-elemek jelzésére.

kutya kapcsolódó keresései

| | |
|--------------------------------------|---------------------------------------|
| kutya tulajdonságai | ingyen elvihető kutya |
| kutya jófogás | kutya képek |
| kutya eladó | kutya fajták |
| kutya testfelépítése | kutya betegségek |

Google-kereső további találati ajánlata a 'kutya' kifejezésre

ebtenyésztő kapcsolódó keresései

[magyar ebtenyésztők országos egyesülete](#)
[meoe törzskönyvezés](#)
[meoe törzskönyvezési osztály](#)
[meoe kutyakiállítás 2018](#)
[meoesz kiállítások 2018](#)
[magyar ebtenyésztők országos egyesülete francia bulldog](#)
[meoesz törzskönyvezési osztály](#)
[kennelnév kiváltás](#)

Google-kereső további találati ajánlata az 'ebtenyésztő' kifejezésre

Mindhárom esetben a találati oldal alján lehet olvasni azt az ajánlatot, amit a keresőmotor kínál a keresőkifejezéshez kapcsolva azzal, hogy a linkekre kattintva tovább találatokat kaphat a felhasználó. A Google itt nyilvánvalóan valamilyen tudásbázis alapján dönti el, hogy milyen javaslatokat tegyen. A fenti példákban láthatóan a korábbi gyakori keresések alapján ad ajánlatot. Azokat a gyakran feltett kérdéseket „ismétli” meg újra, amelyek valahogyan tartalmazzák a konkrét keresőkifejezést is. Ez a folkszónómiákhoz hasonló logika, hiszen a felhasználók által korábban a „rendszerbe tett” információt hasznosítja, és a sok egyéni felhasználói hozzájárulásból gyakorisági alapon választja ki a legrelevánsabbnak tűnő kifejezéseket.

Azt is észre lehet venni, hogy a keresőmotor bizonyos értelemben képes a szinonimitás kezelésére is. Ez ott érhető tetten, amikor az 'eb' kifejezés mellett megjelenik a 'kutya' terminus is az ajánlatok között (a harmadik példában). Ez persze még magyarázható lenne azzal is, hogy a korábbi keresések során gyakran összekapcsolódott a konkrét kereséssel a 'kutyakiállítás' kifejezés, és csak ezért rakta ki a keresőmotor az utóbbit. Ám a következő példában már nyilvánvalóbb, hogy szinonimakezelésről van szó. Az első ábra azt mutatja, hogy milyen találatokat ad vissza a kereső, a második ábrán pedig az látható, hogy milyen további keresőkérdéseket javasol.



magyarország kormányfői



Összes Térkép Képek Videók Hírek Egyebek Beállítások Eszközök

Nagyjából 194 000 találat (0,42 másodperc)

Magyarország kormányfőinek listája – Wikipédia

https://hu.wikipedia.org/wiki/Magyarország_kormányfőinek_listája ▼

Ugrás a(z) **A Magyar Királyság kormányfői** az 1848–49-es forradalom és ... részhez - az első felelős magyar kormány (minisztérium).

Magyarország államfőinek ... · Magyarország Kormánya · Miniszterelnököt

Kategória:Magyarország kormányfői – Wikipédia

https://hu.wikipedia.org/wiki/Kategória:Magyarország_kormányfői ▼

A(z) „Magyarország kormányfői” kategóriába tartozó lapok. A következő 4 lap található a kategóriában, összesen 4 lapból.

Magyarország eddigi miniszterelnökei | tortenelemcikkek.hu

tortenelemcikkek.hu/node/201 ▼

2015. aug. 2. - Magyarország miniszterelnökeinek listája. ... Cikkünk Magyarország miniszterelnökeinek (azaz kormányfőinek) ... **Magyarország kormányfői** ...

Magyar miniszterelnökök a rendszerváltás óta - Terasz | Femina

<https://femina.hu/Terasz> ▼

Magyar miniszterelnökök a rendszerváltás óta - Képgalériánkban Antaltól Orbánig megnézheted a miniszterelnököket. A galéria végén pedig az új kihívók.

Magyarország miniszterelnökei

www.altisk-lelle.sulinet.hu/toriweb/egyeb/Magyarorszagminiszterelnokei.htm ▼

Magyarország miniszterelnökei. 1. Batthyány Lajos 1848. III. 17. - 1848. X. 2. 2. Szemere Bertalan

a Google-kereső találati listája a 'magyarország kormányfői' kifejezésre

magyarország kormányfői kapcsolódó keresései

magyar miniszterelnökök a 20. században

magyarország kormányai 1989 után

magyarország miniszterelnöke 2018

magyarország miniszterelnöke 2017

magyarország államfői

magyarország miniszterelnökei 2017

magyarország kormányzói

miniszterelnök választása

a Google-kereső további találati ajánlata a 'magyarország kormányfői' kifejezésre

Mind a találatok között, mind az ajánlott kifejezések listájában felbukkan a 'miniszterelnök' kifejezés, ami szinonimája a kormányfőnek, és a keresőnek ezt már nyilván tudnia kell, ha ilyen találatokat, illetve ajánlatokat tesz ki az oldalra.

Érdeemes megjegyezni, hogy a Google korábban másfajta logikájú ajánlatokat is tett. Volt, hogy tezaurusokban tárolt tudást mozgósított, és felkínált a keresőkifejezéshez kapcsolható általánosabb vagy speciálisabb jelentésű fogalmakat. Ezt a lehetőséget már jó ideje nem használja ki, aminek nyilván az lehet az oka, hogy a Google által kezelt hatalmas szövegtárban való keresést sokkal

hatékonyabban lehet támogatni a keresők által megfogalmazott kérdésekkel, mint a szemantikailag releváns pontosító vagy tágító fogalmakkal. A társas navigáció erősebb a szemantikus navigáció által kínált lehetőségeknél. Ennek pedig az az oka, hogy a strukturálatlan szövegtengerben nem igazán segítenek a szemantikai finomítások, hiszen a szövegben nem lehet hatékonyan hasznukat venni, viszont a korábbi keresések gyakorisága fontos relevanciaképző információ. Ha valamit sokan kerestek korábban, valószínűleg sokan akarják keresni a jelenben is.

Ez azonban nem jelenti azt, hogy más típusú információk tartományok esetében ne lenne hasznos valamilyen szemantikai támogatás. Mutatunk néhány példát a névterek világából arra, hogy adatbázisok esetén milyen típusú segítséget kínálhat a tudásszervezési rendszerekben tárolt szemantikai információ. A személynévtérben értelmes lehet összegyűjteni azokat az életrajzi adatokat, amelyeket a különböző könyvekben találhatunk meg adott személyekről. Ha tartjuk magunkat ahhoz az alapelvhez, hogy a névtérbe felveszünk minden adatot, amit a források tartalmaznak, ráadásul ezt úgy tesszük, hogy pontosan azt a nyelvi változatot rögzítjük, amelyik a forrásban volt, akkor ezen igény kielégítéséből több minden következik.

Ilyen esetekben gyakran előfordulhat az, hogy az egyes forrásokban ugyanazt az információt változó módon, szinonim alakban adják meg. Például a személyek foglalkozási adatai között az egyik forrás megadhatja egy politikusról, hogy 'miniszterelnök' volt egy adott időszakban, míg egy másik forrás ugyanerre vonatkozóan mondhatja azt, hogy a személy 'kormányfő' volt. Ebben az esetben nyilván nem értelmes megjelölni mindkét adatot foglalkozásként, hiszen tartalmilag ugyanazt mondják el a személyről. Ezt a problémát úgy tudjuk jól kezelni, ha egy tudásszervezési rendszerben megadjuk, hogy a két kifejezés (a 'kormányfő' és a 'miniszterelnök') ugyanannak a foglalkozásnak két nyelvi leírását jelenti, vagyis a KOS-unkbán kezelni kell tudni a kifejezések közti szinonimitást.

Még mindig a foglalkozások példájánál maradva előfordulhatnak olyan esetek is, amikor nem elég a szinonima reláció alkalmazása. Van úgy, hogy egy orvosról megadják, hogy mi pontos szakterülete: az egyik forrásban azt rögzítik, hogy 'fül-orr-gégeorvos', a másikban pedig azt, hogy „fül orvos” és 'gégegyógyász', a harmadikban meg azt, hogy 'gégész'. Vannak tehát különböző foglalkozásaink:

fül-orr-gégeorvos

fül orvos és gégegyógyász

gégész

és kérdés, vajon mit kezdhetünk ezzel a jelenséggel. Az nyilvánvaló, hogy a 'gégész', a 'gégegyógyász' és a 'gégeorvos' kifejezések közé fel kell vennünk a szinonima relációt, és ha valaki 'gégeorvos'-ra keres, akkor is vissza kell adni egy személyt, ha hozzá a 'gégész' vagy a 'gégegyógyász' kifejezés volt rendelve. Ehhez meg kell adnunk egy KOS-ban, hogy a három terminus jelentése megegyezik (tehát egymás szinonimái). Ki kell azonban elégítenünk azt az igényt is, amikor valaki egy általánosabb fogalomra keres (mondjuk: a 'fül-orr-gégeorvos'-ra), és ekkor nyilván elvárhatja, hogy akkor is találatot kapjon, ha egy személy foglalkozásaként „csak” a 'gégész' volt megadva. Ezt úgy lehet kezelni, ha a KOS-ban a 'generikus alárendeltje reláció' mentén összekapcsoljuk az általánosabb és speciálisabb jelentésű fogalmakat arra a világtudásra építve, hogy:

minden gégész IS_A_TPYE_OF fül-orr-gégeorvos

minden gégeorvos IS_A_TPYE_OF fül-orr-gégeorvos

minden gégegyógyász IS_A_TPYE_OF fül-orr-gégeorvos

minden fül orvos IS_A_TPYE_OF fül-orr-gégeorvos

minden fül- és gégeorvos IS_A_TPYE_OF fül-orr-gégeorvos

A fenti „képletek” alapján levonhatjuk azt a következtetést, hogy ha valaki az általánosabb kifejezésre keres, akkor találatot kell kapnia a speciálisabb kifejezésekkel leírt entitások esetén is (az összes fenti példa esetében).

Hasonló megoldással lehet kezelni azt a problémát, amikor a földrajzi helyek esetén szeretnénk megkapni a magyarországi települések listáját úgy, hogy a földrajzi helyek között a településeket különböző típusok sorolhatjuk be, úgymint:

főváros

megyei jogú város

város

nagyközség

község

Minden településhez – egy időben – egy minősítés tartozhat a fentiek közül, de a településlistát csak azon világtudás alapján tudjuk megadni, hogy egy KOS-ban felvesszük azt, hogy a fenti kategóriák mind a ’település’ kategória alá tartoznak fogalmilag.

minden főváros IS_A_TPYE_OF település

minden megyei jogú város IS_A_TPYE_OF település

minden város IS_A_TPYE_OF település

minden nagyközség IS_A_TPYE_OF település

minden község IS_A_TPYE_OF település

A generikus relációhoz hasonló módon hasznosítható a KOS-ban tárolt tudás akkor is, amikor az elemeket a partitív relációval kapcsoljuk össze egymással, igaz, ekkor nagyon figyelni kell arra, hogy a különböző minőségű partitív relációkat ne keverjük össze egymással. A névtéren belül ugyanis többféle értelemben is meg kell különböztetni a partitív reláció altípusait egymástól, és különböző szinteken, különböző entitások között más és más relációtípust kell alkalmaznunk (erről bővebben írunk a térérzékeny keresés lehetőségeit bemutató kutatási jelentésünkben). Nézzünk meg erre is pár példát! Legyen először az a feladat, hogy a földrajzi névtéren belül keressünk az ’Öreghegy’ nevű földrajzi helyekre. Ekkor sok találatot kapunk, mert sok településen belül van ilyen nevű földrajzi egység. A sok találat között azonban lehet olyan szűrési feltételt megfogalmazni (pontosabban fel lehet ennek lehetőségét kínálni a felhasználók számára), hogy csak bizonyos földrajzi téren belüli találatok érdekelnek minket. Mondjuk egy adott megyén belüli ’Öreghegy’-eket keressük. Nézzük az alábbi találati halmazt:

| KERESÉS | |
|---|-------|
| Alsóőreghegy ➔ Föld Európa Magyarország Tolna megye Paksi járás Dunaföldvár Dunaföldvár Alsóőreghegy Mgl. | 14107 |
| Alsóőreghegy vasúti megállóhely ➔ Föld Európa Magyarország Pest megye Gödöllői járás Gödöllő Gödöllő Alsóőreghegy vasúti megállóhely megszűnt vasúti megálló | 89476 |
| Felsőőreghegy ➔ Föld Európa Magyarország Tolna megye Paksi járás Dunaföldvár Dunaföldvár Felsőőreghegy költerület | 38460 |
| Kültelek (Őreghegy) Őreghegy ➔ Föld Európa Magyarország Baranya megye Pécsi járás Abaliget Abaliget Kültelek (Őreghegy) Lh. | 3539 |
| Őreghegy ➔ Föld Európa Magyarország Somogy megye Siófoki járás Balatonendréd Balatonendréd Őreghegy Mgl. | 3970 |
| Őreghegy ➔ Föld Európa Magyarország Baranya megye Pécsi járás Berkesd Berkesd Őreghegy Lh. | 4609 |
| Őreghegy ➔ Föld Európa Magyarország Baranya megye Komlói járás Komló Komló Őreghegy Üd. | 4927 |
| Őreghegy ➔ Föld Európa Magyarország Baranya megye Pécsi járás Orfű Orfű Őreghegy Üd. | 5151 |

találati lista a földrajzi névtérben

A találati listáról készített képernyőkép most „kevés” elemet mutat, a 80-ból csak 8-at, de már itt is látszik, hogy olykor értelmes lehet szűkíteni a találati eredményt (a nyolcvanas listát egyben látva pedig még erősebb lehet az igény a szűrésre). Ekkor lehet megadni azt a további feltételt, hogy csak azok az elemek érdekelnek minket, amelyek adott lokalitáson belül, mondjuk 'Baranya megyé'-n belül helyezkednek el. Ezt az információt ki lehet nyerni a névtér adatai közül, és a megfelelő feltételeket megadva szűrni lehet a találati eredményeken.

Hasonló feladat adódna abban az esetben, ha azt kérnénk, hogy olyan folyóvizet keressünk, amelyek adott lokalitáson belül vannak (adott megyén, településen, településrészen keresztül folynak át).

Arra is mód nyílik, hogy a keresés során „keverjük” a generikus és partitív relációkat, vagyis egyszerre vegyük fel szűrés feltételként a két reláció mentén mozgósítható világtudásunkat. Példa lehet erre az a keresés, amikor adott lokalitáson belül található (partitív szűrés), adott típusba tartozó (generikus szűrés) földrajzi helyeket keressünk. A fenti 'Őreghegy'-es példában a 'Baranya megye'-i partitív szűrés feltételt kiegészíthetjük azzal, hogy csak olyan találatokat várunk el, amelyek adott típusba (mondjuk 'üdülőterület') tartoznak.

Hasonló technikákat alkalmazhatunk az „időbeli keresés” esetén is. A névterekben kezelt adatoknak ugyanis sok esetben vannak, lehetnek időbeli paraméterei, és ezek az adatok ugyanúgy használhatók szűrésre, mint a térbeli koordináták. Például megadhatjuk azt szűrés feltételként, hogy milyen időszakon belül akarunk megtalálni egy adott nevű személyt (mondjuk 'Kovács István'-t). Ezt a születési és halálozási időpontok ismeretében nyilván könnyen megtehetjük, amit kikereshetünk a személynévtér adatbázisából, de olykor használhatunk erre KOS-ban tárolt tudást is, amikor eseményeket rendezünk el valamilyen idősorban. Megadhatjuk azt a feltételt is, hogy 'Kovács István'-okból csak azok érdekelnek minket, akik a 'reformkorban' éltek, és ezt a kérést csak akkor tudjuk teljesíteni, ha valahol (egy KOS-ban) eltároljuk azt a világtudást, hogy a 'reformkor' hol helyezkedik el az időtengelyen.

3.3.1.4.3 Társas navigáció

Amikor egy felhasználó talál egy számára érdekes weboldalt, akkor az oldal címét felveheti a könyvjelzői közé, de elküldheti e-mailben ismerőseinek, barátainak, munkatársainak is. Ha ezt megteszi,

akkor mások számára – akaratlanul – egy olyan navigációs és szűrési lehetőséget biztosít, amely „csak” az érintett közösség tagjai számára áll rendelkezésre. Ez az egyszerű eset szép példát szolgáltat arra a jelenségre, amelyet **társas navigációnak** vagy közösségi navigációnak (social navigation) neveznek. Ezzel a fogalommal azt a fajta eligazodást ragadhatjuk meg, amely során az egyén mások tevékenységét, mások választásait, mások értékeit követi, amikor valaki saját cselekvését a közösség többi tagjának cselekvéseihez igazítva képes eligazodni egy számára ismeretlen térben.

Ez a tér egyaránt lehet valóságos vagy szimbolikus-virtuális. Nézzünk további két példát a társas navigációra – egyelőre a valóságos tér világából. Ha valaki először száll le egy számára ismeretlen repülőtéren, akkor természetesen nem tudja, hol kaphatja meg a csomagját. Mégis gyorsan odatalál a csomagkiadó-szalagokhoz. Ehhez nem kell egyebet csinálnia, mint követni a gépről leszálló utastársait, hiszen ők is ugyanoda igyekeznek, mint az idegen utas, és vélhetőleg ők azt is tudják, hol van a keresett terem. A navigációs probléma megoldását itt a többiek, az idegen utas társai „segítik elő”.

Amennyiben az utazó a szállodában tanácsot kér a helybeliektől, hogy hol talál a saját lakhelyére jellemző, „igazi hazai ízeket”, akkor a helyiek nem biztos, hogy tudnak segíteni neki (sőt, inkább az a valószínű, hogy nem). De ha az utazó a szállodában véletlenül összefut egy régi ismerősével (egy földijével), aki már évek óta abban a városban él, és őt is megkéri, segítsen neki jó vendéglőt találni, akkor ez az ember vélhetőleg „jó” tanácsot fog adni az utazó számára. Egyszerűen azért, mert ebben a kérdésben ők ketten hasonló „ízléssel” rendelkeznek, közös tudásuk van arról, hogy mi számít nekik „hazai íznek”.

A fenti két példa a valóságos életünkben származik, de a társas navigáció jelenségét és annak fontosságát sokan észrevették, megállapították a hálózati kommunikáció világában is (már a kezdetektől fogva). A jelenség megnevezése is a web korai időszakból származik [Dourish & Chalmers 1994].

A társas navigáció olyan szimbólumokat, gyakorlatokat, szabályokat, szabályosságokat alakít ki, amelyek révén az „egyszerű” (valamint jelentés és értelem nélküli) térből ('space') társadalmi tér ('social space') vagy más néven: helyszín ('place') formálódik ki.

Dourish és Chalmers felosztását sokan meghivatkozták, átvették, a jóval későbbi publikációkban is sokszor egyetértőleg tárgyalják (lásd például: [Wesley et al. 2007]). A társas navigáció értelmezését persze pontosították, bizonyos értelemben ki is tágították. Dieberger még 1997-ben kibővítette a Dourish és Chalmers által felvázolt fogalomértelmezést [Dieberger 1997].

"Nagyban befolyásolta őt Tom Erickson 1996-os kijelentése, miszerint a web *társas hipertextként* is jellemezhető. Erickson szerint a személyes honlapok többsége 'érdekes emberekre, helyekre' vonatkozó listákat tartalmaz, amelyek olyan hatással vannak az oldalak látogatóira, hogy szinte felhívják a tulajdonos barátainak, kollegáinak és egyéb "kedvenceinek" megtekintésére." [Chalmers et al. 2004]

A társasági navigáció lényege tehát az alábbi pontokban összegezhető:

- az a személy, akinek a navigációját vizsgáljuk, más embert (vagy embereket) követ (arra megy, amerre mások mentek előtte);
- az(ok), aki(ke)t követnek, a korábbi interakciók során „nyilvánítják ki” a választott dokumentumok relevanciát;
- különböző interakciók lehetségesek (megnézés, meghallgatás, kommentelés, értékelés, megrendelés, megvásárlás stb.), amelyek erőssége eltérhet egymástól.

A társas navigáció megértéséhez mindezek figyelembevételével, elemzésével juthatunk közelebb.

A YouTube adott oldalát mutató képen sárgával bejelöltük azokat a társas navigációs jelzéseket, amelyek sok esetben orientálják a látogatók tevékenységét, választásait.

a YouTube társas navigációs jelzőtáblái, „ösztönzői”

Társas navigációs jel, amikor látjuk, hogy mennyi komment van egy oldalon (itt: 19315), mennyien nézték meg a videót (itt: 135110501), mennyi like-ot kapott a videó (itt: 527010), mennyien iratkoztak fel a feltöltő (itt: a VEVO) csatornájára (itt: 1869218), de bizonyos esetekben az is befolyásolhatja a „fogyasztási” döntésünket, hogy tudjuk, mennyien dislike-olták az oldalt (kíváncsiak lehetünk arra, hogy „mi lehet ebben olyan rossz, hogy annyian kifejezték a nemtetszésüket”).

Az adott oldal fő tartalmához kapcsolt társas navigációs adatait nézegetve elvonhatja a figyelmünket és más oldalra való ugrásra készttethet, ha észrevesszük, hogy a jobb oldalon felkínált további tartalmakhoz kapcsolt mutatók (itt a megnézés-szám) milyen értékeket mutat. Ha hirtelen észrevesszük azt, hogy van egy brutális nézettségi adattal rendelkező másik videó, könnyen előfordulhat, hogy egyből arra az oldalra ugrunk.

Ezek a példák mind a társas navigáció jelenségét (és erejét) mutatják.

A Facebook az a másik szolgáltató, amelynek felületén hemzsegnak a társas navigációs jelzések. A Facebook-posztokhoz mindig láthatók a hozzájuk rendelt értékelések, hozzászólások, megosztások száma, és ebben a folyamatosan hömpölygő poszt-folyamban jobban megragadja a figyelmünket, ha a like és más emotikonok, a megosztások vagy a hozzászólások száma nagyobb, mert azt érezzük, hogy ha sokan áldoztak a figyelmükkel, munkájukkal valamiért, akkor azt nekünk is érdemes lehet megtekinteni.



Facebook-poszt társas navigációs jelei

Korábban említettük az ajánlórendszerek elemzésekor, hogy ezek a rendszerek is az értékeléseket tartalmazó információt szedik össze, azokat aggregálják, és azok alapján teszik meg a javaslatukat. De ha így tesznek, az voltaképp nem jelent mást, mint hogy a többiek korábbi értékeléseit valahogyan felmutatva „késztetnek” minket arra, hogy kövessük a többieket.

A Google keresőmotor színre lépése után meglepően gyorsan bekövetkező piaci sikere talán legnagyobb mértékben a újfajta relevanciaképző algoritmus, a PageRank alkalmazásának köszönhető. A PageRank voltaképpen a weboldalakon elhelyezett linkek iteratív és súlyozott aggregálásával képzett tetszési index. Ha egy oldalra sok, azon belül minél több jelentős oldalról érkező link mutat, akkor az előbbre kerül a relevancia-rangsorban. Ez abban az értelemben nem mutatkozik meg úgy, hogy a felhasználók látnák és emiatt követnék ezt a fajta népszerűséget, de a Google keresőmotor használatába mégis bele van valahogy égetve ez a közösségi jelleg: attól lesz valami népszerű, és akkor választják sokan az oldalt, ha a közösség múltbeli tevékenysége (itt a linkkészítés) hatással lehet az adott felhasználó navigációs döntésére a jelenben. Márpedig a Google esetében erről van szó.

3.3.1.4.3.1 Értékelő interakció

A társas navigáció során a korábbi interakciókban megnyilvánuló értékeléseket, eltérő mértékű elköteleződéseket kell figyelembe venni. Első pillantásra nem minden interakció tűnik értékelésnek, de ha az interakció egyes típusait sorba állítjuk, akkor látszik, hogy az ember és a dokumentum viszonya a különböző interakciótípusokban mindig tartalmaz valamiféle elköteleződést, bevonódást, tehát minden interakció hordoz valamiféle értékmozzanatot.

Vannak persze olyan interakciók, amelyek esetében direkt értékelésről van szó. Ilyen például az **értékelés** (rating) vagy a **sorba rendezés** (ranking). Az előbbi esetben a felhasználók kardinális rendezést hajtanak végre valamilyen mérték szerint (számszerűen értékelnek), ezzel szemben a sorba

rendezés során ordinális rendezésről beszélünk (amikor csak egymásutániságot fejeznek ki két vagy több elem között). Ezek az értékelések a szolgáltatás dokumentumaira vonatkoznak, és az értékeléseket arra használják, hogy segítségükkel a szolgáltatás egészén belül lehessen szűrni. A Slashdot szolgáltatásban például az alacsony átlagminősítéssel rendelkező hozzászólásokat (hozzászólókat) ki lehet szűrni, ezáltal áttekinthetőbbé, követhetőbbé válik a rendszer.

De rengeteg más interakció lehetséges, és ezekben – ahogy említettük –nem mindig lehet észrevenni a meghúzó értékmozzanatot. Pedig a következőkben pont azt mutatjuk be, hogy az egyes interakciótípusok különböző elköteleződésekkel „tartalmaznak”, és pont ezek a „digitális jelek” (digital mark) használhatók ki a társas navigáció során.

A teljes igénye nélkül nézzünk meg néhány tipikus interakciótípust:

- befogadni egy dokumentumot (listening, viewing, reading)
- vásárlás (shopping)
- könyvjelző elhelyezése (bookmarking)
- kattintás (clicking)
- értékelés (rating)
- sorba rendezés (ranking)
- link elhelyezése a szövegben (linking)
- bejegyzés (posting)
- kommentálás (commenting)
- link elhelyezése videóra (responding to video)
- levél küldése (sending email)
- oldal elküldése (sending web page)
- oldal tovább küldése (forwarding web page)
- feltöltés (uploading)
- letöltés (downloading)

Az általánosítás érdekében vannak interakciók, amelyek „ugyanazt” jelentik, bár másként hivatkoznak rájuk. Dokumentumtípustól függően mást mondunk, ha megnézünk egy képet, elolvassuk egy szöveget vagy meghallgatunk egy zeneszámot, de mindegyik interakció lényege az, hogy az adott dokumentum tartalmát befogadjuk (interakciós és értékelési szempontból mindegy, hogy egy fórum, egy blog, egy newsgroup vagy egy portál szöveges anyagaihoz fűzünk kommentet). Ebből a szempontból vizsgálva tehát egyszerűsíthetünk a képen, ha a korábban már bemutatott információtipológiára támaszkodva „összevonunk” pár interakciótípust.

Fontos az is, hogy különbséget tegyünk kétféle tartalomtípus között. A navigáció célja mindig valamilyen dokumentum megtalálása, de ehhez mindig valamilyen metainformációt használunk. A web világában van egy speciális (és rendkívül fontos) metainformáció, a *link*, amely az adott dokumentum címét mutatja. Az interakció egyik típusa pont az ilyen linkek elhelyezését jelenti. Ekkor ne csak azokra a „hagyományos” linkekre gondoljunk, amelyeket a weboldalakra ágyaznak be az oldal szerkesztői. Ide tartozik például a játsszasi listák (playlist) elkészítése, a kedvencek közé emelése és ezek felsorolása, más multimédia dokumentumokra való mutató (respond to video) stb.

A fentiek alapján tehát az alábbi értékeléseket „fedezhetjük fel” az egyes interakciókban a társas navigáció során:

- kedvenclista (favorites), playlist (játsszasi lista)
- toplisták
 - tartalom-toplisták
 - leggyakrabban fogyasztott (nézett, hallgatott, olvasott, letöltött)
 - legjobbra értékelt
 - legfrissebb
 - egyéni kedvenclistákban leggyakrabban felbukkanó tétel
 - felhasználó-toplista
 - a legújabb/legutóbb közreműködő felhasználók

- a legújabb/legutóbbi fórumhozzászólást, kritikát, minősítést, kommentárt közzétevő felhasználók
- a legújabb/legutóbbi fájlt, tartalmat megosztó felhasználók
- a legújabb/legutóbbi sávzélességet megosztó felhasználók
- a legtöbbet közreműködő felhasználók
 - a legtöbb fórumhozzászólást, kritikát, minősítést, kommentárt közzétevő felhasználók
 - a legtöbb fájlt vagy legnagyobb mennyiségű tartalmat megosztó felhasználók
 - a legnagyobb sávzélességet megosztó felhasználók
- a legnagyobb reputációval/bizalommal rendelkező felhasználók
- a legtöbb baráttal rendelkező felhasználók
- a legkevésbé közreműködő felhasználók listája
- az előző lista ellentéte
- „direkt értékelés” (evaluation)
 - linking, valued list: kedvenclista (favorites), playlist (játsszási lista)
 - rating,
 - ranking

Az interakcióban megnyilvánuló elköteleződések azért fontosak számunkra, mert ezek „szolgáltatják” az iránymutatást a társas navigáció során.

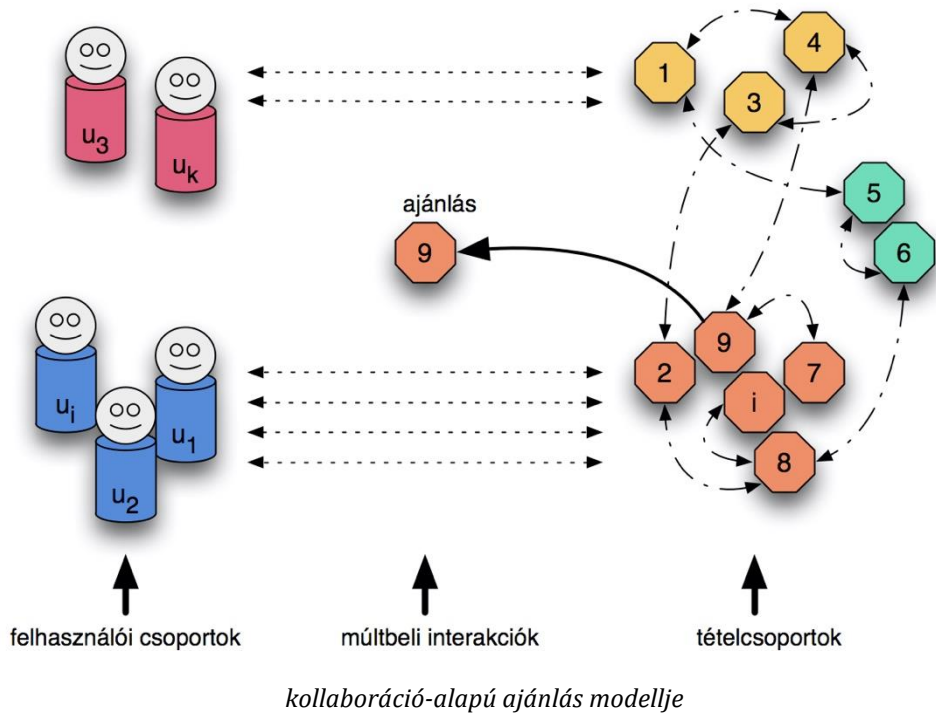
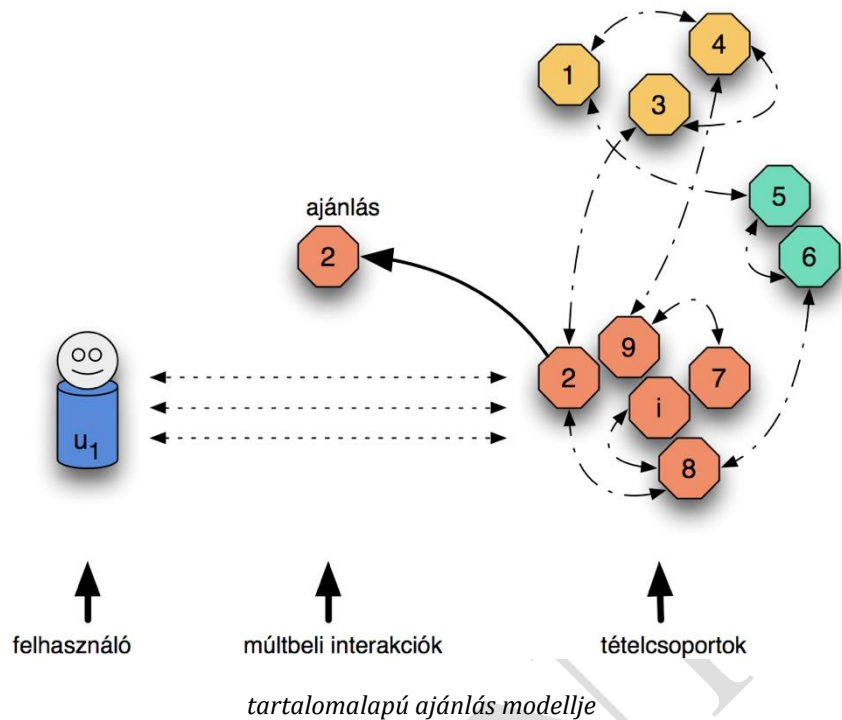
3.3.1.4.3.2 Ajánlórendszer

A társas navigáció megértéséhez tehát abból kell kiindulnunk, hogy a jelenség lényegét a felhasználók korábbi interakcióiban tetten érhető preferenciakinyilatkoztatások jelentik. A társas navigáció két típusa különíthető annak alapján, hogy **explicit** vagy **implicit rendszerekről** van-e szó. Az előbbieken a preferenciát meg lehet szerezni a felhasználóktól, míg az utóbbiak esetében a felhasználók preferenciáit áttételesen kell kikövetkeztetni.

Akárhogy is van, az ilyen rendszereket **ajánlórendszereknek** (recommendation system) nevezzük, mert mindig arról van szó, hogy – explicit vagy implicit módon – más felhasználók interakcióit úgy is felfoghatjuk, mint egyfajta ajánlásokat a későbbi felhasználók számára. Az ilyen rendszerek típusait aszerint találhatjuk meg, hogy mi alapján valósulnak meg ezek az ajánlások:

- **tartalomalapú ajánlás** (content-based recommendation): a felhasználóknak saját múltbeli preferenciáikhoz hasonló tételeket ajánlanak;
- **kollaboráció-alapú ajánlás** (collaborative recommendation): a felhasználóknak hozzájuk hasonló más felhasználók múltbeli preferenciáihoz hasonló tételeket ajánlanak;
- **kevert megoldás** (hybrid approaches): ezek a módszerek keverten tartalmaznak tartalomalapú és kollaboráció-alapú ajánlásokat.

A lényegi különbség a tartalom-, illetve a kollaboráció-alapú ajánlások között van. Ezek sémáját mutatjuk be a következő két ábrán:



A legelső ajánlórendszer a *Grundy system* volt [Adomavicius et al. 2005], [Rich 1997], bár sokan hivatkozzák meg elsőként a Xerox PARC műhelyében kifejlesztett *Tapestry* rendszert. Utóbbi levelezőrendszerekben (hírcsoportokban, hírlevelekben) kínálta a felhasználók számára azt a lehetőséget, hogy megjelöljék, ha egy hírt érdekesnek, jónak (vagy érdektelennek, rossznak) gondoltak, és a preferenciamegjelölésekből kiindulva ajánlásokat tudtak tenni a közösség más tagjai számára az egyéni profilok hasonlósága alapján [Goldberg et al. 1992]. Már a *Tapestry* fejlesztői is jelezték, hogy

a rendszerük bármilyen információfolyam esetében működőképes lehet, nem csak levelezőrendszerekben lehetne alkalmazni. És valóban, a későbbiekben megjelenő ajánlórendszerek mind a Tapestry működéséhez hasonló logikát követtek.

A GroupLens, a Video Recommender és a Ringoadjusted voltak az első olyan rendszerek, amelyekben a predikcióra már kollaboratív szűrés technikát használtak [Adomavicius et al. 2005]

Már a Tapestry megjelenésekor összekapcsolták az ajánlórendszer fogalmát a *közösségi szűrés* (collaborative filtering) fogalmával. Nyilván azért, mert egyfelől az egyéneknek szóló ajánlatokat csak azon az alapon lehet megtenni, hogy a közösség egésze végzi az értékelést, másfelől az így létrejövő ajánlások nem jelentenek egyebet a felhasználók számára, mint hogy a teljes kínálatból csak a számukra nagy eséllyel érdekes, fontos egységek lesznek elérhetők, amely a szűrési feladatnak egy lehetséges megoldása. A közösségi szűrés két típusát különböztethetjük meg:

- aktív közösségi szűrés (active collaborative filtering)
- passzív közösségi szűrés (passive collaborative filtering)

Az **aktív közösségi szűrés** (active collaborative filtering) esetén a felhasználók értékelik a termékeket, dokumentumokat, és az egyéni értékelések aggregált mutatója láthatóvá válik a közösség egésze számára, ez pedig megkönnyíti mások számára a döntéshozatalt. Ebben a helyzetben is hasonló problémák vannak, mint a personalizáció esetében.

Az **első-értékelés problémájaként** (First-Rater Problem) hivatkoznak arra a hiányosságra, hogy a rendszerbe érkező új elemeket értelemszerűen még senki sem értékelhette, és amíg ez a helyzet nem változik, ezek az egységek nem is kerülhetnek be a „körforgásba” (vagyis a többiek sem látják, ők sem ítélik meg ezeket). Értékelés hiányában pedig a rendszer nyilván nem képes kapcsolatot teremteni a különböző egységek, illetve az egységek és a felhasználók között. Az ajánlórendszer mechanizmusa így nem tud beindulni.

A másik rendszerprobléma szintén „első lépés” jellegű, csak ekkor az új felhasználóval „van gond”. A **hidegindítás problémájaként** meghivatkozott esetben (Cold-Start Problem) az újonnan belépő felhasználót azért nem lehet sokáig „megragadni” a rendszerben, mert amíg egyetlen értékelést sem adott le, addig a rendszer semmilyen módon nem tudja kitalálni az új felhasználó preferenciáit, hiszen ehhez nincs semmiféle összehasonlítási alapja az új ügyfél esetében.

A **passzív közösségi szűrés** (passive collaborative filtering) során a felhasználó különböző tevékenységeit lehet megfigyelni, amelyek az adott termék, szolgáltatás, dokumentum stb. iránti elköteleződést, relevanciatalajdonítást jelzik. Ilyen felhasználói tevékenységek lehetnek:

- megvásárolni valamely terméket;
- folyamatosan, ismételtelen felkeresni, használni, letölteni, elmenteni, kinyomtatni valamely terméket, oldalt, dokumentumot;
- hivatkozni, linkkel rámutatni egy oldalra, egy site-ra;
- több alkalommal, ismételtelen keresni valamit.

Az ajánlórendszerek a vásárlási és fogyasztásbeli hasonlóságuk alapján próbálnak meg ajánlatokat tenni a felhasználók, vásárlók, fogyasztók számára. Nemcsak dokumentumok, hanem bármilyen vásárolható, fogyasztható termék vagy szolgáltatás esetében működnek az ilyen metodikák, ezért a legáltalánosabban ezeket a 'katalógustétel' (catalog item) megnevezéssel illelhetjük.

A különböző logikájú ajánlórendszer-algoritmusok mindig valamilyen hasonlóságot keresnek a tranzakciós folyamatokban. Mindig vásárlási, értékelési, kattintási, figyelemmegragadó eseményt figyelnek meg, vagyis valamilyen forgalmi információt elemeznek. A felhasználók (vásárlók, értékelők) valamilyen jószággal kapcsolatban a figyelmükkel fizetnek, olyan tevékenységet hajtanak végre, amely valamilyen mértékű elköteleződést mutat a termék iránt. Az elköteleződés megnyilvánulhat a megtekintés, a kattintás, az értékelés, a vásárlás, a megrendelés stb. aktusában, és ez az elköteleződés egyben kifejezi azt is, hogy a felhasználó számára az adott jószág releváns. A jószágból megint csak sokfélélt elképzelhetünk. Lehet termék, lehet szolgáltatás, azon belül lehet valamilyen dokumentum (video, zeneszám, könyv, cikk stb.).

A hasonlóságkeresés többféle algoritmus révén valósulhat meg, amelyek alapvetően abban különböznek egymástól, hogy inkább a felhasználók vagy inkább a jószágok közti hasonlóságot veszik-e inkább figyelembe. A legelterjedtebb algoritmusok az alábbiak [Linden et al. 2003]:

- hagyományos közösségi szűrés (traditional collaborative filtering)
- klasztermodell (cluster model)
- keresés- vagy tartalomalapú módszertan (search-based method/content-based method)
- termék alapú közösségi szűrés (item-to-item collaborative filtering)

A **hagyományos közösségi szűrés** a felhasználók közötti hasonlóság elvén működik: a felhasználók és a dokumentumok közötti kapcsolatokat (vásárlást, értékelést, kattintást, megtekintést stb.) egy hatalmas mátrixba építik, majd a felhasználók, vásárlók között hasonlóságokat keresnek. majd a hasonló vásárlói, fogyasztói, felhasználói viselkedéseket, mintázatokat megtalálva már képesek ajánlásokat megfogalmazni az aktuális felhasználó számára.

A **klasztermodell** a felhasználókat egy klaszterező eljárással (vagy más, nem-felügyelt tanuló algoritmus segítségével), offline (nem-valósídejű, aszinkron) módon hasonlósági osztályokba rendeli, és az online (valósídejű, szinkron) felhasználó profilját (azaz a termékekhez való viszonyát leíró vektort) az így szegmentált felhasználói csoportok profiljaival összehasonlítja, hogy megtalálja a felhasználó profiljához leginkább hasonló klaszterprofil.

A **keresés- vagy tartalomalapú módszertan** (search-based method, content-based method) a dokumentumok, termékek, katalógustételek közötti szemantikai kapcsolatokat kihasználva képes ajánlani. A módszer abból indul ki, hogy a termékeknek, dokumentumoknak vannak metaadatai, amelyek alapján minden egyes katalógustételt el lehet helyezni az adott dokumentumgyűjtemény metaadatainak szemantikai terében. Ebben a szemantikai térben az egyes egységek között előfordulhatnak hasonlóságok, azonosságok, amelyek mentén már lehet ajánlásokat tenni. Ez a módszer így ajánlhat más dokumentumokat az azonos, vagy hasonló formai leíró metaadatai alapján (azonos alkotó, szerző, rendező, zeneszerző, előadó más alkotásait), de a rendszer javasolhat más műveket a tartalmi leíró metaadatok hasonlósága, egyezése révén (műfaji hasonlóság, tárgyszavak, kategóriák vagy címkék szerinti kapcsolódások mentén). Ez a metodika csak akkor lehet működőképes, ha a felhasználónak nincs túlzottan sok már megvásárolt vagy értékelt tétele, mert ilyen esetekben a szemantikai kapcsolatok száma kezelhetetlenül nagyvá válik.

Look for Similar Items by Category

[Nonfiction](#) > [Philosophy](#) > [General](#)
[Nonfiction](#) > [Social Sciences](#) > [Anthropology](#) > [Cultural](#)
[Science](#) > [General](#)
[Science](#) > [History & Philosophy](#) > [General](#)

Look for Similar Items by Subject

[Philosophy of mathematics](#)
 [Social & political philosophy](#)
 [Science](#)
 [Archaeology / Anthropology](#)
 [Science/Mathematics](#)
 [Anthropology - Cultural](#)
 [General](#)
 [Philosophy & Social Aspects](#)
 [Philosophy / General](#)
 [Science / Philosophy & Social Aspects](#)
 [Science : Philosophy & Social Aspects](#)
 [Social Science : Anthropology - Cultural](#)

Find books matching ALL checked subjects

i.e., each book must be in subject 1 AND subject 2 AND ...

az amazon.com könyvtárház szemantikai kapcsolódásokon alapuló ajánlórendszere

A **termék alapú közösségi szűrési algoritmus** (item-to-item collaborative filtering) abban tér el a hagyományos közösségi szűréstől, hogy nem a felhasználók, hanem a termékek közötti hasonlóságot keresi (és használja ki az ajánlatok kidolgozásához). Két termék közötti hasonlóság alapja ilyenkor az a tény, hogy a két terméket ugyanaz a vásárló vette meg (vagy értékelt pozitívan). A termékekre offline-

aszinkron módon felépíthető hasonlósági mátrix segítségével így már valós időben is nagyon gyorsan lehet ajánlatokat kínálni a felhasználók számára.

Customers Who Bought Related Items Also Bought Page 1 of 9

az amazon.com könyvtárház 'Item-to-Item Collaborative Filtering' ajánlórendszere

Customers Who Bought Items in Your Shopping Cart Also Bought Page 1 of 4

az amazon.com könyvtárház 'shopping cart' ajánlórendszere

Your Recent History (What's this?) Page 1 of 13

Recently Viewed Items

After viewing product detail pages or search results, look here to find an easy way to navigate back to pages you are interested in.

Look to the right column to find helpful suggestions for your shopping session.

[View & Edit Your Browsing History](#)

Continue shopping: Recommended for You

az amazon.com könyvtárház 'személyes történeten alapuló' ajánlórendszere

A négyféle ajánlórendszer-logika más előkészítő fázisokat követel meg, más számítási igényekkel rendelkezik, más relevanciaminőséget (pontosságot és kiterjedtséget) képes biztosítani, más skálázási lehetőségeket rejt magában [Linden et al. 2003], de ezek kifejtése, szakszerű ismertetése már nem a jelen beszámoló feladata. Azokat az algoritmusokat, matematikai, számítástechnikai módszereket, amelyek a közösségi ajánlórendszereket működtetik, matematikusok, informatikusok fejlesztik. Ők azt értik, azt tudják, hogy miként működnek a gépek, az algoritmusok. Ez a gondolatmenet arra próbált meg rávilágítani, hogy mi, hogyan és miért történik az emberek világában.

3.3.1.5 Az információelérés típusai

Az eddigieket összefoglalva most már megadhatjuk, hogy milyen fontosabb információelérési lehetőségeket, technikákat, mechanizmusokat különíthetünk el egymástól. Az eddigiek alapján mondhatjuk, hogy a dokumentumokhoz való hozzáférést, a dokumentumokban tárolt információk elérését biztosítani lehet:

- előzetesen, aszinkron módon, szűréssel feldolgozott, böngészéssel elérhető dokumentumtartományokban:
 - brandelt tartalmak elkülönítésével (YouTube, Flickr, Digg, Twitter, Index, origo, last.fm)
 - közösségi ajánlórendszerekkel, társas navigációval
 - perszonalizációval: az egyedi felhasználóhoz rendelt szűrőfeltételek figyelembevételével
 - aktív személyi szűréssel: a felhasználó által megadott szűrőfeltételekkel;
 - passzív személyi szűréssel: a felhasználó viselkedéséről gyűjtött adatok szűrőfeltételekké konvertálásával;
 - előrerendezéssel, osztályozással, katalogizálással, tartalmi-szemantikai navigációval
- valós időben, szinkron módon, kereséssel feltárható dokumentumtartományokban:
 - a dokumentumok tartalmaiban
 - karakteres kereséssel írásszöveg-dokumentumban
 - gépi beszédfelismerésen alapuló transzkriptálás után beszédszöveg-dokumentumokban
 - gépi dallam-, ritmus- és harmóniafelismerésen alapuló kereséssel zenei dokumentumokban
 - gépi zörejelismerésen alapuló kereséssel hangdokumentumokban
 - gépi képfelismerésen alapuló képmintázat szerinti kereséssel álló- és mozgóképes dokumentumokban
 - a dokumentumok metaadataiban

3.4 Gyakorlati következmények

Ebben a fejezetben azt vizsgáljuk meg, hogy az előző fejezetekben tárgyalt elméleti problémák gyakorlati megoldására tett javaslatokat hogyan hasznosítjuk a névtérrendszer felépítésekor. Végig kell mennünk azon, hogy miket kell tennünk ahhoz, hogy az általános megfontolásokat, javaslatokat saját helyzetünkre konkretizálva tudjuk a tényleges gyakorlatban alkalmazni.

3.4.1.1 A névtérben használt kiemelt relációk pontosítása

A projekt első szakaszában tulajdonnévtérket kell létrehozunk. Ezek közös sajátossága, hogy individuumokat, példányokat kell kezelni bennük. A későbbiekben építeni kell majd köznévtérket, tudásszervezési rendszereket is, amelyekbe már döntően általános fogalmakat (típusokat, osztályokat) kell felvenni. Már az első szakaszban szükségünk van arra, hogy a tulajdonnévtér entitásainak leírásához általános fogalmakat, és tudásszervezési rendszerekben rögzített világtudást használjunk fel. Mindezek miatt nagyon fontos, hogy a rendszer egészén belül olyan relációfogalmakat használjunk, amelyek konzisztens módon tudnak egy általános relációelméletre támaszkodni. Ebben a fejezetben azokat a szempontokat gyűjtjük össze, amelyek segíthetnek ebben a törekvésünkben.

Rögzítsük újra a kiindulópontunkat: a dokumentumok, nevek, névhordozók egyértelmű leírása és a minél pontosabb kereshetőség érdekében ügyelni kell arra, hogy mind a névtérben, mind a névtér által támogatott adatrendszerekben, mind a keresést támogató tudásszervezési rendszerekben minél pontosabb relációfogalmakat használjunk.

A korábban bemutatott relációk (generikus, partitív, eleme, példánya, szinonimája, homonimája stb.) között több olyan is van, amelyek esetében különbséget érdemes tenni aszerint, hogy milyen típusú

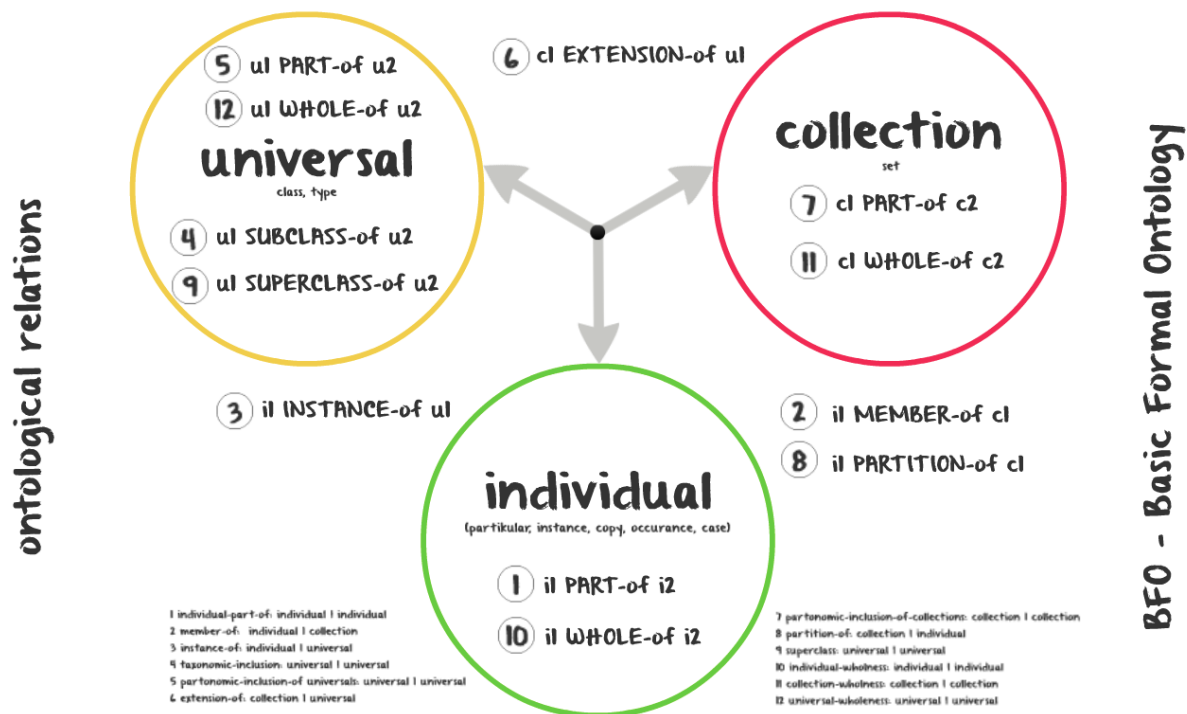
relátumokat kapcsolnak össze. Kiváltképp fontos megtenni ezt a distinkciót a partitív reláció esetében, mivel itt lehet a legnagyobb keveredést előidézni a pongyola fogalomhasználattal.

A BFO (Basic Formal Ontology) ajánlásait követve a legfontosabb relációk esetében pontos és fontos megkülönböztetéseket tehetünk (Bittner et al. 2004). Az első lényeges figyelembe veendő szempont szerint háromfajta fogalomtípust kell elkülönítenünk egymástól, úgymint:

- **individuum**, példány (individual)
- **típus**, univerzálé, osztály (universal)
- **gyűjtemény** (collection)

A háromféle fogalomtípus között különféle relációk definiálhatók, amelyek olykor ugyanazon relációtípusba tartoznak, de a relátumok minőségében különböznek egymástól, ezért el kell különíteni őket egymástól.

Az alábbi ábra mutatja azt, hogy a három alapvető kategória között milyen relációkat kell definiálnunk (az ábrán az angol nyelvű terminusokat tüntettük fel, hogy igazodjunk az alkalmazott elmélet terminushasználatához).



a három alapkategória és a köztük értelmezett relációk

A 'típus' ('universal') és a 'példány' ('individual') fogalmak közti különbséget nem szükséges magyaráznunk, sok tudományterületen használják ezt a fogalom párt (még ha sokszor más terminusokat is alkalmaznak). A két fogalom közti kapcsolatot azzal jellemezhetjük, hogy minden tulajdonság, amit a típusra feltételezünk, megjelenik (meg kell, hogy jelenjen) a típusfogalom terjedelmébe tartozó individuumfogalomban.

A 'típus' és a 'individuum' közti kapcsolatot írhatjuk le a 'példánya' relációval.

Fontosabb kérdés az, hogy miért van szükség a 'gyűjtemény' ('collection') fogalomra a modellben. Egy példa segítségével tudjuk ezt szemléltetni. Ha vesszük a pártlistás adatokat a választási adatbázisból, akkor láthatjuk, hogy könnyen előfordulhat az az eset, amikor nem a pártokhoz mint individuumokhoz kell a szavazatokat rendelni, hanem olyan pártlistákhoz, amelyeken több párt neve is szerepel. Ebben az

esetben pártok egy csoportját kell megragadnunk, nem az egyedi pártokat, amit a modellbe felvett 'collection' entitás segítségével tudunk megoldani. Egy másik példát is említhetünk: a könyvtári világban nem csak a személyeket és a testületeket szokás leírni, de igény van arra is, hogy családokat lehessen adatokkal ellátni. Azt tudjuk, hogy a 'család' fogalmán emberek egy halmazát értjük, de erre az entításra nem állíthatjuk sem azt, hogy személy, sem azt, hogy testület lenne. A családokat viszont leírhatjuk úgy, hogy azt mondjuk, a család emberek valamilyen gyűjteménye. A gyűjtemény mindig valamilyen individuumok halmaza. Fontos viszont azt is hangsúlyozni itt, hogy a gyűjteményekbe nem mindig kell azonos minőségű dolgokat felvenni. Gyűjtemény lehet egy konkrét személy, két konkrét testület és három konkrét földrajzi hely.

A gyűjtemény és a példány közti kapcsolatot fejezi ki a – halmazelméleti értelemben vett – 'eleme' reláció. Az 'eleme' és a 'példánya' reláció közti különbség egyfelől abban fejezhető ki, hogy az individuumok egyszer egy gyűjtemény-, másszor egy típusfogalomhoz kapcsolódnak, másfelől megkövetelhetjük azt, hogy a 'példánya' relációval csak olyan kapcsolatot fejezzünk ki, amelyben a reláció bal oldalán szereplő összes individuum osztozik egy közös tulajdonságban (ami alapján a típusfogalom alá lehet őket sorolni. Az 'eleme' reláció esetében ezt az utóbbi feltételt nem követeljük meg.

Ez a különbség ad magyarázatot arra is, hogy miért vannak, lehetnek, sőt, kell, hogy legyenek olyan gyűjtemények, amelyek legalább egy szempontból azonos minőségű elemekből állnak. Akkor van szükség ilyen értelemben vett gyűjteményre, amikor egy típus előfordulásaira akarunk rámutatni. Valamely típusfogalom előfordulásait (példányait) olyan gyűjteménynek tekinthetjük, amelynek összes eleme rendelkezik a típusba tartozást jelző adott tulajdonsággal. A 'típus' és a 'gyűjtemény' közti reláció fogja egybe a típusfogalom terjedelmét, extenzióját, azért a köztük levő relációnak az 'extenziója' nevet adhatjuk.

A 'példány' és a 'gyűjtemény' fogalma között értelmezhető még egy további reláció is, amit a partitív reláció egyik altípusaként definiálhatunk. Olykor szükség lehet arra, hogy a példányok egy csoportját úgy kössük össze egymással, hogy azok a teljes gyűjteményen belül egy jól elkülöníthető halmazát alkossák. Ha így teszünk, partíciókat hozhatunk létre, amelyek – ideális esetben – teljesen diszjunktak, és együttesen lefedik a teljes gyűjteményt. Ezt a kapcsolatot írhatjuk le a 'partíciója' relációval.

Az eddigiekben azokat a relációkat mutattuk be, amelyek a három főkategória között érvényesek. A továbbiakban nézzük át azt, hogy milyen relációkat lehet definiálni az egyes kategóriákon belül.

Előzetesen rögzíthetjük azt, hogy egyfelől az általános generikus és partitív relációkat lehet speciális jelentéssel a három kategórián belül értelmezni, másfelől a e két reláció esetében mindig értelmes és hasznos lehet azok inverzéről beszélni, így a továbbiakban mindig relációpárokat fogunk bemutatni.

Mindhárom kategórián belül értelmezhetünk egy-egy partitív relációpárt, a partitív alárendeltje (része) és a partitív fölérendeltje (egésze) relációt.³⁷ Nem idézünk itt fel minden definitív mozzanatot a korábbi kutatási jelentésünkben bemutatott mereológiaiából, de azt az egy tételt kiemeljük, miszerint a partitív reláció meghatározásában tételezzük a reláció tranzitivitását. Erre a későbbiekben még hivatkozni fogunk. A relációpárok két-két tagja közt mindig fennáll az inverzitás, amit a következőképpen fejezhetünk ki.

partitív alárendeltje $(x,y) \equiv$ partitív fölérendeltje (y,x)

A típustartományban értelmezett partitív reláció két típusfogalom között teremt kapcsolatot. Példaként vehetjük az alábbiakat:

a fark része a kutyának

a ló része a ménesnek

a szarvasmarha része a gulyának

a megye része az országnak

³⁷ A partitív reláció formális meghatározását megadtuk a projekt egy korábbi kutatási jelentésében.

Ezek az állítások azt írják le, hogy minden kutyára igaz, hogy ha van farka a kutyának, akkor az a kutya része, minden ménes lovakból, minden gulya szarvasmarhákból áll, minden megye az országhatárokon belül van, így minden, ami a kutyafarokhoz tartozik, az a kutyához is tartozik, minden, ami a lóhoz tartozik, az a méneshez is tartozik, minden, ami szarvasmarhához tartozik, az a gulyához is tartozik, és minden, ami a megyében van, az az országban is van. Fontos itt, hogy a partitivitás csak a megadott fogalmak között áll fent. Elképzelhető, hogy egy legelésző ménes lovai közé beballag egy tehén, és ekkor mondhatjuk azt, hogy egy csoportban vannak (egy gyűjteményt alkotnak), de ettől még a tehenet nem tartjuk a ménes részének.

A korábbi kutatási jelentésben kifejtett relációelméletben elkülönítettük az általános partitív reláción belül két altípust, ami a most bevezetendő – más értelemben vett – altípuspárok esetében is értelmes altípuszálásra ad lehetőséget. A partitív relációt leíró elméletben elkülönítik egymástól az **atomos** és a **nem-atomos** partitív relációkat, amelyeket nem szabad egymással felcserélhető módon használni. Előbbire példa a ló és a ménes, a tehén és a gulya vagy a csellós és a vonósnégyes közti partitív reláció, míg utóbbira a fark és a kutya, a pata és a ló vagy a zenész keze és a zenész közti partitív kapcsolat. A két altípus közti felcserélhetőség tilalmát pedig egy példával szemléltethetjük. Nézzük az alábbi három mondatot.

a zenész keze része a zenésznek

a zenész része a zenekarnak

a zenész keze nem része a zenekarnak

Ez a három állítás mind igaz, viszont ha fent akarjuk tartani a partitív relációra megállapított tranzitivitás érvényességét, márpedig fent kell tartanunk, akkor ez a helyzet ellentmondásosnak tűnhet, hiszen a három állításban nem teljesül a tranzitivitás. A megoldás csak az lehet, hogy a fenti három kijelentésben alkalmazott partitív relációfogalmat kettéválasztjuk, és a két fogalmat külön-külön és csak a megfelelő helyen használjuk. Ha felveszünk egy 'nem-atomos_része' és egy 'atomos_része' relációt az általános része reláció alá, akkor a fenti kijelentéseket pontosíthatjuk.

a zenész keze nem-atomos_része a zenésznek

a zenész atomos_része a zenekarnak

a zenész keze nem atomos_része a zenekarnak

a zenész keze nem nem-atomos_része a zenekarnak

A négy állítás mindegyike megáll önmagában és nincs köztük ellentmondás sem, mert a tranzitivitást két eltérő reláció közt nem kell elvárunk. Az első (a zenész kezére és a zenészre vonatkozó) kijelentés a 'nem-atomos_része' relációt használja, míg a második (a zenészre és a zenekarra) vonatkozó állítás az 'atomos_része' relációt tételezi. A harmadik és negyedik mondat viszont nem igaz, egyszerűen azért nem, mert a zenész keze és a zenekar között sem az 'atomos_része', sem a 'nem-atomos_része' relációt nem értelmezhetjük. Az ellentmondát viszont így feloldottuk, mivel nincs „tranzitivitás-kényszer”, ha az első két állítás különböző relációt használ.

A típusfogalmak és a partitív reláció kapcsolatáról eddig elmondottak természetesen elmondhatók a fogalmak terjedelmébe sorolt példányok esetében is, de akkor már más partitív relációt kell az individuális entitások között felvennünk. Mondhatjuk a következőket.

ez a konkrét fark része a Bodri kutyának

a Pejko nevű ló része a Béla ménesének

a Riska nevű szarvasmarha része a szomszéd gulyájának

Fejér megye része Magyarországnak

Az itt alkalmazott partitív relációfogalmakat természetesen ugyanúgy és ugyanazért fel kell bontani atomos és nem-atomos altípusokra, mint ahogy és amiért ezt tettük a típusfogalmakhoz tartozó 'rész' reláció esetében. Új relációk bevezetésére pedig egyfelől azért van szükségünk, mert a relációk relátumai itt más kategóriákba esnek, mint az előző esetben, másfelől pedig értelmetlen lenne típusfogalmat használni individuális fogalmak között.

A gyűjtemények kategóriáján belül értelmezhető partitív reláció individuumok gyűjteményei közti rész-egész kapcsolatokat kell kezeljen, ami nem igazán teszi lehetővé, hogy ezen a területen atomos partitív relációt alkalmazzunk, bár ennek az állításnak elméleti bizonyítását még nem tudjuk megadni. A következő példát mutatjuk be az itt alkalmazható rész-egész relációra.

a magyarországi városok halmaza része a magyarországi települések gyűjteményének

Budakeszi, Szentendre és Pomáz része a Pest megyei települések halmazának

Szentendre és Szolnok része az 'sz' betűvel kezdődő települések halmazának

Eddig egyszer sem adtunk arra példát, hogy a partitív alárendeltje (rész) reláció mellett hogyan lehet értelmezni ennek inverz párját, a partitív fölérendeltje relációt. Most mutatunk egyet.

az 'sz' betűvel kezdődő települése halmaza tartalmazza Szentendre és Szolnok településeket

Ennek analógiájára a korábbi esetekre is hasonló módon lenne elvégezhető az inverzrelációk képzése. Egyetlen reláció maradt még ki az ismertetésből, ami – talán – a legfontosabb az összes közül. Ez a generikus alá- és fölérendeltje relációpár. Ez a reláció csak a típusfogalmak tartományán belül értelmezhető. A tudásreprezentálás legfontosabb relációjáról van szó, hiszen a generikus fogalmak segítségével tudunk új fogalmakat bevezetni. Vegyünk pár példát erre is.

a kutya is-a-type-of állat

a repülő is-a-type-of közlekedési eszköz

az ember is-a-type-of élőlény

az atomos_része reláció is-a-type-of része reláció

A korábbi kutatási jelentésben megadtuk a generikus reláció formális meghatározását is, amiből csak annyit idéznénk most fel, hogy a generikus reláció másodrendű, hiszen típusfogalmak közt teremt kapcsolatot, és ennek leírásához arra van szükség, hogy amikor kifejezzük a formuláját, akkor ne csak a generikus relációval összekapcsolt két típusfogalom terjedelmébe tartozó előfordulásokra szűkítsük az univerzális kvantor hatókörét, de arra is szükség van, hogy kimondjuk egy második univerzális kvantor alkalmazásával, hogy a generikus reláció minden összekapcsolt típusfogalompár esetében érvényes kell legyen.

A generikus alárendeltje (is-a-type-of) relációra is megadhatjuk annak inverzét, amire ugyanolyan logika mentén írhatunk fel egy formulát, mint ahogy azt a partitív reláció inverze kapcsán tettük.

generikus alárendeltje $(x,y) \equiv$ generikus fölérendeltje (y,x)

A generikus reláció adja a kezünkbe annak lehetőségét, hogy ha nincs olyan típusfogalmunk, amely alkalmas lenne a világ valamely szegmensére vonatkozó tudásunk reprezentálására, akkor új fogalmat hozunk létre a generikus reláció segítségével úgy, hogy veszünk egy fölöttes fogalmat és megadjuk az

új, definiálandó fogalom jelentését az – arisztotelészi értelemben vett -differentia specifica meghatározásával. Egy példával szemléltetve mindezt vegyük az 'autó' fogalmát, amelyre az önvezető autók megjelenéséig azt mondhattuk, hogy olyan közlekedési eszköz, amit gépi erővel lehet helyváltoztatásra használni úgy, hogy közben szükség van a – emberi – 'vezető' fogalmára is, ám amikor létrehozuk az önvezető autókat, akkor utóbbiakra be kell vezetni egy új fogalmat, ami már nem kívánja meg a 'vezető' fogalmát. Ekkor azt mondhatjuk, hogy az önvezető autó olyan autó, amely haladni képes emberi vezető nélkül is. Ez a minőség lesz az, ami specifikálja az önvezető autó jelentését.

3.4.1.2 Reláció típus és relációpéldány

A webes szolgáltatásokkal szemben egyre komolyabb elvárásként fogalmazódik meg az, hogy a mögöttes adatbázis sémája minél inkább igazodjon az RDF-triplet logikához, ahhoz a modellezési filozófiához, amely szerint a világról szóló tudásunkat háromelemű RDF-állítások formájában kell megfogalmazni. A névtér projekten belül nem teljes körűen, de elég kiterjedten igazodunk ehhez az elváráshoz. Sok adattípus esetében relatíve könnyen lehet teljesíteni ezt az elvárást. Vannak azonban olyan adatkörök, amelyek esetében nem triviális, hogy miként lehet megfelelni a triplet-logika által támasztott követelménynek.

Ahhoz, hogy a névterek számára minél általánosabb sémát definiálhassunk, valahogyan kezelni kell azt a problémát, amely olyan adatkörök leírása esetében merül fel, amikor az adattáblák sémái „bonyolultabbak”, amit másként úgy is kifejezhetünk, hogy a sémainformációt jellemző predikátumoknak háromnál több argumentuma van. Ilyen esetekben kérdés az, hogy az ilyen típusú adatokat hogyan lehet „átírni” tripletekre. Kiindulásként nézzünk meg pár példát azért, hogy világossá váljék, mi is a feladat ezen a területen!

választási adatok

| esemény | település | pártlista | szavazatok száma |
|---------|------------|-------------|------------------|
| 2018 | Szentendre | Fidesz-KDNP | 8384 |

településstatisztikai adatok

| esemény | település | adattípus | értéke |
|---------|-----------|----------------------|--------|
| 2015 | Pomáz | munkanélküliek száma | 232 |

filmográfiai adatok

| film | személy | szerepkör | karakter |
|------------|--------------|-----------|----------|
| Üvegtigris | Rudolf Péter | színész | Lali |

más névterek azonosítói

| névtér | típus | földrajzi hely | azonosító |
|-------------------|------------|----------------|-----------|
| KSH Helységnévtár | aktuális | Dunaújváros | 33214 |
| KSH Helységnévtár | történelmi | Sztálinváros | 22341 |

személyek életrajzai különböző forrásokban

| név | forrás | pozíció | életrajz |
|-------------|--------------------------|----------|-----------------------------|
| Balogh Béla | Magyar Életrajzi Lexikon | 24.oldal | 1946.02.06-án született ... |
| Balogh Béla | Ki kicsoda | 33.oldal | Budakeszin nőt fel |

A fenti példák mind olyan több (kettőnél több) argumentumú relációkat (predikátumokat) szemléltetnek, amelyek nem írhatók át közvetlenül és könnyen a triplet-logikába. A kérdés az, hogy hogyan valósítható meg a legáltalánosabb módon a transzformáció. A következő fejezetekben először megmutatjuk, hogy milyen általános, filozófiai megoldást lehet adni a problémára, majd bemutatjuk, hogy a névtérrendszeren belül hogyan lehet azt alkalmazni, milyen következmények fakadnak ebből az „átállásból”.

3.4.1.2.1 A mintát adó filozófiai megoldás: az eseménypéldány fogalma

A fent leírt probléma megoldására használhatunk egy elég régi filozófiai probléma megoldására tett javaslatot (Davidson 1967). Az eredeti filozófiai problémát egy predikátumlogikai példával szemléltethetjük. Vegyük a következő eseményt:

‘János megkeni a kenyerét.’

Ezt fogalmilag egy kétargumentumú predikátummal írhatjuk le:

megken(János, kenyér)

Előfordulhat, hogy többet tudunk és többet is akarunk leírni az eseményről. Ha pontosítani szeretnénk az esemény leírását, azt többféleképpen tehetnénk. Mondhatnánk, például, az alábbiakat:

‘János megkeni vajjal a kenyerét.’

‘János megkeni lassan a kenyerét.’

‘János 2018.05.30-án megkeni a kenyerét.’

Ha bármelyik kiegészítést vagy akár ezek aggregátumát vesszük, akkor mindig új predikátumokat (fogalmi modelleket) kell felvinnünk a modellünkbe, mivel az eltérő argumentumszámú predikátumokat meg kell különböztetni egymástól. A fenti „új” argumentumokat mindet figyelembe véve a „leghosszabb” predikátum így nézne ki:

megken(János, lassan, kenyér, vaj, 2018.05.30)

Ha megengedjük, hogy egy predikátum különböző argumentumszámmal is használható legyen, akkor ennek a megoldásnak az lesz a problémája, hogy, egyfelől előzetesen rögzíteni kell az összes lehetséges argumentumú predikátumot modellben, vagyis előre egy „teljes”, de legalább is a leggazdagabb predikátummodellt kell felépíteni, másfelől ezzel a megoldással nem lehet biztosítani azt, hogy következtetni lehessen a specifikusabb predikátumokból az általánosabb predikátumokra vonatkozóan, pedig ezt nyilvánvaló módon meg kéne tudnunk tenni. Arra, hogy pontosan mit is jelent az utóbbi, nézzünk meg egy példát.

Tegyük fel, hogy az alábbi predikátum tartalmilag igaz.

megken(János, lassan, kenyér, vaj, 2018.05.30)

Ha ez így van, akkor abból nem tudunk következtetni egy másik predikátum igazságára, pedig az érezhetően kapcsolatban van az előzővel.

megken(János,kenyér)

A világtudásunk alapján nyilvánvaló, hogy ha igaz az első predikátum, akkor igaznak kell lennie a másíknak is, logikai eszközökkel azonban ezt nem tudjuk levezetni az eddig alkalmazott megoldások alapján. A következtetésképtelenségünk oka az a tény, hogy a két predikátumot formális okok miatt egymástól teljesen függetlennek kell tartani egyszerűen azért, mert más argumentumszámmal rendelkeznek.

Davidson a következő megoldást javasolta a probléma megszüntetésére. Bevezetjük az eseménypéldányok (e) fogalmát és ezekre vonatkoztatjuk az eseménypredikátumok (relációk) hatókörét. Ha vannak eseménypéldányok, amelyek különböző eseménypredikátumokkal írhatók le (különböző eseményosztályokba tartoznak), akkor:

- az individuális eseményekre lehet kvantorokat alkalmazni
- az individuális eseményeket több eseményosztályba be lehet sorolni
- egy konkrét individuális eseményt úgy lehet leírni, hogy logikai diszjunkcióval összekötjük azokat az eseményosztályokat, amelyek alá besorolható az adott eseménypéldány

A fentiek azt jelentik, hogy az esemény leírásakor az összekapcsolt eseményosztályok mindegyikében szerepelni kell az individuális eseménynek. Nézzük meg, hogy lehet ilyen módon leírni az elemzett predikátumainkat. Első megoldásként azt mutatjuk be, amikor az eredeti, két-argumentumú predikátumból indulunk ki.

$$\exists e(\text{megken}(\text{János}, \text{kenyér}, e) \wedge \text{kenőanyag}(\text{vaj}, e) \wedge \text{kenési_idő}(2018.05.30, e) \wedge \text{kenés_módja}(\text{lassan}, e))$$

A kiindulópont az, hogy feltételezzük, létezik egy 'e' – individuális – esemény, amelyhez tetszőleges számú argumentumot lehet rendelni. Ekkor az eseményt magát úgy írjuk le, hogy olyan predikátumokat képzünk, amelyekbe mindig betesszük az új argumentumot, és mindig mellérendeljük azt az – eseménypéldány – argumentumot, amelynek segítségével a konkrét eseményt tudjuk azonosítani. Ez utóbbi fogja össze az összes argumentumot. Az így dekomponált argumentumszerkezetben olyan argumentumok vannak, amelyek tetszőleges módon diszjunkcióba állíthatók egymással úgy, hogy a megoldás az összes kombinatorikai lehetőséget kezelni képes. Ha egy – összetett – állítás igaz, és valamelyik komponensét elhagyjuk belőle, a „csonka” diszjunkció akkor is igaz marad, ami már azt jelenti, hogy ezzel a megoldással a fent jelzett problémát megszüntetjük.

Davidson javaslatának lényege, hogy mindig három komponenset kapcsoljunk össze úgy, hogy egy komponens minden predikátumban ugyanaz lesz. Ez utóbbi mozzanat teremti meg a kapcsolatot az egyébként egymástól logikai értelemben különböző predikátumok között. Ezt a logikát akkor érvényesítjük teljes körűen a kiinduló példánkban szereplő predikátum esetében, ha magát az alappredikátumot is felbontjuk. Ekkor a következőképpen módosul az előző formula.

$$\exists e (\text{kenő_ágens}(\text{János}, e) \wedge \text{megkent}(\text{kenyér}, e) \wedge \text{kenőanyag}(\text{vaj}, e) \wedge \text{kenési_idő}(2018.05.30, e) \wedge \text{kenés_módja}(\text{lassan}, e))$$

Ez a formula annyiban különbözik az előzőtől, hogy felbontottuk a 'megken(valaki, valamit)' predikátumot két egy-argumentumú (a 'kenő_ágens(valaki), illetve a 'megkent(valamit)') predikátumra, és mindkettőhöz hozzátettük a eseménypéldányra mutató argumentumhelyet. Ez a megoldás annyiban jobb, mint az előző, hogy ekkor már arra is következtetni tudunk, hogy ki volt az, aki megkente a kenyeret, illetve mi volt az, amit megkent János. Az előző megoldásban ez a következtetés még nem volt lehetséges.

Ha a második megoldást alkalmazzuk, akkor az összes predikátum azonos szerkezetűvé válik. Mindig van egy 'tartalmas' argumentumhely, és van egy olyan, amelyben az – összekötő kapcsolatot biztosító – eseményindividuumot azonosítjuk.

3.4.1.2.2 Relációpéldány a triplet-logika szolgálatában

Az előző fejezetben bemutatott, a kortárs filozófiában kidolgozott megoldás logikájával teljesen izomorf a web világában elterjedt, a világtudásunkat RDF-tripletek segítségével leíró modell, amit a Linked Data kezdeményezés is átvett és a gépek közti adatsere szabványos modelljévé tett.

Az előző fejezetben felvázolt megoldás – bizonyos átértelmezésekkel – alkalmazható a névtérbe kerülő relációk, predikátumok esetében is. A korábbi példáinkon segítségével megmutatjuk, hogyan. Vegyük az első példát a választási adatok köréből.

| esemény | település | pártlista | szavazatok száma |
|---------|------------|-------------|------------------|
| 2018 | Szentendre | Fidesz-KDNP | 8384 |

Ez az adatrekord, szövegszerűen, szintagmatikailag jól-formált mondatként úgy lenne megfogalmazható, hogy:

a 2018-as parlamenti választásokon a pártlistán a Fidesz-KDNP pártszövetség Szentendrén összesen 8384 szavazatot kapott

Ez az alábbi – négy-argumentumú – predikátumformátumra írható át.

pártlistás_szavazat(2018, Szentendre, Fidesz-KDNP, 8384)

Ebben a példában pontosan követhetjük Davidson javaslatát, és mondhatjuk azt, hogy egy eseménypéldány felvételével, illetve az alappredikátum dekomponálásával elvégezhetjük a következő transzformációt.

$$\exists e \text{ (választás_ideje(2018, } e) \wedge \text{ választás_helye(Szentendre, } e) \wedge \text{ választás_jelöltje(Fidesz-KDNP, } e) \wedge \text{ választás_eredménye(8384, } e))$$

A településstatisztikai adatok esetében az alábbi rekordot „kellene” átírni.

| esemény | település | adattípus | értéke |
|---------|-----------|----------------------|--------|
| 2015 | Pomáz | munkanélküliek száma | 232 |

A rekord predikátumformátuma:

munkanélküliség(2015, Pomáz, 232)

A predikátum dekompozíciója:

$$\exists e \text{ (munkanélküliség_ideje(2015, } e) \wedge \text{ munkanélküliség_helye(Pomáz, } e) \wedge \text{ munkanélküliség_száma(232, } e))$$

A harmadik példánkban az alábbi filmográfiai adatokat kell átalakítanunk.

| film | személy | szerepkör | karakter |
|------------|--------------|-----------|----------|
| Üvegtigris | Rudolf Péter | színész | Lali |

A rekord predikátumformátuma:

stáb(Üvegtigris, Rudolf Péter, színész, Lali)

A predikátum dekompozíciója:

$$\exists e (\text{film_címe}(\text{Üvegtigris}, e) \wedge \text{film_szereplője}(\text{Rudolf Péter}, e) \wedge \text{film_karaktere}(\text{Lali}, e))$$

Amikor más névterek azonosítóit kell számon tartanunk, akkor azt az alábbi módon tehetjük meg adatbázisban hagyományosan.

| névtér | típus | földrajzi hely | azonosító |
|-------------------|------------|----------------|-----------|
| KSH Helységnévtár | aktuális | Dunaújváros | 33214 |
| KSH Helységnévtár | történelmi | Sztálinváros | 22341 |

A rekord predikátumformátuma:

névtér_azonosító(KSH Helységnévtár, történelmi, Dunaújváros, 22341)

A predikátum dekompozíciója:

$$\exists e (\text{névtér_neve}(\text{KSH Helységnévtár}, e) \wedge \text{azonosító_típus}(\text{történelmi}, e) \wedge \text{földrajzi_hely}(\text{Dunaújváros}, e) \wedge \text{azonosító}(22341, e))$$

Ha a személyek – különböző forrásokban található – életrajzait az alábbi tábla sémája szerint tároljuk, akkor a következő transzformációra lehet szükségünk.

| név | forrás | pozíció | életrajz |
|-------------|--------------------------|----------|-----------------------------|
| Balogh Béla | Magyar Életrajzi Lexikon | 24.oldal | 1946.02.06-án született ... |
| Balogh Béla | Ki kicsoda | 33.oldal | Budakeszin nőt fel |

A rekord predikátumformátuma:

életrajz(Balogh Béla, Magyar Életrajzi Lexikon, 24. oldal, 1946.02.06-án született ...)

A predikátum dekompozíciója:

$\exists e$ (személy(Balogh Béla, e) \wedge forrás(Magyar Életrajzi Lexikon, e) \wedge pozíció(24.oldal, e) \wedge életrajz(1946.02.06-án született ..., e)))

A fenti transzformációs példák mintájára tetszőleges számú argumentummal rendelkező predikátumot át lehet írni a triplet-logika elvárásainak megfelelően, csak az a kérdés, vannak-e olyan szempontok, amelyek miatt nem teljesen egységesen és nem teljes körűen érdemes az RDF-kompatibilitást megteremteni.

Mielőtt a triplet-logika alkalmazásának hatókörét tárgyalnánk, bemutatjuk azt a megoldást, amellyel az RDF-állításokat kezeljük a névtérrendszeren belül. Ebben a részben támaszkodni fogunk pár olyan állításra, amelyet a később fejezetekben fejtünk csak ki bővebben, mert az a névtérrendszer nyelvi modelljével kapcsolatos.

A névtérrendszeren belül elkülönítettünk egy olyan nyelvi sémát, amelyben a nyelvi konstrukciókra vonatkozó alapadatokat tároljuk. Ebben a sémában van „hely” az általános fogalmak megragadására (a 'concept' tábla), ahová a fogalmak jelentéseit, referenciáit lehetne direkt módon beírni (valamilyen formális nyelv alkalmazásával). Ezt a jelentésepítést azonban nem végezzük el a projekt keretében, mert ez a feladat egyfelől nem tartozik a projekt célrendszerébe (ami egyébként igen nagy munkát kívánna meg), másfelől a névtérrendszernek a jelenlegi szakaszában erre nincs is szükség. Ebbe a táblába csak általános fogalmakat írunk, az individuumokra referáló individuális „fogalmak” tárolását (és azonosítását) másként oldjuk meg. Általános fogalmak lehetnek nevek is, relációk is, sőt, bármi más, általános célokra alkalmazott nyelvi konstrukciókat reprezentáló fogalmak is.

A tudásszervezési rendszerekben tárolt világtudás is csak indirekt módon képes „érzékelteni” a fogalmak jelentéseit azáltal, hogy számon tartja a fogalmak egymáshoz való viszonyát a különböző relációkon keresztül. Az ilyen kapcsolatokat a nyelvi sémában a 'concept_concept' táblában helyezük el, aminek rekordjai azonnal átfordíthatók klasszikus RDF-állításokra. Ezen a módon azonban csak az általános fogalmak közti általános kapcsolatokat kezeljük. Azért fontos ezt hangsúlyozni, mert az itt és így tárolt világtudás mindig csak valamilyen általános, sematikus, predikátum-jellegű tudást jelent, mindenfajta individuális – tér- és időbeli – kötöttség nélkül. Itt tehát predikátumtípusok (névtípusok, relációtípusok) vannak, amelyek önmagukban nem elégségesek ahhoz, hogy a névtérrendszeren belül rögzítendő individuális állításokat egyértelműen leírassuk. Utóbbihoz arra van szükség, hogy legyenek – a predikátumtípusok példányaiként értelmezhető – predikátumindividuumaink.

A predikátumindividuumokat önálló táblában vesszük fel, és mindig összekapcsoljuk őket a megfelelő predikátumtípussal. Ez a predikátumpéldány megfeleltethető a davidsoni elméletben kulcsszerepet játszó eseménypéldánynak. Ezzel tudjuk összekapcsolni a többargumentumú predikátumok dekompozíciója során kialakított komponenseket egymással. Nézzünk meg egy példát erre, vegyük a korábban már elemzett esetet a választási adatokról! Akkor a dekomponálás eredményeként a következő formulát kaptuk – predikátumlogikai szinten.

$\exists e$ (választás_ideje(2018, e) \wedge választás_helye(Szentendre, e) \wedge választás_jelöltje(Fidesz-KDNP, e) \wedge választás_eredménye(8384, e))

A kérdés itt már csak az, hogy ezt hogyan lehet adatbázisban reprezentálni. A fenti kompozíció négy darab, két-argumentumú predikátumból áll. Utóbbiakat egyértelműen át lehet forgatni RDF-tripletekbe az alábbiak szerint.

RDF(választás_ideje, 2018, e)

RDF(választás_helye, Szentendre, e)

RDF(választás_jelöltje, Fidesz-KDNP, e)

RDF(választás_eredménye, 8384, e)

A fenti állításokban a tripletek utolsó tagja mindig arra a predikátumpéldányra mutat, amit felvettünk a predikátumpéldány táblába azért, hogy össze tudjuk kapcsolni a komponenseket egymással. Az RDF-állítások első paramétereként mindig egy általános fogalmat leíró predikátumtípust kell megadnunk. A középső komponensbe kerülnek azok a konkrét értékek, amelyek az adott predikátumtípus adott példányához tartoznak.

Fontos szempont, hogy a predikátumpéldányok felvételekor mindig meg kell adnunk azok típusát, mert csak az utóbbi ismeretében tudjuk meghatározni később, hogy az individuális adatokat milyen kompozíciós szerkezetbe kell beírni. Ez a kompozíciós szerkezet általános információ, ezért ezt tárolhatjuk a 'concept_concept' táblában, ahol a predikátumtípusok közötti kapcsolatokat írjuk le. Itt köthetjük össze az eredeti, sok-argumentumú predikátumtípust és a dekomponálás eredményeként létrehozott két-argumentumú predikátumtípusokat egymással.

* * *

A névtér projekt számára volt és van egy külső elvárás, hogy ti. mind modellezési, mind fejlesztési kérdésekben minél jobban igazodjunk a webes trendekhez. Ezen elvárások egyike azt fogalmazza meg, hogy a névtér adatsémájával minél jobban kompatibilissé váljunk a triplet-logikán alapuló modellhez. Ezt a gyakorlati igényű és szempontú elvárást támogatni lehet a fent bemutatott, elméleti alapvetéssel, csak az általános megoldás aktualizálni kell a névtér környezet sajátos elvárásaihoz.

A névtér általános modellje felépíthető úgy, hogy – ha nem is teljesen, de – nagy mértékben kielégítsük az RDF-logika elvárásait. Ahol eltérünk az triplet-megoldásoktól, ott is könnyen a transzformáció. Ahhoz, hogy a tényleges megvalósítás módja, értelme láthatóvá váljék, szükség van egy olyan kiterőre, amelyben bemutatjuk a névtérrendszer nyelvi modelljét, és ezen belül tisztázzuk a névtérekben és a KOS-ban tárolt tudás egymáshoz való viszonyát.

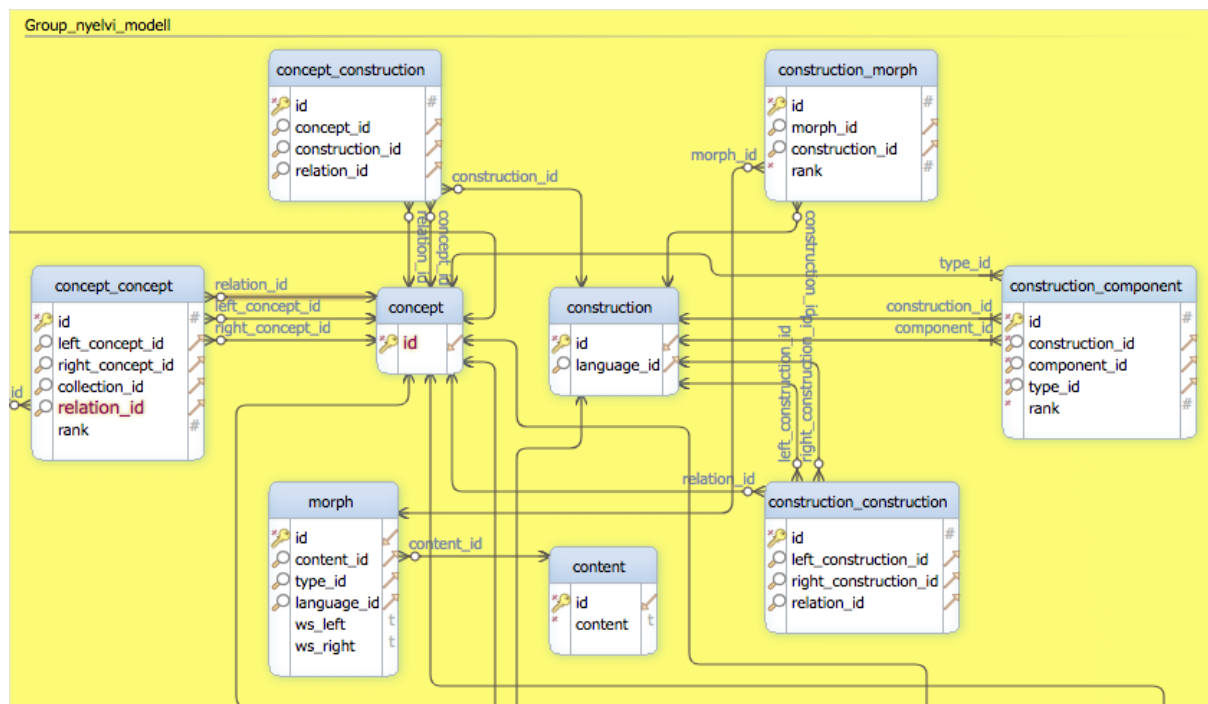
3.4.1.3 A névtérekben és KOS-ban tárolt tudás viszonya

A névtérrendszeren belül önálló sémában helyezzük el minden nyelvi információt. Ennek az a legfontosabb indoka, hogy egyetlen helyen legyen tárolva minden nyelvi adat. Ennek érdekében – a nyelvi sémába beépítünk minden rendelkezésre álló – különböző nyelvi adatforrásokból beszerezhető – nyelvi információt, a lehetséges karakterláncok, a morfémákat, az ezekből képezhető konstrukciókat.

A névtérrendszer számára szükséges tudásszervezési rendszereket is a nyelvi sémába építjük, hiszen ezek is nyelvi erőforrások. Ahhoz, hogy pontosan definiálni tudjuk, hogy miként lehet hasznosítani a névtéren belüli keresések során a KOS-okban tárolt tudást, először be kell mutatnunk, hogy milyen logika alapján épül fel egyfelől a névtér általános modellje, illetve másfelől a nyelvi séma, és ezután térhetünk ki arra, hogy a KOS-ban tárolt tudást hogyan lehet kiszedni a rendszerből.

3.4.1.3.1 Névtérmodell, nyelvi modell

A nyelvkezelés ismertetését kezdjük azzal az ábrával, amely mutatja a nyelvi adatok modelljét a névtérrendszer elkülönített sémáján belül.



a nyelvi szegmens sémája a névtéren belül

A nyelvi modellben négy nyelvi entitást különítünk el: a tartalom, a morféma, a konstrukció és a fogalom adattípusát, és természetesen kezelnünk kell a közöttük értelmezhető kapcsolatokat is.

A **tartalom** ('content') entitásban tároljuk azokat az elemi nyelvi egységeket (szavakat, morfémákat, toldalékokat, írásjeleket stb.), amelyekből valamilyen nyelven valamilyen nyelvi konstrukció felépíthető. Ebben a táblában olyan karaktersorozatokat építhetünk, amelyekben nincs szóköz, és nincsenek középpontozási jelek, illetve fontos még, hogy ebben a táblában kisbetűvel írunk le minden karaktert. Ehhez a szinthez még nem rendelünk nyelvi kötést. Ha egy adott karakter(sorozat) több nyelven is használatban lehet, akkor is csak egyszer szerepelhet itt. Azt sem tudhatjuk még ezen a szinten, hogy milyen morfológiai jellemzői, illetve milyen jelentése lehet egy adott nyelven az adott karaktersorozatnak. Nézzünk pár példát:

eleven

gyula

vár

A **morféma** ('morph') entitás szintjén megadhatjuk egy adott tartalmi egység morfológiai jellemzőit (egy kiválasztott nyelven). A morféma szintjén már nyelvet kell a content-hez rendelni, és mindenképpen meg kell adni annak szófaji besorolását, valamint a nyelvi konkatenálás során várható viselkedését. Ezen a szinten még mindig nem tudunk jelentést kezelni, de tudjuk, hogy a morféma hogyan „viselkedik” nagyobb nyelvi konstrukciókba ágyazva. Ha egy nyelvi egység azonos alakú, de eltérő morfológiai minőségű, akkor azok mind különböző morfémák lesznek. A morfémák között szerepelnek a toldalékok, írásjelek is, amelyekre az jellemző, hogy önállóan nem jelenhetnek meg egy nyelvi megnyilatkozáson belül, csak az „önálló” megjelenésre képes szavakhoz illeszthetnek. Az előbbi példánk alapján a következő rekordok kerülnek be a morféma entitásba.

eleven (angol, számnév)

eleven (magyar, melléknév)

vár (magyar, ige)
vár (magyar, köznév)
Gyula (tulajdonnév)
gyula (köznév)
k (többszám jele)
k (betű)
t (tárgyrag)
- (kötőjel)

A harmadik szinten a **konstrukció** fogalma áll ('construction'), ami a tényleges nyelvi kommunikációban használt nyelvi egységet írja le. Minden nyelvi megnyilatkozás konstrukciókból tevődik össze. Egy konstrukció két forrásból „táplálkozik”. Egyfelől morfémből áll össze, de itt már megengedhető, hogy a konstrukció több szóból álljon. Másfelől a konstrukciónak itt tulajdoníthatunk jelentést, amit a konstrukció a fogalom ('concept') entitástól vehet át. A konstrukciók többféle módon állhatnak elő: állhatnak egyetlen morfémből (szótövek esetén), össze lehet őket fűzni több morfémből is őket (összetett vagy toldalékolt szavak esetén), de lehetnek összetett konstrukciók, amikor egy konstrukció több egy-morfémás konstrukcióból tevődik össze.

árva (magyar, köznév: szótő, egyelemű, egy-morfémás konstrukció)
árvákat (magyar, köznév: toldalékolt szó, egyelemű, több-morfémás konstrukció)
Árvákat Támogató Alapítvány (magyar, tulajdonnév: testületi név, összetett konstrukció)
Petőfi Sándor (magyar, tulajdonnév: személynév, összetett konstrukció)
Gyula (magyar, tulajdonnév: keresztnév, egyelemű konstrukció)
Gyula (magyar, tulajdonnév: családnév, egyelemű konstrukció)
gyula (magyar, szótő, köznév: méltóságnév, egyelemű konstrukció)
Gyula (magyar, szótő, tulajdonnév: földrajzi név, egyelemű konstrukció)

A konstrukciókat „két lépésben” építhetjük fel. Az egyelemű konstrukciókat a morfémből illeszthetjük össze ('construction_morph'), az összetett konstrukciókat pedig az egyelemű konstrukciók összefűzésével kaphatjuk meg ('construction_component').

Mindig meg kell adni, hogy a konstrukció milyen nyelvhez kapcsolódik. Az esetek döntő többségében a konstrukció nyelve megegyezik a morféma (vagy morféma) nyelvével, amit lehetne „örököltetni”, azonban bármikor előfordulhat olyan eset, amikor egy összetett konstrukción belül „más nyelvű” komponens is szerepel, ezért az összetett konstrukciók esetében nem lehet a nyelvi kapcsolatot származtatott értéként kezelni. Példa rá:

Klement Gottwald Villamossági Gyár (magyar, összetett)
Klement (német, egyelemű)
Gottwald (német, egyelemű)
Villamossági (magyar, egyelemű)

Gyár (magyar, egyelemű)

A fenti összetett konstrukció első két tagja német, míg a második kettő magyar – egyelemű – konstrukció, de ettől még az egész konstrukció nyilván magyar, de ezt a minősítést nem tudjuk levezetni a komponensek „tulajdonságaiból”, ezért ezt külön meg kell adni.

A konstrukciók jelentését a fogalom ('concept') táblából származtathatjuk: minden konstrukció esetén meg kell adni, hogy milyen fogalomhoz kapcsolható ('concept_construction'). A névkonstrukciók közül a köznevek, illetve a tulajdonnevek más relációval kapcsolódhatnak össze egy fogalommal. Egy köznév a hozzá kapcsolódó fogalom jelentését reprezentálja valamely nyelven. Ugyanazt a fogalmi jelentést egy másik konstrukció is reprezentálhatja. Ha ugyanazon a nyelven használt konstrukciókról van szó, akkor szinonimitásról beszélünk, eltérő nyelvű konstrukciók esetében pedig egymás fordításainak tekinthetjük őket.

kutya és eb (magyar és magyar, szinonimitás)

kutya és dog (magyarul és angol, fordítás)

A tulajdonnevek – definitív minőségüknek köszönhetően – másfajta relációval kapcsolódnak a fogalmakhoz. Saul Kripke tézisének elfogadva mondhatjuk (Kripke 2010), hogy a tulajdonnevek merev jelölők, vagyis egyetlen funkciójuk az, hogy referáljanak valamilyen individuális entitásra. Ebből levonhatjuk azt a következtetést, hogy a tulajdonnevek valamilyen névtípus (személynév, testületi név, földrajzi név stb.) példányainak tekinthetők. Igaz ugyan, hogy ezek a nevek is reprezentálnak valamit (a referált individuumot), de lényegesebb az a tulajdonságuk, hogy egyedi jellegük, az a minőség, hogy tovább már nem bonthatók, nem képezhető belőlük újabb fogalom. Ezzel élesen szemben állnak a köznevek, amelyek esetén mindig fennáll az az elvi lehetőség, hogy az általuk reprezentált fogalom alá – a generikus reláció segítségével – újabb fogalom (illetve az azt reprezentáló konstrukció) képezhető. A fogalmak jelentését kétféle módon adhatjuk meg. Először a saját ontológiai elköteleződéseinknek megfelelően valahogy definiálhatunk csúcsgogalmakat, azaz létrehozhatunk egy csúcsontológiát, majd az így alkotott, legfelső szintű fogalmak segítségével alárendelt fogalmakat képezve újabb fogalmakat definiálhatunk úgy, hogy kifejezzük ezek viszonyát egymáshoz, illetve a csúcsgogalmakhoz. A nyelvi konstrukciókkal kifejezhető világtudást a fogalmak közti kapcsolatok segítségével írhatjuk le. Ezt a tudást írhatjuk be a 'concept_concept' táblába, ahol két fogalmat kapcsolhatunk össze egy relációval (ami szintén a fogalom!). Ez a reláció (tábla) alkalmas arra, hogy a tudásszervezési rendszerekben felépített világtudást kezeljük.

A nyelvi modellben van még egy tábla ('construction_construction'), ami a konstrukciók közti viszonyt írja le. Amikor van két tulajdonnevünk, és ezek között létezik valamilyen kapcsolat, amit nem lehet kifejezni a fogalmakhoz való viszonyuk alapján, akkor lehet erre a táblára szükség. Mutatunk egy példát erre:

Katalin IS-INSTANCE-OF keresztnév (tulajdonnév)

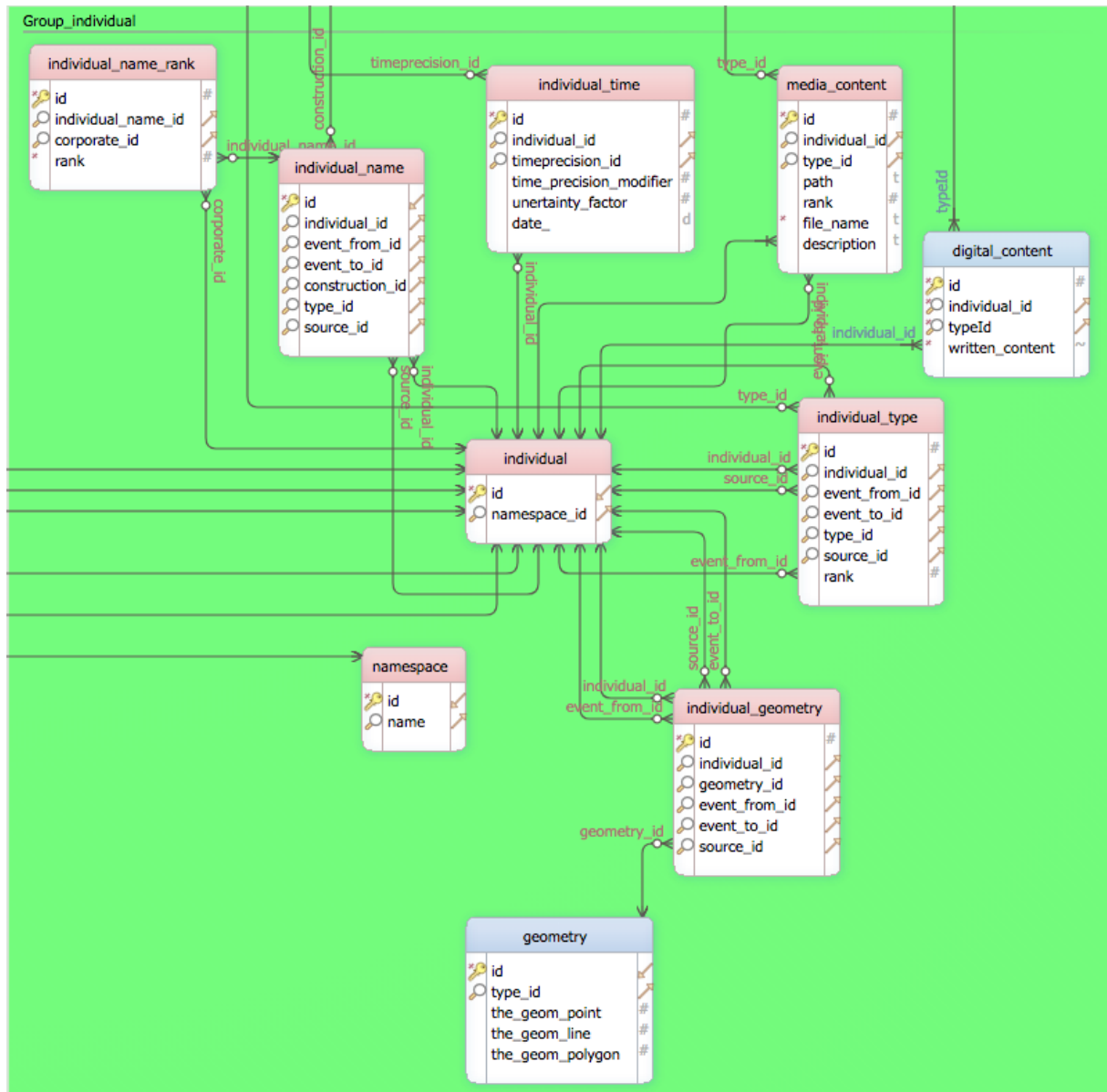
Kati IS-INSTANCE-OF becenév (tulajdonnév)

Kati IS-BECENÉV-OF Katalin

Mind a 'Katalin', mind a 'Kati' női személynév, amelyekről a fogalmi szinten annyit tudunk mondani, hogy ezek példányai valamilyen tulajdonnévnek, de ez a tudás nem fedi le azt a tudást, hogy a 'Kati' a 'Katalin' beceneve. Ennek rögzítésére szükséges a konstrukciók közti kapcsolatot leíró tábla.

A névtér központi részében kezeljük azokat az individuális entitásokat, amelyek a névterek lényegét alkotják: a különböző névtípusok alapján elkülönített névterekbe tartozó individuális megnevezéseket (tulajdonneveket), a nevekkel referált individuális névhordozókat, valamint azokat az individuális entitásokat, amelyek egyes névhordozók jellemzéséhez szükségesek lehetnek.

A nevekre lehetnek példák a 'Petőfi Sándor', Sztálinváros', 'Magyar Szocialista Munkáspárt' konstrukciók, a névhordozókra a 'Petőfi Sándor'-ként ismert magyar költő, az 1950-es években 'Sztálinváros'-ként hivatkozott magyarországi város, a 'Magyar Szocialista Munkáspárt' névvel illetett állampárt 1956 és 1989 között. Az azonosítás miatt szükséges individuális adatokra hivatkozhatunk példaként individuális eseményekre: mikor született, mikor halt meg egy adott személy, mikortól meddig volt valamilyen neve egy adott földrajzi helynek vagy egy testületnek stb. ('individual_time'), a földrajzi helyek azonosításához pedig igénybe vehetjük azokat a földrajzi-geometriai adatokat, amelyek segítségével lokalizálni tudjuk a geoindividuumokat a földrajzi térben ('geometry').



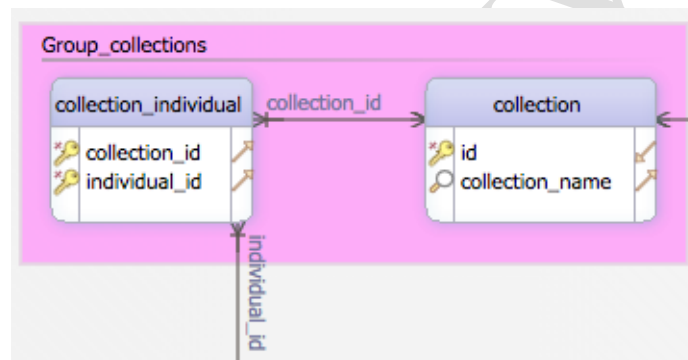
a névtér magjában levő entitások

Minden (tulajdon)névtérbe kerülő individuális entitásnak kötelezően meg kell adnunk azt, hogy milyen névtérbe ('namespace') tartozik, illetve milyen típusba, milyen általános fogalom alá ('individual_type') sorolható. A típusmegjelölésből felvehetünk több rekordot is, ha arra egy adott névtér esetében szükség van. A névhordozók és megnevezések (nevek) közti pontos megkülönböztetés kedvéért külön

azonosítjuk a névhordozókat és a megnevezéseket (neveket) az 'individual', illetve 'individual_name' táblák segítségével.

A névterek közé fel lehet venni dokumentumokat (filmeket, képeket, könyveket, zeneműveket stb.), így ezek példányai is bekerülhetnek a névhordozók közé, és lehet neveket (címeiket) adni nekik. A dokumentumok tartalmainak elérhetőségét kétféle módon érdemes kezelnünk, hogy igazodhassunk a különböző dokumentumtípusok eltérő vonásaihoz. Mivel az audiovizuális dokumentumokat (képeket, filmeket, hanganyagokat) nem akarjuk adatbázisban tárolni, így esetükben arra van szükség, hogy meg tudjuk adni azt, hogy ezek hol érhetőek el a fájlrendszer szintjén ('media_content'). A szöveges dokumentumokat (vagy legalább is azok egy részét) szeretnénk adatbázisban tárolni azért, hogy a névtérhez kapcsolódóan többféle módon is meg tudjuk jeleníteni, kereshetővé tenni azokat. E célból ezeket a szövegegyeségeket külön tároljuk ('digital_content').

A névtér magjában individuális entitásokat, illetve ezek individuális megnevezéseit (neveit) kezeljük. Előfordulhat olykor, hogy nem konkrét individuumokat, hanem ezek egy csoportját kell megragadnunk. Ennek kezelésére a névtérrendszer egyik főkategóriáját, a 'gyűjtemény' ('collection') fogalmát alkalmazhatjuk.³⁸ A gyűjtemény is individuális entitás, akárcsak az elemei, viszont a gyűjtemény és individuum más főkategóriába tartozik, amit kifejez az a pusztán tény, hogy köztük egy aszimmetrikus kapcsolat, az 'eleme' reláció vehető fel.³⁹



gyűjteményentitás a névtérben

A különböző típusba tartozó individuális entitásoknak definiálni lehet gyűjteményeket, és bizonyos helyzetekben ezeket „keverten is lehet használni. Vegyük példaként a parlamenti választások adatkörét, ahol a szavazók pártlistákra adhatják le a voksaikat, amikor a pártlista jelenthet egyetlen pártot, de olykor többet is. Utóbbi esetekben mindig gyűjteményes fogalmat kell alkalmaznunk. Másik példa lehet a gyűjteményekre a személyek esetén alkalmazható család fogalom is.

Fidesz – KDNP pártlista

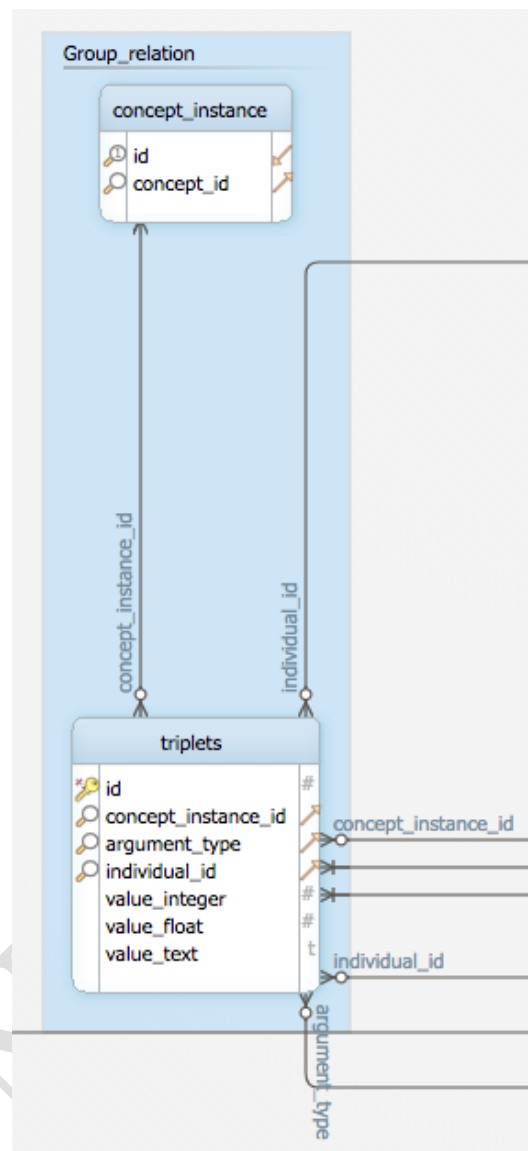
Rákóczi család

Az első példában két individuális testület (a 'Fidesz' és a 'KDNP') együttesen, gyűjteményként hivatkozható, hiszen nem külön-külön szavaztak rájuk, de nyilván fontos az a tudásunk is, hogy az adott pártlista két konkrét, individuális testület „összekapcsolásából” áll. A második példa a könyvtári világ számára lehet fontos, ahol a személyek mellett elkülönítve kezelik a családokat, amiket megint csak gyűjteményes fogalomként tudunk pontosan megragadni.

³⁸ A gyűjtemény (és néhány más, fontos, ide kapcsolódó) fogalomról a jelentés egy másik fejezetében bővebben írunk.

³⁹ A főkategóriákról szóló fejezetben foglalkozunk a köztük definiálható kapcsolatokkal is, és ott „élesen” elkülönítjük egymástól a 'példánya' és az 'eleme' relációt.

A névterekben kezelt individuális entitások közti kapcsolatokat egyetlen – a RDF-triplet logikához igazodó – táblában rögzítjük.



a névtér-kapcsolatok kezelése RDF-tripletekben

Az individuális kapcsolatok kezelésének módját alátámasztó elméleti megfontolásainkat a kutatási jelentés egy másik fejezetében bontjuk ki. Minden individuális entitások közti kapcsolat individuális relációnak minősíthető, amit egy relációpéldány vagy fogalompéldány ('concept_instance') táblában veszünk fel, és ezzel az individuális relációpéldánnyal kötjük össze a kapcsolat minden elemét. Ahány argumentumú kapcsolatról van szó, annyi rekordot kell felvenni az RDF-szerű mondatok leírásához ('triplets'). Azt, hogy hányas kötésben kell leírunk egy ilyen kapcsolatot, onnan lehet tudni, hogy az általános fogalmak szintjén megadjuk, hogy valamely predikátum (fogalom) hány argumentummal rendelkezik. Ezt a nyelvi séma 'concept_concept' táblájában rögzíthetjük.

3.4.1.3.2 A nevek kezelése a nyelvi modellben

A névtéradatrendszeren belül élesen elválasztjuk egymástól a tulajdon- és köznévteteket, aminek tükröződnie kell az alapsémában is. Mivel a kétféle névtér közti különbség arra vezethető vissza, hogy milyen (általános vagy individuális) kategóriába tartozó entitásokat kezelünk bennük, ezt le kell

képeznünk a névtér modelljében. Ezt úgy valósítjuk meg, hogy az egyes névtérbe tartozó entitások adatait külön szegmensben tároljuk. Mindez azt is jelenti, hogy magukat az entitásokat is külön azonosítjuk.

A közneveket – a névterektől teljesen elkülönített – nyelvi sémában kezeljük. Ennek a sémának alapvetően az a fő funkciója (és értelme), hogy minden releváns nyelvi információt itt (és csak itt) tároljunk. Mind a köznevek, mind a tulajdonnevek nyelvi konstrukciók, amelyek morfológiai, szintaktikai tulajdonságait, valamint a fogalmi kötését, fogalmi kapcsolódásait hasonlóan kell kezelni. Egy korábbi kutatási jelentésben már bemutattuk, hogy a tulajdonnevek milyen minőségükben és hogyan térnek el a köznevektől (a tulajdonnevek referenciája mindig – értsd: minden lehetséges világban – ugyanaz, és erre azt mondják, hogy a tulajdonnevek merev jelölők. Egy másik fontos tézisünk volt, hogy a tulajdonnevek nem önmagukban „válnak képessé” erre a szerepre, hiszen a tulajdonneveknek nincs semmi olyan különleges vonása, amelynek alapján – ebben a dimenzióban – meg tudnánk különböztetni őket a köznevektől. Saul Kripke elméletre támaszkodva viszont azt mondhatjuk (Kripke 2010, hogy nem a tulajdonnevek azok, amelyek révén megvalósul a merev jelölés, hanem az a megnevezési aktus, amely a nevet az adott dologhoz (névhordozóhoz) rendeli, illetve az a névhasználati gyakorlat, amely folyamatosan fenntartja azt a közös tudást, hogy a dolog és a név össze van kapcsolva egymással. A név helyett ezért használja Kripke a **megnevezés** kategóriáját, ami egy esemény jellegű fogalom, és ebben a minőségében sokkal jobban illeszkedik az adott névhasználói közösség folyamatos megnevezési gyakorlatához. Erre a megkülönböztetésre már csak azért is szükségünk lehet, mert amikor a tulajdonneveket hozzá akarjuk rendelni az általuk reprezentált dolgokhoz (a névhordozókhoz), akkor arra van szükségünk, hogy meg tudjuk adni, hogy ez a kapcsolat mettől meddig van van érvényben. Ezt a fajta időbeliséget, időhöz kötöttséget azonban nem tudjuk sem a névhordozóhoz, sem a tulajdonnévhez kapcsolni. A névhordozóhoz azért nem, mert annak nyilván nem lehet inherens tulajdonsága a neki tulajdonított név, magát a nevet pedig azért nem láthatjuk el időbeliséggel, mert a neveknek nem inherens tulajdonsága az idő. A megnevezés viszont, ami az események kategóriájába tartozik, már könnyen és lényege szerint kapcsolható az időhöz.

A köznevek és a tulajdonnevek is nyelvi konstrukciók, tehát a nyelv elemi egységeiből, a morfémákból kell felépíteni őket. A morfémák mindig adott nyelven belül értelmezhetők, és a nyelvre jellemző morfológiai, szintaktikai szabályok mentén lehet belőlük összetett morfémákat, kifejezéseket létrehozni. A nyelvi konstrukciók „tartalmát” azáltal nyerik el, hogy a nyelvhasználói közösség valamilyen jelentéseket, referenciákat tulajdonít nekik, ami viszont független a nyelvtől. A következő konstrukcióknak teljesen eltérő a szintaktikai, morfológiai alakja, mégis ugyanaz a jelentése, referenciája a különböző nyelveken:

dog (angol)

kutya (magyar)

eb (magyar)

hunt (német)

A konstrukciók alakját, formátumát a morfémák segítségével felépítve szükség van arra is, hogy jelentést, referenciát tudjunk hozzájuk rendelni. Ezt a névtér modelljében úgy oldjuk meg, hogy definiálunk egy 'concept' entitást. Ezzel azokat a jelentéseket próbáljuk megragadni, jelezni, amelyek a különböző nyelvű – konstrukciók referenciáit írják le. A fenti konstrukciók mindegyike a 'kutyaság' fogalom terjedelmébe eső dolgokra (a kutyákra) referál, tehát ezeknek a konstrukcióknak a – nyelvi modellben felvett – azonosítói mind a 'concept' entitás 'kutyaság'-ot jelző rekordjára. Ezt a 'concept_construction' relációval fejezhetjük ki. Ebben a kapcsolatban ugyanarra a 'kutyaság' fogalomra hivatkoznak a különböző nyelven felvett konstrukciók ('kutya', 'dog', 'hunt'), de ugyanoda mutatnak az azonos nyelvű, eltérő alakú, de megegyező jelentésű konstrukciók is ('kutya' és 'eb'). A fogalmak és konstrukciók közti kapcsolatban jelölni kell a kapcsoló relációt is, mert ez névtípusonként változhat.

Mondhatjuk, hogy a nyelvi konstrukciók a – nyelvfüggetlen – 'concept' entitástól nyerik el a jelentésüket.

A tulajdonnevek és köznevek közti különbség abban is megnyilvánul, hogy ezek másfajta módon, másfajta reláció mentén kapcsolódnak a 'concept' entitáshoz. A köznevek mint konstrukciók – bizonyos értelemben – reprezentálják a fogalom jelentését. Emiatt – egy adott nyelvet tekintve – minden olyan konstrukció, ami nem szinonim egy másikkal, szükségképpen más fogalomra mutat. A szinonim konstrukciók mutathatnak ugyanarra a fogalomra, ebben az értelemben a konstrukciók száma nagyobb, mint a fogalmaké, de azt azért mondhatjuk, hogy a nagyságrendjük durván megegyezik.

A tulajdonnevekkel kicsit más a helyzet. Ha egy tulajdonnév valamely konkrét individuumra referál, akkor mondhatjuk azt, hogy nincs igazán tartalmas jelentése. Emiatt a tulajdonnév mint nyelvi konstrukció és a vele összekapcsolható fogalom között másfajta relációt kell érvényesítenünk. Mivel a tulajdonnevek adott típusú (tulajdon)név példányaiként értelmezhetők, ezért a konstrukciók és fogalmak közti kapcsolatra azt kell mondanunk, hogy az a 'példánya' relációval fejezhető ki.

Ez a „pici” különbség a tulajdon- és köznevek között nagyon nagy változást okoz a számosságban. A tulajdonnevek és a hozzájuk kapcsolható fogalmak között több nagyságrendi különbség van. Ez abból fakad, hogy a példa reláció mentén csak annyi különböző fogalomhoz kapcsolódhatnak a tulajdonnevek, amennyi tulajdonnévtípust definiálunk. A 'földrajzi név' mint fogalomnak magyar nyelven is többszáz ezer példa lehet, amit úgy tudunk megragadni, hogy az összes földrajzi nevet mint konstrukciót az egyetlen 'földrajzi név' fogalomhoz kötjük a 'példánya' relációval. Ugyanilyen vagy legalább is hasonló mértékű számossági aszimmetria van a többi tulajdonnévtípus esetében is. Keresztnévből relatíve kevesebb van. Lehet azon vitatkozni, hogy hány ezer, de mindenképpen kevesebb, mint tízezer. A jelenlegi ismereteink alapján a családnevek százazres nagyságrendben lehetnek a magyar nyelvben, a személynevek számosságára (a család- és keresztnévek kombinációira) egyelőre nem tudunk becsléseket mondani, de könnyen elképzelhetőnek tartjuk, hogy ez a szám elérheti a milliós nagyságrendet. Vélhetőleg a testületi név példányainak számossága a legnagyobb, de erre vonatkozóan még nincs megbízható számszerű adataink. A személyek, a személynévtér esetéhez hasonlóan nincs arra szükség, hogy egy nemzeti testületi névtérbe minden valaha létezett, illetve az összes most létező testületet felvegyünk, de ha egy erőteljes szűrést feltételezünk, akkor is a sokszázazres nagyságrendet valószínűsíthetjük azon testületek számosságára vonatkozóan, amelyek bekerülhetnek a testületi névtérbe. A testületi nevek számosságát befolyásolja az a tény, hogy az ilyen nevek kiosztásának más szabályai vannak a személynevekhez képest. Amíg a személyneveknél nincs olyan elvárás, hogy a nevek legyenek országosan különbözők, addig a testületi nevekre nézve elég régóta van egy ilyen elvárás.

Összegezve a fentieket: néhány – az általános fogalom alá tartozó 'névtípus'-hoz ('személynév', 'családnév', 'keresztnév', 'földrajzi név', 'testületi név' stb.) tehát több ezer, több tízezer, többszáz ezer névpéldány tartozik. Ha csak a konstrukciók szintjén vizsgáljuk a számossági kérdést, akkor kijelenthető, hogy az individuális entitásokra referáló tulajdonnevekből több van, mint az általános fogalmakra referáló köznevekből, igaz, amikor a szintagmatikus tengely mentén komponálunk meg egy nyelvi konstrukciót, akkor a világtudásunk nyelvi reprezentálására sokkal több új konstrukciót alkothatunk, mint amennyit a nyelvi rendszerek alapjait jelentő szótárakba, tudásszervezési rendszerekbe felvesszünk.

* * *

Amiatt, hogy a közneveket és tulajdonneveket – többféle értelemben is – eltérő módon kezeljük a nyelvi modellben, nem egyértelmű, hogy miként lehet kezelni bennük azokat a tudásszervezési rendszereket, amelyekben egyaránt vannak általános és individuális fogalmak. A következő fejezetben ezt a kérdéskört vizsgáljuk meg.

3.4.1.3.3 A tudásszerkezési rendszerek kezelése a nyelvi modellben

Amikor azt a kérdést vizsgáljuk, hogy a tudásszerkezési rendszerekben rögzített világtudást hogyan lehet használni a névtérrendszeren belüli keresés támogatására, akkor pontosan tisztázni kell a tudásszerkezési rendszerek és a névterek viszonyát. Ehhez az alábbi tézisekből indulhatunk ki.

- Korábban már kifejtettük, hogy a tudásszerkezési rendszerek – a típusuktól függő – rögzített sémával rendelkező adatbázisként értelmezhetők.
- Az előző fejezetekben bemutattuk azt, hogy a tudásszerkezési rendszerek adatait a nyelvi sémában rögzítjük.
- Eddig nem foglalkoztunk vele, de könnyen belátható, hogy a tudásszerkezési rendszerek adatai gyakorlatilag szinte automatikusan áttehetőek RDF-sémákba, hiszen a KOS-ok általános szerkezete illeszkedik a triplet-logikához (mindig két terminust kapcsolnak össze egy relációval).
- A névterek között elkülönítettük a köznévtereket és a tulajdonnévtereket egymástól azon az alapon, hogy milyen neveket (közneveket vagy tulajdonneveket) tartalmaznak, aminek persze további fontos következménye volt az is, hogy a hozzájuk kapcsolható relációk típusában is eltértek egymástól: a tulajdonnévterek legfontosabb relációja a 'példánya' reláció, ami egyfelől azokat a egyedi és konkrét megnevezéseket fogja egybe, amelyek révén a névhordozókhoz neveket rendelünk, másfelől az adott típusú névhordozók előfordulásait kapcsolja az adott névtérhez.
- A tulajdonnévterek legbelső magjának leírására a terminuslista (mint a tudásszerkezési rendszerek egyik típusa) alkalmas. A tulajdonnévterek belső magja terminuslista.
- A tulajdonnévterek példányainak egyértelmű azonosításához szükséges adatok, azok az adatok, amelyek már kívül esnek a szűk értelemben vett tulajdonnévterek határain, és különösen azok az adatok, amelyek már a névterek közti kapcsolatokat írják le, egyre inkább „igénylik” azt, hogy általános fogalmakkal reprezentálható világtudás alapján értelmezzük őket. Ezt az igényt a névtér modelljében úgy próbáljuk meg kielégíteni, hogy a elkülönítjük a tulajdonnévterek individuális példányaira vonatkozó adatokat, illetve az ezek leírására szolgáló – általános világtudást reprezentáló, a nyelvi modellben kezelt – adatokat egymástól.

A nyelvi modellben kezeljük tehát az általános fogalmakat, valamint a köztük létrehozható relációtípusokat, amelyeket a 'concept' entitásban lehet felvenni. Ide minden általános fogalmat be lehet írni, ami a nyelvi konstrukciók jelentésének kifejezéséhez szükséges lehet. Ez a szempont önmagában még nem „teremt meg” azt az igényt, hogy az általános fogalmak között legyen valamilyen struktúra felépítve. A személynévtéren belül a foglalkozások kezeléséhez szükséges van egy foglalkozási klasszifikációra, a földrajzi névtéren belül szükség van a földrajzi helyek tipológiájára, a könyvcímtéren vagy a filmcímtéren belül is szükség lehet a műfajok, dokumentumtípusok és a művek más alapadatait leíró tipológiákra. Ezekre az igényekre mindre definiálni lehet – speciális célú és terjedelmű – tudásszerkezési rendszereket, amelyeket mind a nyelvi modell megfelelő helyére lehet beírni. A 'concept_concept' tábla az, ahol rögzíteni lehet az általános fogalmak közti általános kapcsolatokat.

Ha a tulajdonnévterek számára fontos mini-KOS-ok számára alkalmas modellt teremtünk, akkor ezt a modellt természetesen már használni lehet arra is, hogy bármilyen tudásszerkezési rendszer fel lehessen építeni benne. Mivel az általános fogalmak között szerepelnek mindazok a relációk, amelyek akármelyik tudásszerkezési rendszerhez szükséges lehet, az általános fogalmak közé pedig tetszőleges új fogalmat fel lehet venni, ezért a KOS-ok építésének nincs elvi akadálya.

A klasszikus tudásszerkezési rendszerek építésének egyetlen „problémáját” az jelentheti, hogy a tulajdonnévterek és a KOS-ok építésének logikája abból egy szempontból nem kompatibilis egymással. A teauruszok, a klasszifikációs rendszerek „megengedik” azt, hogy individuális és általános fogalmak egyaránt legyenek a rendszerben, míg a névtér aktuális modelljében ezek élesen elkülönülnek egymástól. Nem feloldhatatlan ez az „ellentmondás”, több megoldás is elképzelhető a szükséges transzformációra. Definiálni lehet olyan kapcsolatokat, amelyeken keresztül össze lehet fűzni a különböző táblákban tárolt, különböző minőségű adatokat egymással. Kijelenthetjük tehát, hogy tetszőleges tudásszerkezési rendszer felépíthető olyan módon, hogy részben a nyelvi modellben rögzített általános fogalmakból, illetve a névterekben tárolt individuális fogalmakból kapcsoljuk össze.

3.4.1.3.4 Névterek és a nyelvi modell viszonya

Minden tulajdonnévtérben legalább két alkalommal igénybe kell venni a példány relációt. Először minden tulajdonév esetében kötelezően meg kell adni, hogy milyen névtípusba (személynév, földrajzi név, testületi név stb.) tartozik, másodsor pedig azt is meg kell mondanunk, hogy a névhordozó mint individuum milyen entitástípus (személy, földrajzi hely, testület stb.) alá sorolható be. Mindkét esetben minimum egy-egy fogalmat, illetve – adott nyelven, például magyarul – minimum egy-egy nyelvi konstrukciót kell a névhez, illetve a névhordozóhoz rendelni. Ez a két hozzárendelés azt jelenti, hogy első lépésben felvesszünk egy-egy individuális entitást a névhordozók, illetve a megnevezések terébe.

Az nem szorul magyarázatra, hogy miért kell a névhordozók példányait önállóan azonosítani. A nevek, pontosabban a megnevezések példányosítása a névtér projekt újszerű megoldása, ami nem magától értetődő, ezért szükségesnek látszik itt némi indoklás. A nyelvi modell elkülönítésének legfőbb értelme az, hogy egyetlen sémán belül kezeljük minden nyelvi információt. Ide kerül minden nyelvi kommunikáció során felhasználható nyelvi konstrukció (köznevek, tulajdonnevek, igék, jelzők, határozószavak, toldalékok, központosítási jelek stb.). természetesen a tulajdonnevek is szerepelnek a nyelvi modellben. Ha a 'Horváth Béla' nyelvi konstrukciót használhatjuk személyek referálására (márpedig használhatjuk), akkor annak ott a helye a nyelvi modellben. Tudjuk azt is, hogy ezt a tulajdonnevet többször igénybe vehetjük személyekre való utalásként, viszont ezt a többszöri használatot nem lehet a nyelvi modellben kezelni, mert nem lehet, nem érdemes annyiszor felvenni oda az adott kifejezést, ahányszor azt használják a személynévtérben belül (vagy bármilyen más névhasználati kontextusban).

További problémát jelent, hogy a névterekben azért sem igazán használhatunk tulajdonneveket, mert azokhoz nem lehet időbeli információt lehet rendelni, hiszen a neveknek nincs (nem lehet) inherens időfüggése, időbeli lokalizációja. Márpedig sok esetet ismerünk, amikor a tulajdonnevek és a névhordozók között egyértelműen létezik valamilyen időbeli kapcsolat.

Ezt a látszólagos ellentmondást azzal tudjuk feloldani, hogy deklaráljuk, a névterekben nem neveket, hanem **megnevezéseket** rendelünk a névhordozókhoz. Ha így teszünk, akkor egyfelől nincs szükség a tulajdonnevek többszörös felvételére a nyelvi modellen belül, elég csak egyetlen névalakot felvenni oda és akárhányszor lehet – a megnevezéseken keresztül – rájuk hivatkozni. Másfelől a megnevezésekhez mint individuális eseményekhez már lehet időadatot rendelni, hiszen az eseményeket mindig el lehet helyezni (lokalizálni lehet) a idő dimenziójában. Ezt a modellezési döntést hosszabban kifejtettük egy korábbi kutatási jelentésben, a részletesebb kifejtést, az alaposabb indoklást ott lehet elolvasni.

3.4.1.4 Relevanciakezelés a névtérben

Az elméleti fejezetekben körbe jártuk a relevanciakezelés problémáját, elkülönítettük az információs tartományokon belül megállapítható távolsági mértékekre, illetve felhasználók által kinyilvánított értékjelzésekre támaszkodó relevanciaértelmezéseket egymástól. A következőkben azt fejtjük ki, hogy hol és hogyan használhatjuk ki az érték-, illetve mérték-alapú relevanciakezelés lehetőségeit a névtereken belüli keresések során.

3.4.1.4.1 Értékkezelés a névtérben

Az értékjelzés-alapú relevanciakezelés során azt használhatjuk ki, hogy a felhasználók értékjelzéseket „adnak le” az információs térben való „mozgásuk” során, és ezeket összegyűjtve, különböző technikákkal aggregálva visszaforgathatjuk a következő navigációs események támogatására.

A kérdés az, hogy az aktuális felhasználók névtereken belüli megnyilvánulásai alapján milyen értékjelzésekre számíthatunk, és ezek alapján hogyan tudjuk támogatni a későbbi látogatók navigációs tevékenységét. Az alábbi lehetőségeket tekintjük át röviden.

- gyakorisági ajánlat
 - legtöbbször keresett kifejezés névterenként, együttesen
 - legtöbbször megnézett oldal
 - legtöbbször megnézett térkép, grafikon
- szerkesztői ajánlat
- véletlen ajánlat

A következő alfejezetekben a fenti három ajánlatot elemezzük.

3.4.1.4.1.1 Gyakorisági ajánlat

A társas navigációról szóló elméleti fejezetben áttekintettük a felhasználói tevékenységéből kinyerhető értékjelzéseken alapuló relevanciakezelés lehetőségeit. Ez a relevanciakezelési mód valamilyen gyakoriságon alapuló, algoritmusok által kidolgozott ajánlatot jelent. Az ilyen technológiák mindig a forgalmi metaadatok alapján dolgozzák ki az ajánlataikat. Minél inkább értékelő jellegű felhasználói interakcióról van szó, az így begyűjtött felhasználói forgalmi adat annál inkább megfelel a társas navigációs jelzés céljaira. Bizonyos értelemben minden interakciónak tulajdoníthatunk egyfajta értékelő jelleget, hiszen a média egészére, így a webes kommunikációra is igaz, hogy a felhasználó a figyelmével fizet, és mivel ez a figyelem „korlátos, ritka jószág”, bármilyen dokumentum, információ, adat, amit a felhasználó a figyelmével „kitüntet”, „értékesebb” azokhoz az egységekhez képest, amelyekre nem jut figyelem. Természetesen az interakciók között lehet (és kell) különbséget tenni a bennük megnyilvánuló értékelköteleződés erőssége mentén. Az erősebben értékelő interakciók közé sorolhatjuk – többek között – a tetszésnyilvánítást (like-olást vagy vagy dislike-olást), a szavazást, a vásárlást.

A névterekben viszonylag kevés erősen értékelő jellegű interakcióra lehet számítani, hiszen itt legtöbbször a világtudásunk egy apró szeletét kínáljuk fel az érdeklődőknek, amivel kapcsolatban ritkán szoktunk értékelköteleződéseket kinyilvánítani. Bizonyos oldalakon fel lehet kínálni a tetszésnyilvánítás lehetőségét, leginkább ott, ahol nem elemi, hanem összetettebb adatokat prezentálunk (mondjuk a választási adatokat interaktív térképen mutató oldalakon).

A névtereken belül a leggyakoribb (és ezáltal leginkább mérhető) interakció vélhetőleg a látogatók keresőtevékenysége lesz. A felhasználók kereséseit nyilván számolni kell, hogy az ilyen forgalmi adatokból ki lehessen számolni a **legtöbbször keresett kifejezéseket** mutató indikátort. A gyakorisági adatokat több helyen és több módon is hasznosítani lehet. Egyfelől a **keresésgyakorisági** adatokat felhasználhatjuk a találati oldal rendezése során. Fel lehet kínálni olyan rendezési opciót, amely a keresésgyakorisági érték alapján rendezi el a találatokat. De lehet olyan modult kitenni a névtér valamelyik pontjára, ahol a leggyakoribb keresőkifejezéseket mutatjuk meg.

A keresésgyakorisági értékeket nyilván névterenként érdemes venni, de meg lehet próbálkozni azzal is, hogy valamilyen összetett, valamennyi névtérre vonatkozó gyakorisági értéket is számolunk, és a névterek „közös” oldalán ezt az értéket ugyanúgy hasznosítjuk, mint a konkrét névterek saját mutatóit. A keresésgyakorisági mutató számolásakor figyelembe kell venni az időt is, hogy az időbeli index alapján tudjuk kezelni a „keresési divat” jelenségét: mindig lehetnek olyan keresések, amelyek adott időszakban valamilyen dokumentum, adat, információ iránti nagyobb – gyakran csak időleges – érdeklődést jelzik.

A kereséshez képest erősebb értékelő interakciónak minősíthetjük azt, amikor a látogató meg is néz egy oldalt a találati listában felkínáltak közül. Erre az interakcióra is lehet mértéket képezni, és a **legtöbbször megnézett oldal** indikátorát ugyanolyan módon érdemes kezelni, visszaforgatni a névtér felületére, mint a keresésgyakoriság mutatószámát. Meg kell csinálni az indikátort névterenként és össznévtér szinten is, lehet a találati lista rendezésében hasznosítani, és lehet olyan felületet készíteni, ahol a nézésgyakorisági mutató alapján kínáljuk fel külön a „legnépszerűbb” tartalmakat a látogatók számára. A névterek alapoldalain a névtértípusok szerint kezelt névhordozók alapadatai és nevei érhetők el. Sok esetben ezek is fontosak a felhasználók számára, de biztosan lesznek olyan látogatók, akik nem – szöveges – alapadatokért, hanem „tartalmasabb” jellegű dokumentumokért, szövegekért, képekért, térképekért érdeklődnek inkább. Az ilyen igények kielégítésére érdemes a névtereken belül kezelt, „nagyobb figyelemre számot tartó” dokumentumokat, képeket, térképeket külön kezelni, és saját mértéket számítani az esetükben. A **legtöbbször megnézett dokumentum** (életrajz, térkép, grafikon, portré stb.) mutatóját ugyanúgy többféleképpen lehet kezelni, mint a többi forgalmi indikátort.

Az eddig bemutatott indikátorok mintájára lehet olyan mutatókat is definiálni, amelyeket nem a gyakoriság, hanem a „frissesség” alapján számolunk ki. Ilyenkor a leggyakoribb helyett a legfrissebb (legutolsó) keresőkifejezést, látogatott oldalt, találatot mutathatjuk meg. A **legutolsó keresés, legutóbb megnézett dokumentum** indikátorait ugyanúgy lehet kezelni, mint a többi mutatót.

3.4.1.4.1.2 Szerkesztői ajánlat

A predigitális korban a legáltalánosabb szűrési szempontok egyikét azok a szerkesztői, kurátori, mecénási, kiadói döntésekben megnyilvánuló értékjelzések jelentették, amelyek a dokumentumok (könyvek, cikkek, zeneművek, filmek, képek stb.) nyilvánosságban való megjelenését lehetővé tették. A különböző dokumentumtípusok esetén a nyilvánosság elé, majd a hagyományos archívumokba azok a dokumentumok juthattak el, amelyek a különböző kulturális kánonképzési csatornákon keresztül el tudták nyerni a megfelelő döntéshozók támogatását. A kánonképző mechanizmusok működtetése során hozott döntéseket nem kevésszer partikuláris elfogultságok terhelték, amit az utókor gyakran és joggal kritizálhatott, de ettől még ez a rendszer hosszú időn keresztül működőképes kiválasztási mechanizmusként funkcionált. A digitális kor – több szempontból – radikálisan megváltoztatta a hagyományos kánonképzés lehetőségeit. A digitális eszközök mindenki számára megkönnyítették új dokumentumok létrehozását, a hálózati összekapcsoltságnak köszönhetően pedig sokkal könnyebb lett a dokumentumok terjesztése a közönség felé. Ez az átalakulás sok – bár nem minden – szempontból valóban „főlösgessé” tette a régi kánonképző mechanizmusok igénybevételét. Mindezek ellenére leszögezhetjük, hogy a hagyományos kánonképző technikákat (de legalább is azok egy részét) érdemes lenne fenntartani a digitális korszakban is, hiszen láttuk, hogy az információs tartományokon belüli navigációhoz milyen fontosak a relevanciaképző metódusok, és megítélésünk szerint a régi, hagyományos kanonizálási eljárások között vannak olyan is, amelyeket érdemes „átemelni” a digitális korszakba is.

A névtereken belül biztosítani lehet megfelelő véleményvezéreknek, hogy ők válogathassanak a „gondjukra bízott” információs tartomány elemi közül, és azokat adott felületre kitehessék. Az ilyen oldalak relevanciáját a véleményvezérek által képviselt értékészlet adná, és ez nyilván a látogatók azon része számára lenne releváns, amelynek tagjai az adott értékközösségbe tartoznak.

3.4.1.4.1.3 Véletlen ajánlat

A relevanciakezelés kívánalma maga arra a megfontolásra épül, hogy az információs tartományt, a dokumentumokat el lehet (és ha lehet, akkor el is kell) rendezni egy vagy több fontossági szempont szerint. Ez talán plauzibilis feltételezés, mégsem érdemes arra törekedni, hogy kizárólagosan kövessük ezt a logikát. Előfordulhatnak olyan esetek az életünkben, amikor egy véletlenszerűen elének kerülő „ajánlat” mindennél érdekesebb, relevánsabb lehet számunkra, és ezt a lehetőséget érdemes a névtereken belül is kihasználni.

A hagyományos könyvtári világban ezt teszi lehetővé a szabadpolcra kihelyezett könyvek közti böngészés, de azt nyújtja nekünk az a polc is, ahová a legutóbb megjelent könyveket vagy az olvasók által éppen most visszahozott könyveket pakolják. Ezeket a véletlen által egymás mellé rendezett tételeknek kell minősítenünk, és ennek mintájára érdemes felkínálnunk a névtereken belül is azt a lehetőséget, hogy a látogatók számára elérhető legyen egy olyan felület, ahol a véletlen által „generált” belépési pontokat kínálunk fel. A véletlenszerűen kiválasztott elemeket tartalmazó oldalaknak, felületeknek ugyanazt a funkcionalitást kell kínálniuk, amit a hagyományos keresés során visszaadott találati oldalak nyújtanak.

Természetesen felmerül a kérdés, hogy milyen adatokat érdemes ilyen célokra használni az egyes névtereken belül, hiszen nyilvánvalóan nem mindegyik adattípus lehet alkalmas erre a „feladatra”.

- a személynévtérből fel lehet kínálni egy véletlenszerűen kiválasztott személy legfontosabb életút-adatait, illetve rövid életrajzát;
- a személynévtérből ki lehet tenni a véletlen alapján egy személy portréképét vagy – ha van – egy videóját;
- a testületi névtérből fel lehet kínálni egy véletlenszerűen kiválasztott testület legfontosabb alapadatait, illetve rövid történetét;
- a testületi névtérből ki lehet tenni egy, a testületet jellemző képet, mondjuk egy logót;
- a földrajzi névtérben a „véletlen oldalon” megjelenhetnek bármelyik földrajzi hely alapadatai, esetleg térképen mutatva a hely elhelyezkedését a földrajzi térben;
- a földrajzi névtérből véletlenszerűen lehet választani azon képek, címerek, zászlók közül, amelyek településekhez, megyékhez, országokhoz stb. kapcsolhatók;

- a könyvcímterekből értelemszerűen a könyvek borítóit, esetleg fülcímzövegeit és persze a könyvek alapadatait lehet a véletlen alapján „előrehozni”;
- a filmterekből a filmekhez kapcsolódó alapadatok, standfotók, plakátok, filmrészletek vagy akár a teljes filmek is kitehetők a „véletlen oldalra”;
- a névterek legalsó szintje fölé építhető további adatkörök, mint például a választási adatok vagy a társadalomstatistikai adatokból véletlenszerűen kitett elemi egységek (táblázatok, térképek) nagyon sokat segíthetnek az ilyen adatbázisok tartalmának feltárásában, az érdeklődés felkeltésében.

A véletlenszerű ajánlatok a szó szoros értelmében nem mondhatók relevánsnak, hiszen sem felhasználói értékelköteleződést, sem valamilyen közelségi mértéket nem találhatunk mögöttük, mégis érdemes próbálkozni az ilyen technológiákkal, hogy a teljes információs tartomány bármelyik szegmense, bármelyik eleme megragadhassa – legalább véletlenszerűen – a felhasználók figyelmét.

3.4.1.4.2 Mértékkezelés a névtérben

A keresőtechnológiák világában valamilyen közelségi mértéket számolva, és azt az információs tartomány elemi egységeihez rendelve próbálják biztosítani azt, hogy a találatok relevánsabbak tűnjenek a felhasználók számára. Ebben a fejezetben összeszedjük azt, hogy a névtereken belül milyen közelségi mutatókat számolhatunk, és hol tudjuk ezekkel növelni a keresés hatékonyságát.

3.4.1.4.2.1 Szintaktikai közelség kezelése

Amikor nyelvi alapú keresésekről van szó, a keresés hatékonyságát többféle technikával lehet növelni még akkor is, ha a nyelvi kifejezések szintaktikai tulajdonságait vesszük figyelembe a keresés során.

Az egyik szempont szerint fel lehet ajánlani a teljes vagy részleges egyezésen alapuló keresés lehetőségét. Nyilvánvaló, hogy a teljes egyezésen alapuló keresés nagyobb relevanciával rendelkezhet a részleges kereséshez képest, de ez nem lehet oka arra, hogy csak ezt az opciót preferáljuk. A teljes egyezésen alapuló keresés sok esetben kevés találatot ad, esetleg nem ad vissza semmit, és ilyenkor jól jöhet a kevésbé precíz, ám nagyobb felidézési arányt produkáló keresési lehetőség.

Részben a teljes vs. részleges egyezésen alapuló keresés területére tartozik, de érdemes külön kezelni a betűérzékeny vagy betűérzéketlen keresés lehetőségét. A kis- és nagybetűk megkülönböztetését azért érdemes – részben – külön opcióként kezelni, mert itt nem a karaktersorozat elemeiben, hanem azok formai jellemzőiben lehet eltérés, és a kétféle keresési opció a betűérzékenységre figyelembe vételén múlik.

A keresés relevanciáját növelhetjük olyan technikákkal is, amelyek a keresőkifejezéseket „értelmezik”, formálják át, és a keresést többször is végrehajtják. Ezek a technikák bizonyos értelemben valamifajta névváltozat-kezelésnek tekinthetők. Ide tartozik a kötőjeles vagy kötőjel-nélküli alakok, az egybeírt vagy különírt alakok, a „hivatalos elírások” vagy a „felhasználói elírások”. Nézzünk meg ezekre pár példát!

Különösen a földrajzi és testületi nevek világában gyakran előfordul, hogy a nevekben kötőjelek, betűközök vannak, és az ilyen neveket a névhasználói gyakorlatokban többféleképpen írják. Elvileg mindig van egy hivatalos, szabályos alak, de a tényleges gyakorlatban ettől még használatban tartják a „szabálytalan” alakokat is.

Ürmös-hát vagy Ürmöshát (kötőjel-használat)

Napospart vagy Napos part (különírás-egybeírás)

Mivel nem tudhatjuk, hogy a felhasználók vajon a szabályos vagy a szabálytalan alak szerint keresnek-e, az ilyen esetekben az a bölcsőbb, ha a keresőrendszer mindkét alak szerint hajtja végre a keresést, és úgy ad vissza találatot. Ez egyfajta relevanciakezelésként minősíthető, hiszen a felhasználók számára találatot ad vissza attól függetlenül, hogy ők milyen „pontosan” tették fel a keresőkérdéseiket.

Olyan esetek is lehetségesek, amikor a névtérben a különböző – akár hivatalos – forrásokból származó nevek alakjai eltérnek egymástól. Ezeket a „hivatalos hibákat” is kezelni kell, mert a felhasználók megint csak bármelyik változat alapján is kereshetnek.

Bánfi-szivattyútelep vagy Bánkfi-szivattyútelep („hivatalos elírás”)

Arra is fel kell készíteni a keresőrendszert, hogy „felhasználó-oldali hibát” (de legalább is azok egy részét) is kezelni tudja

Dunaújváros helyett Dunaujváros („felhasználói elírás”)

Dunaújváros helyett Dunaújvros („felhasználói elírás”)

A felhasználó-oldali hiba lehet véltelen elgépelés eredménye és lehet helyesírási hiba is (a fenti két példában az első helyesírási, a második gépelési hiba). Mivel a magyar nyelven sok ékezetes magánhangzót használunk, és ezek a tulajdonnevekben is gyakran előfordulnak, sok esetben számolhatunk azzal, hogy a keresőkifejezések – a helyesírásukat tekintve – rossz alakban érkeznek be. A keresőmotoroknak kezelnie kell tudni ezt a helyzetet, és minden lehetséges „ékezetes hiba” esetén fel kell ismernie, hogy a látogató vajon milyen kifejezésre akarhatott keresni akkor is, ha tévedett a beírás során. Ezt a magánhangzók (és bizonyos mássalhangzók, mint a 'j' és 'ly') esetében meg lehet csinálni. Más típusú értelmezési feladat adódik a keresőmotor számára akkor, amikor a felhasználó nem helyesírási, hanem gépelési hibát vét ('Dunaújváros' helyett 'Dunaújvros'-t ír). Ennek a felhasználó-oldali hibának a kezelését úgy lehet megoldani, hogy valamilyen szóhasonlósági mértéket definiálunk, és ha két szóalak (a „helyes” és a „helytelen”) adott közelségben van egymáshoz, akkor a „rosszul megadott” keresőkifejezést is figyelembe vesszük a keresés során úgy, hogy a „helyes alak” szerint keressünk (és adunk vissza találatokat).

A szavak, kifejezések szintaktikai szerkezetével kapcsolatban további „hibázási” lehetőséget jelent az, hogy egy összetett név esetében a felhasználó esetleg rossz sorrendben adja meg a név komponenseit. Ezt is korrigálni kell tudnunk, amit úgy tehetünk meg, hogy a keresés során nem ragaszkodunk a komponensek helyes sorrendjéhez, a felhasználó által megadott összetett keresés egyes elemeit a sorrendiség nélkül vesszük figyelembe. Ez természetesen nem jelenti azt, hogy a felhasználó által megadott sorrendet ne lehessen beszámítani. Ha olyan találatok vannak, amelyek ugyanazokat a névelemeket tartalmazzák, csak más sorrendben, akkor nyilván értelmes lehet aszerint rendezni az ilyen találatokat, hogy előbbre kerüljenek azok az elemek, amelyek jobban igazodnak a felhasználói keresőmintához.

A nyelvi konstrukciók világában a szintaktikai közelséget értelmezhetjük úgy is, hogy a találati elemek alfabetikus távolságát vesszük alapul. Ez a relevanciakezelés nem az információ tartományon belüli keresésre, szűrésre, hanem az ezek eredményeként visszaadott találati halmaz elrendezésének módjára vonatkozik. Előfordulhatnak esetek, amikor érdemes alkalmazni ezt a technikát is a találati oldalakon. Amikor a találati oldalon ki kell tenni a találati halmaz elemeit, akkor mindig fontos kérdés, hogy milyen sorrendben jelenjenek meg a találatok. Ha rendelkezésre áll megfelelő relevanciakezelési lehetőség, akkor maga az algoritmus biztosítja az elemek megjelenítési sorrendjét. Vannak azonban olyan esetek, amikor nem áll rendelkezésre elég információ a sorrend kialakítására. Ilyenkor érdemes lehet a találati oldalra irányított egységekhez kapcsolt adatok alapján többféle rendezési lehetőséget felkínálni. Az egyik opció nyilván az ábécé szerinti rendezés. Sok esetben lehet kényelmes a felhasználók számára, ha a találati halmaz elemeit ebben a sorrendben tekinthetik át. Előfordulhat azonban az is, hogy az alfabetikus rendezés nem segít. Például akkor, amikor ebben a dimenzióban teljesen (vagy nagyon) egyező elemek szerepelnek a találati listában (mondjuk a földrajzi nevek esetében százas nagyságrendben szerepel az 'Öreghegy' név).

A találati oldalakra kitett egységek adattartalmától függően azonban mindig adódhatnak másfajta rendezési lehetőségek az ábécé szerinti rendezés mellett. Például a személyek esetében a sok, azonos 'Szabó István' név esetén nem releváns az ábécé szerinti sorba állítás, viszont érdekes lehet a születési időpont szerint elrendezni őket, míg a földrajzi helyek esetében el lehet rendezni az azonos nevű elemeket, mondjuk, a partitív fölérendeltjeik neve alapján, esetleg a típusba sorolásuk szerint.

3.4.1.4.2.2 Szemantikai közelség kezelése

A szűk értelemben vett névterek szemantikailag gyengék, laposak annak köszönhetően, hogy alapvetően tulajdonnevekkel jelölt individuális entitások alapadatait tartalmazzák. Van azonban néhány olyan információ az alapadatok között, amely alkalmas lehet az individuális entitásokhoz kapcsolható szemantikai téren belüli közelségi viszonyok kezelésére (és ezzel szemantika-alapú relevanciakezelésre). Ezeket nyilván érdemes kihasználni a keresőoldalakon.

A személynévtérben a személyekhez rendelt foglalkozások felkínálják annak lehetőségét, hogy az általuk kifejezett szemantikai térben lehessen valamilyen szemantikus közelséget kezelni. Vegyük azt a példát, amikor a felhasználó az 'agrokémikus' foglalkozásra keres (értsd: azt szeretné tudni, hogy a személynévtérben szereplő személyek közül 'kik az agrokémikusok'). Ekkor nyilván vissza kell adnunk (és vissza is tudjuk adni) azokat a személyeket, akikhez pontosan ez a foglalkozás hozzá van rendelve. Ha viszont tudjuk, hogy szemantikailag két olyan foglalkozásmegnevezésünk van, amely alapján találatokat kellene visszaadnunk, de szintaktikailag a második alak nem illeszkedik a keresőkifejezéshez, akkor valamit lépünk kell.

agrokémiai kutató

agrokémikus

Ha a felhasználó számára csak az 'agrokémikus' kifejezéshez szintaktikailag illeszkedő (szintaktikailag helyes) találatokat adjuk vissza, akkor szemantikailag gyengébb eredményt produkálunk ahhoz képest, amikor visszaadjuk az 'agrokémiai kutató' keresőkifejezéshez illeszkedő találatokat is. A világtudásunk alapján ugyanis tudjuk, hogy aki 'agrokémikus'-ra keres, annak megfelelő találat lesz az is, ami az 'agrokémiai kutató' mintára illeszkedik. Ezt a szemantikai közelséget ki lehet fejteni tudásszervezési rendszerek segítségével, és ha így teszünk, akkor a KOS-ban tárolt tudást hasznosítani lehet a keresések hatékonyságának szemantikai alapon történő javítására.

Ha a foglalkozásokat kifejező nyelvi konstrukciókból felépítünk egy foglalkozási struktúrát, amelyben kifejezzük a köztük levő szinonimitási vagy generikus alá-, fölérendelési viszonyokat, akkor ebből a szerkezetből mindig kiolvashatjuk azokat a konstrukciókat (fogalmakat), amelyekre támaszkodva szemantikailag relevánsabbá tehetjük a találati eredményeket. A fenti példára hivatkozva: ahhoz, hogy a felhasználónak visszaadjuk a személyeket akkor is, ha a foglalkozás alapján visszaadott találati lista elemei szintaktikailag és szemantikailag nem pontosan illeszkednek a felhasználói keresőkifejezésre, az szükséges, hogy eltávolítsuk valahol (egy KOS-ban) a következő állítások valamelyikét.

agrokémiai kutató IS-A-TYPE-OF agrokémikus

agrokémiai kutató IS-A-SYNONYM_OF agrokémikus

A két állítás némileg eltérő ontológiai elköteleződést fejez ki, igazából az adott információs rendszer céljaitól függhet az, hogy mikor melyiket választjuk. Abban viszont nincs különbség a két állítás (és elköteleződés) között, hogy mindkét esetben ugyanúgy lehet hasznosítani az állításokban rejlő tudást a keresés szemantikai pontosítására. Azt kell csak tenni, hogy a két állítás alapján nem csak az 'agrokémikus' kifejezésre, hanem – a KOS-ból kiolvasott kapcsolat alapján – a 'agrokémiai kutató' konstrukcióra is keresünk a foglalkozások és személyek közti kapcsolatot kifejező információs tartományban.

További példaként hivatkozhatunk még azokra az esetekre, amelyeket a KOS-ban alapú keresésről írott fejezetben mutattunk be.

A személynévtérben a foglalkozásokhoz hasonló szemantikai pontosításra ad lehetőséget a személyek nemzetiségi, állampolgársági „besorolása” alapján történő keresés. Ekkor azt a KOS-ban tárolt tudást mozgósíthatjuk, amely a nemzetiségi, állampolgárságot reprezentáló kifejezéseket (fogalmakat) kapcsolják össze, például:

talján IS-SYNONYM-OF olasz

A földrajzi névtéren belül értelem szerűen másfajta információ alapján végezhetünk szemantikai közelségen alapuló kereséspontosítást. A legnyilvánvalóbb példa lehet erre a települések típusai között felállítható szemantikai kapcsolatok hasznosítása.

A földrajzi névtér számára az a kívánatos megoldás, hogy a települések típusba sorolását a lehető legpontosabb kifejezésekkel adjuk meg. Ha ezt tesszük, akkor a – magyar – települések jellemzéséhez az alábbi listából választhatunk típusleíró információt:

község

nagyközség

város

megyei jogú város

főváros

A listából egyet és csak egyet kell kiválasztanunk, hogy a lehető legpontosabb leírását adjuk meg a településnek ebben a dimenzióban. Ha viszont így teszünk, akkor bizonyos keresések esetén könnyen szemantikailag gyenge találati eredményekre számíthatunk. A 'megyei jogú város'-ra vagy a 'fővárosra' keresve (teljes illeszkedést feltételező keresések esetében) nem kapjuk vissza sem a megyei jogú városokat, sem a fővárost akkor, amikor valaki a 'város'-ra keres, miközben tudjuk jól, hogy mind a megyei jogú városok, mind a főváros a város fogalom alá tartoznak, tehát szemantikailag indokolt lenne ezeknek a városoknak is szerepelniük a találati listában. A megoldás nyilvánvaló: a KOS-ban eltárolt tudást kell hasznosítanunk a keresés feltételeinek bővítésére és ezzel a szemantikai pontosság növelésére. Ehhez fel kell venni a tudásszervezési rendszerünkbe a következőket:

nagyközség IS-A-TYPE-OF község

megyei jogú város IS-A-TYPE-OF város

főváros IS-A-TYPE-OF város

Ezek az állítások alkalmasok arra, hogy a fent említett pontosításokat elvégezhessük, és relevánsabb találati eredményeket adjunk vissza.

* * *

A névtérrel alapadatrendszerként, egyfajta alapzatként értelmezve adódik a lehetőség, hogy további adatkörök is beépítsünk a névtérrendszer egészébe (választási adatok, archontológiai adatok, társadalomstatisztikai adatok stb.). Ezek az adatkörök már komolyabb szemantikával rendelkeznek, többfajta általános adattípus különíthető el bennük. Ilyen esetekben további lehetőségek adódnak a különböző adattípusok közti szemantikai távolságok kezelésére, és a névtérrendszer általános keresőjéhez többféle kiegészítést, szemantikai bővítést lehet illeszteni. Ezek megtervezése, leírása azonban egy következő kutatási jelentés tárgya lesz.

3.4.1.4.2.3 Térbeli közelség kezelése

A térérzékeny keresés technikái azt használják ki, hogy a földrajzi térben lokalizálható entitások közelségét ki lehet használni arra, hogy a megadott térbeli közelségi feltételeknek megfelelő találatokat adjanak vissza. A térbeli közelség kezelése nyilvánvalóan alkalmas arra, hogy a találatokhoz relevanciamértéket rendeljünk, és ez alapján javítsuk a keresések minőségét, a felhasználók elégedettségét.

A térérzékeny keresés lehetőségeiről egy másik kutatási jelentésben adunk számot.

3.4.1.4.2.4 Időbeli közelség kezelése

A térérzékeny keresés technikái azt használják ki, hogy a névterekben leírt entításokhoz rendelhető – és az időtengelyen elhelyezhető – események közt létezik és elemezhető az időbeli közelséget kifejező kapcsolat, amit természetesen fel lehet használni szűrésekre a keresések során.

Az időérzékeny keresés lehetőségeiről egy másik kutatási jelentésben adunk példákat.

MUNKANYAG

4 Tér- és időérzékeny keresés

4.1 Térérzékeny keresés

A nemzeti névtér teljes rendszere egyfelől és alapvetően egymástól jól elkülöníthető tulajdonnévterekből (mint személynévtér, testületi névtér, földrajzi névtér stb.) áll, de a nemzeti névtér része lehet egy vagy több köznévtér is. A köznévtér fogalma nem határozható meg olyan pontosan mint a tulajdonnévtér fogalma, a köznévterekben rögzített világtudás részben a felhasználási céloktól, részben a névtereket megtervező és felépítő emberek ontológiai elköteleződéseitől függ az, hogy milyen ismeretterületre vonatkozóan, milyen csúcskategóriákat elfogadva, milyen szerkezetben tárolják a köznévterekbe írt tudást.

A tulajdonnévterekbe individuális entitások adatait vesszük fel, és ennek a ténynek fontos következményei vannak. A konkrét individuumokra ugyanis az jellemző, hogy mint a létező dolgok előfordulásait, mindig lokalizálni tudjuk a tér és idő dimenziója mentén. A névterekbe felvett individuumok esetében ez a minőség a gyakorlatban úgy realizálható, hogy az individuális entitások alapjellemzőit mindig elhelyezhetjük a tér és/vagy az idő tengelyei mentén. A tulajdonnévterek ezen minősége “felkínálja” azt, hogy a névterekbe tartozó individuumok keresése során kihasználjuk azt a lehetőséget, hogy a entitáselőfordulások térbeli és/vagy időbeli elhelyezkedését is figyelembe vesszük a keresések során. Az ebből adódó következő kérdés az, hogy milyen módon lehet támogatni a térérzékeny, illetve időérzékeny keresés tevékenységét. A következő fejezetekben ezt a kérdést vizsgáljuk meg alaposabban. Vizsgálódásunk elsősorban a térérzékeny keresés lehetőségeinek vizsgálatára irányul, az időérzékeny keresés lehetőségeit csak röviden elemezzük.

4.1.1.1 Partitív reláció

A térérzékeny keresés a partitív reláció mentén értelmezhető. Ezt a relációtípust régóta elemzik, a mereológia elmélete a partitív reláció szabatos definiálásával, a reláció egyes altípusainak pontos elkülönítésével, a partitív reláció alkalmazása során keletkező problémák elemzésével, megoldásával foglalkozik.

4.1.1.1.1 A partitív relációt leíró elmélet: mereológia

A rész/egész relációval vagy másként a partitív alárendeltje/partitív fölérendeltje relációval a mereológia elmélete foglalkozik. A része (partitív alárendeltje) reláció az egésze (partitív fölérendeltje) inverze, és ez a relációkettős is tranzitív, hierarchikus – akár csak a generikus reláció. A tranzitivitás azonban csak a pontosan definiált – és egymástól is jól elkülönített – altípusokra érvényes, nem általában a partitív relációra. Sok esetben érvényesül a tranzitivitás. Vegyünk két példát!

Ha Érd Pest megye része, és Pest megye Magyarország része, akkor Érd is Magyarország része.

Ha az ujjam a kezem része, és a kezem a karom része, akkor az ujjam a karom része is.

Könnyen lehet azonban olyan példát találni, amikor a partitív reláció intranzitivitást tapasztalhatunk.

A zongorista keze része a zongoristának, a zongorista része a zenekarnak, de a zongorista keze nem része a zenekarnak.

Az ellentmondásosnak tűnő helyzet magyarázata egyszerű: többféle partitív reláció létezik, és ezek között nem érvényes a tranzitivitás. Olyan mereológiai elméletre van ezért szükség, amely kezelni, magyarázni tudja ezt a jelenséget. A versengő mereológiai elmélet közül itt a Peter Simons és Archello Varzi nevéhez köthető elméletet fogadjuk el (Simons 2000, Varzi 1996).

A partitív relációra igaz az alábbi három tulajdonság.

mereológiai axiómák

| | |
|------|---|
| (a1) | $\forall x(P(x, x))$ |
| (a2) | $\forall x\forall y(P(x, y) \wedge P(y, x) \rightarrow x = y)$ |
| (a3) | $\forall x\forall y\forall z(P(x, y) \wedge P(y, z) \rightarrow P(x, z))$ |

Az első három axióma rögzíti, hogy a P partitív reláció reflexív (a1), antiszimmetrikus (a2) és tranzitív (a3). A következő lépésben a P relációval néhány új mereológiai relációt definiálunk, majd az újakkal még újabbakat hozunk létre:

| | | |
|--------------|---|---------------------------|
| (1) PP(x, y) | $\forall x\forall y(P(x, y) \wedge \neg P(y, x))$ | valódi része |
| (2) O(x, y) | $\forall x\forall y\exists z(P(z, x) \wedge P(z, y))$ | átfedése |
| (3) U(x, y) | $\forall x\forall y\exists z(P(x, z) \wedge P(y, z))$ | kívüllevősége |
| (4) OX(x, y) | $\forall x\forall y(O(x, y) \wedge \neg P(x, y))$ | részhatáros átfedése |
| (5) UX(x, y) | $\forall x\forall y(U(x, y) \wedge \neg P(x, y))$ | részhatáros kívüllevősége |
| (6) PO(x, y) | $\forall x\forall y(OX(x, y) \wedge OX(y, x))$ | valódi lefedése |
| (7) PU(x, y) | $\forall x\forall y(UX(x, y) \wedge UX(y, x))$ | valódi kívüllevősége |

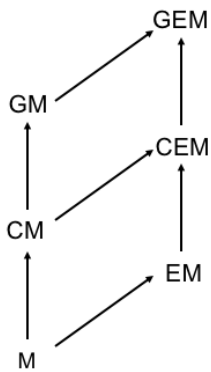
Sorban definiáljuk a következő relációkat: valódi része (1), átfedése (2), kívüllevősége (3), részhatáros átfedése (4), részhatáros kívüllevősége (5), valódi lefedése (6), valódi kívüllevősége (7). Ezek után a P része, O átfedése és U kívüllevősége relációk segítségével először meghatározhatjuk a kiterjeszthetőség (erős) elvét (a4), majd az (a5)-(a7) mereológiai axiómák elfogadásával definiálhatjuk a zárt (extenzionális) mereológia elméletét.

| | |
|------|--|
| (a4) | $\forall x\forall y(\neg P(x, y) \exists z(P(z, x) \wedge \neg O(z, y)))$ |
| (a5) | $\forall x\forall y(U(x, y) \rightarrow \exists z\forall w(O(w, z) \leftrightarrow (O(w, x) \vee O(w, y))))$ |
| (a6) | $\forall x\forall y(O(x, y) \rightarrow \exists z\forall w(P(w, z) \leftrightarrow P(w, x) \wedge P(w, y)))$ |
| (a7) | $\forall x\forall y\exists z((P(z, x) \wedge \neg O(z, y)) \rightarrow \exists z\forall w(P(w, z) \leftrightarrow (P(w, x) \wedge \neg O(w, y))))$ |

További axiómák felvételével egyrészt definiálhatjuk az általános (extenzionális) mereológia elméletét (a8) másrészt meghatározhatjuk az atomos (a9), illetve atom-nélküli mereológiákat (a10).

| | |
|-------|---|
| (a8) | $\exists x\phi \rightarrow \exists z\forall y(O(y, z) \leftrightarrow (\exists x(\phi \wedge O(y, x)))$ |
| (a9) | $\forall x\exists y(P(y, x) \wedge \neg \exists zPP(z, y))$ |
| (a10) | $\forall x\exists y(PP(y, x))$ |

Nem szükséges a fenti axiómák mindegyike ahhoz, hogy egy mereológiai elméletet definiálni lehessen. Varzi többfajta mereológiai elméletet különített el más és más axiómák együttes elfogadásával.



| | |
|-----|---|
| M | mereológia: (a1)-(a3) |
| EM | extenzionális mereológia: (a1)-(a4) |
| CM | zárt mereológia: (a1)-(a3), (a5)-(a7) |
| CEM | zárt extenzionális mereológia: (a1)-(a7) |
| GM | általános mereológia: (a1)-(a3), (a8) |
| GEM | általános extenzionális mereológia: (a1)-(a4), (a8) |

Ezek a mereológiai elméletek különböző erősségű és tartalmú ontológiai elkötelezettségek mentén érvényesek. A fenti hat elmélet egyikében sem szerepel azonban az az utolsó két mereológiai axióma, amelyek alapján elkülöníthetjük az atomos és nem-atomos partitív relációkat egymástól. Az (a9) és (a10) jelű tételek alkalmazásával a fenti elméletek tovább bonthatók atomos vagy atom-nélküli elméletekre.

A partitív reláció is tranzitív. Jeleztük azonban, hogy a tranzitivitás nem általában a partitív reláció egészére, hanem annak valamelyik típusán belüli kapcsolatokra igaz. A fenti formulák alapján most már mondhatjuk azt, hogy a partitív reláció tranzitivitása mindig csak valamelyik – fent definiált – mereológiai elméleten belül érvényesül. A partitív reláció tranzitivitása akkor válik érvénytelenné, amikor egy atomos és atom-nélküli partitív relációra akarjuk érvényesíteni. Az emberi test és a testrészek között része viszonyt atom-nélküli partitív relációval írhatjuk le, ezzel szemben a testületek és személyek közti szervezeti tagságot atomos partitív reláció alkalmazásával tudjuk megragadni. A kétféle rész-egész relációt azonban nem szabad összekeverni, különben ahhoz a tranzitivitási deficithez jutunk, amit korábban már bemutattunk a zenekari tag zongorista keze kapcsán. Természetesen lehet többféle partitív relációt használni egy rendszeren belül, de ilyen esetekben a tranzitivitás létezését csak adott partitív reláción belül várhatjuk el, a – részben – eltérő minőségű partitív relációk közt nem számíthatunk rá. Mindebből az következik a névtérépítésre vonatkozóan, hogy – szükség esetén – többféle partitív relációt kell definiálni és használni, hogy az ilyen keveredést elkerüljük.

A partitív reláció talán a leggyakrabban használt reláció, és ezért könnyen előfordul, hogy különféle használati kontextusokban némileg másként, vagyis más jelentések mentén használják ugyanazt a terminust. Ebből mindig adódhatnak problémák, ezért érdemes tisztázni azt, hogy a névtérrendszeren belül milyen partitív relációkat különítünk el egymástól. Első lépésként a legfontosabb alapkategóriákat kell definiálnunk, hogy utána pontosan el tudjuk különíteni az eltérő kategóriák közt alkalmazott partitív relációkat egymástól, és meg tudjuk mondani azt, hogy miben érhető tetten a köztük levő különbségek.

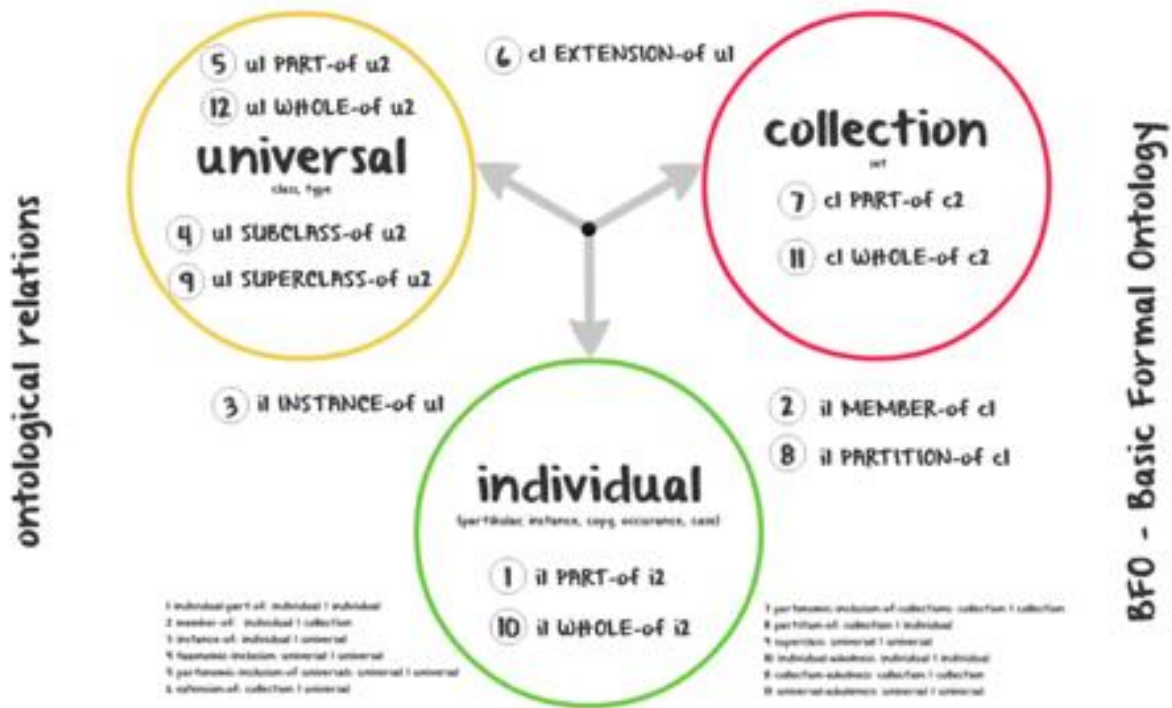
4.1.1.1.2 Ontológiai alaprelációk

A névtérrendszer számára elfogadtuk és át is emeltük a BFO (Basic Formal Ontology) projekt ajánlásait a legfontosabb kategóriákra és a köztük levő relációkra vonatkozóan (Bittner et al. 2004). Az BFO “útmutatása” szerint háromfajta fogalomtípust kell elkülönítenünk egymástól, úgymint:

- **individuum**, példány (individual)
- **típus**, univerzálé, osztály (universal)
- **gyűjtemény** (collection)

A háromféle fogalomtípus között különféle relációk definiálhatók, amelyek olykor ugyanazon relációtípusba tartoznak, de a relátumok minőségében különböznek egymástól, ezért el kell különíteni őket egymástól.

Az alábbi ábra mutatja azt, hogy a három alapvető kategória között milyen relációkat kell definiálnunk (az ábrán az angol nyelvű terminusokat tüntettük fel, hogy igazodjunk az alkalmazott az elmélet terminushasználatához).



A BFO fogalomértelmezéseit egy másik kutatási jelentésben mutattuk be, ahol részletesen elemeztük, mit, hogyan lehet definiálni, értelmezni a csúcskategóriák esetében. Itt csak a partitív relációval kapcsolatos nézeteinket fejtjük ki bővebben.

Mindhárom kategórián belül értelmezhetünk egy-egy partitív relációpárt, a partitív alárendeltje (része) és a partitív fölérendeltje (egésze) relációt. A relációpárok két-két tagja közt mindig fennáll az inverzitás, amit a következőképpen fejezhetünk ki.

$$\text{partitív alárendeltje } (x,y) \equiv \text{partitív fölérendeltje } (y,x)$$

A típusstartományban értelmezett partitív reláció két típusfogalom között teremt kapcsolatot. Példaként vehetjük az alábbiakat:

a fark része a kutyának

a ló része a ménesnek

a szarvasmarha része a gulyának

a megye része az országnak

Ezek az állítások azt írják le, hogy minden kutyára igaz, hogy ha van farka a kutyának, akkor az a kutya része, minden ménes lovakból, minden gulya szarvasmarhákból áll, minden megye az országhatárokon belül van, így minden, ami a kutyafarokhoz tartozik, az a kutyához is tartozik, minden, ami a lóhoz tartozik, az a méneshez is tartozik, minden, ami szarvasmarhához tartozik, az a gulyához is tartozik, és minden, ami a megyében van, az az országban is van. Fontos itt, hogy a partitivitás csak a megadott fogalmak között áll fent. Elképzelhető, hogy egy legelésző ménes lovai közé bebállag egy tehen, és ekkor mondhatjuk azt, hogy egy csoportban vannak (egy gyűjteményt alkotnak), de ettől még a tehenet nem tartjuk a ménes részének.

A korábbi kutatási jelentésben kifejtett relációelméletben elkülönítettük az általános partitív reláción belül két altípust, ami a most bevezetendő – más értelemben vett – altípuspárok esetében is értelmes altípusálásra ad lehetőséget. A partitív relációt leíró elméletben elkülönítik egymástól az **atomos** és a **nem-atomos** partitív relációkat, amelyeket nem szabad egymással felcserélhető módon használni. Előbbire példa a ló és a ménes, a tehen és a gulya vagy a csellós és a vonósnégyes közti partitív reláció, míg utóbbira a fark és a kutya, a pata és a ló vagy a zenész keze és a zenész közti partitív kapcsolat. A

két altípus közti felcserélhetőség tilalmát pedig egy példával szemléltethetjük. Nézzük az alábbi három mondatot.

a zenész keze része a zenésznek

a zenész része a zenekarnak

a zenész keze nem része a zenekarnak

Ez a három állítás mind igaz, viszont ha fent akarjuk tartani a partitív relációra megállapított tranzitivitás érvényességét, márpedig fent kell tartanunk, akkor ez a helyzet ellentmondásosnak tűnhet, hiszen a három állításban nem teljesül a tranzitivitás. A megoldás csak az lehet, hogy a fenti három kijelentésben alkalmazott partitív relációfogalmat kettéválasztjuk, és a két fogalmat külön-külön és csak a megfelelő helyen használjuk. Ha felveszünk egy 'nem-atomos_része' és egy 'atomos_része' relációt az általános része reláció alá, akkor a fenti kijelentéseket pontosíthatjuk.

a zenész keze nem-atomos_része a zenésznek

a zenész atomos_része a zenekarnak

a zenész keze nem atomos_része a zenekarnak

a zenész keze nem nem-atomos_része a zenekarnak

A négy állítás mindegyike megáll önmagában és nincs köztük ellentmondás sem, mert a tranzitivitást két eltérő reláció közt nem kell elvárunk.

A típusfogalmak és a partitív reláció kapcsolatáról eddig elmondottak természetesen elmondhatók a fogalmak terjedelmébe sorolt példányok esetében is, de akkor már más partitív relációt kell az individuális entitások között felvennünk. Mondhatjuk a következőket.

ez a konkrét fark része a Bodri kutyának

a Pejko nevű ló része a Béla ménésének

a Riska nevű szarvasmarha része a szomszéd gulyájának

Fejér megye része Magyarországnak

Az itt alkalmazott partitív relációfogalmakat természetesen ugyanúgy és ugyanazért fel kell bontani atomos és nem-atomos altípusokra, mint ahogy és amiért ezt tettük a típusfogalmakhoz tartozó 'része' reláció esetében. Új relációk bevezetésére pedig egyfelől azért van szükségünk, mert a relációk relátumai itt más kategóriákba esnek, mint az előző esetben, másfelől pedig értelmetlen lenne típusfogalmat használni individuális fogalmak között.

A gyűjtemények kategóriáján belül értelmezhető partitív reláció individuumok gyűjteményei közti rész-egész kapcsolatokat kell kezeljen, ami nem igazán teszi lehetővé, hogy ezen a területen atomos partitív relációt alkalmazzunk, bár ennek az állításnak elméleti bizonyítását még nem tudjuk megadni. A következő példát mutatjuk be az itt alkalmazható rész-egész relációra.

a magyarországi városok halmaza része a magyarországi települések gyűjteményének

Budapest, Szentendre és Pomáz része a pest megyei településeknek

Szentendre és Szolnok része az 'sz' betűvel kezdődő települése halmazának

Eddig egyszer sem adtunk arra példát, hogy a partitív alárendeltje (része) reláció mellett hogyan lehet értelmezni ennek inverz párját, a partitív fölérendeltje relációt. Most mutatunk egyet.

az 'sz' betűvel kezdődő települése halmaza tartalmazza Szentendre és Szolnok településeket

Ennek analógiájára a korábbi esetekre is hasonló módon lenne elvégezhető az inverzrelációk képzése. A fő ontológiai kategóriák, valamint a köztük kapcsolatot teremtő – a partitív reláció általános fogalma alá sorolható – altípusok bemutatása után vizsgáljuk meg alaposabban, hogy azon a szakterületen, ahol mindig is a térbeli viszonyok (tehát a partitív reláció) kérdéseivel foglalkoztak, a térinformatikában milyen tudás halmazódott fel, és ebből mit hasznosíthatunk mi a névtér projekt megvalósítása során.

4.1.1.2 Térinformatikai áttekintés

A helyzeti információk segítségével a meglévő, nagy mennyiségű gazdasági, társadalmi és környezeti információk felhasználását és használatuk hatékonyságát jelentős mértékben növelni lehet. Ezáltal a társadalom előtt álló kihívásokat jobban meg tudjuk érteni és kezelni, mellyel a fenntartható gazdasági, társadalmi és környezeti fejlődést támogatjuk.

A relációs adatbázis modell napjainkban általánosan elterjedt megoldásnak számít, a legtöbb GIS (Geography Information System) szoftver ezt a megoldást alkalmazza az alfanumerikus geometria típusú adatok tárolására.

A relációs modell lényege az a megközelítés, hogy a felhasználó számára áttekinthető és egyszerűen kezelhető táblázatokba lehessen elhelyezni az adatokat, amely nagyon hasonlít a hagyományos adatszervezés struktúrájához.

A modellben elhelyezkedő táblázatok között kapcsolatokat hozhatunk létre, így több, egymással összefüggő táblázat feldolgozásával olyan aggregált lekérdezéseket, virtuális táblázatokat hozhatunk létre, melyek segítségével az összetartozó információkat egyszerűen jeleníthetjük meg.

Az adatok elérésére lekérdező nyelvek alakultak ki, melyek mind logikájukban, mind pedig formális megjelenésükben az Oracle adatbázis kezelő rendszer SQL (Structured Query Language) lekérdező nyelvéhez hasonlítanak. A különböző adatbázisrendszerek esetében a struktúrált lekérdező nyelv különböző dialektusait használva férhetünk hozzá és módosíthatjuk adatainkat.

A korszerű alfanumerikus adatbázisokat SQL segítségével érhetjük el, de ami szintén fontos, hogy a földrajzi információs rendszerek (GIS) szoftverek szintén fel vannak készítve arra, hogy a téradatokat egy interfész segítségével közvetlenül relációs adatbázisban tárolhassák. Lehet ez az adatbázis akár lokális adatbázis (file-adatbázis pl. MS Access, SQLite) vagy nagy adatbáziskezelő rendszer egy szerveren (PostgreSQL, MySQL, MariaDB, Oracle...). Az alkalmazható adatbázisok és térinformatikai szoftverek köre mára jelentősen kibővült. A térbeli objektumok tárolására bevezetett adattípusok és azok kezelésének módja azonban a megoldások többségében az OGC (Open Geospatial Consortium) által kiadott szabványokon alapul, így a különböző megoldásokban az alapelvek jelentős része megegyezik. Az OGC célja egy olyan kollaboratív szervezet fenntartása, mely felel és közösen tevékenykedik azért, hogy a térbeli adatok kezelésére, megosztására alkalmas szabványokat fejlesszenek és hozzanak létre.

A mindennapi életben használt és tárolt adatok 80%-ának van térbeli vonatkozása és ennek megjelenítésére a térképet használjuk, mint általános eszközt. Mára a papír alapú térképektől ezen a területen is eljutottunk a digitális tartalmakig, amelyek szolgáltatására alkalmas adatrendszerek kidolgozása vált és válik szükségessé.

Kialakultak azok a legfontosabb térképtípusok, melyekre támaszkodva fejlődtek és működnek a különböző mérnöki tervező szervezetek. A térképek egy másik típusa azonban a gazdaságot, a szociális-foglalkoztatási szférát, a népesség-nyilvántartást, az egészségügyet, a környezetvédelmet, a meteorológiát és más tematikus ágazatot szolgál.

Leegyszerűsítve három nagy csoportra oszthatjuk az alkalmazott térképeket:

- geodéziai nagyméretarányú térképek
- topográfiai térképek
- tematikus térképek.

A nagyméretarányú térképek közvetlen méréseken alapulva készülnek mérnöki, illetve pontos nyilvántartási feladatok ellátására. Méretarányuk 1:500 és 1:5.000 közé esik. A mérési eredmények minimális általánosítással és szimbolikával kerülnek ábrázolásra. Az eltolt ábrázolás nem engedélyezett. A topográfiai térképek méretaránya 1:10.000-tól 1:200.000-ig terjed. A méretarány csökkenésével az általánosítás foka nő, azaz a valós tartalmat egyszerűsítéssel csökkentik. A Föld felszín mesterséges és természetes objektumainak ábrázolása mellett adminisztratív, gazdasági tematikákat is tartalmazhat. Ennek megfelelően ábrázolásmódja gazdag, melyet színek és szimbólumok segítségével valósít meg.

A nagyobb méretarányú térképek esetében közvetlen felmérést alkalmaznak, azonban kisebb méretarány esetében már kartográfiai módszerekkel, általánosítással hozzák létre a térkép tartalmát. Ezek a térképek alkalmazhatnak eltolást, torzítást melynek hatására a térképen megjelenített objektumokról nem lehet pontosan tudni, hogy az a valós objektum pontos helyét ábrázolja-e. Ez megengedett ennél a térképtípusnál, mivel a térkép feladata a tájékoztatás.

A kutatás szempontjából legfontosabb térképtípus, a tematikus térkép, amely gyakran kis méretarányú (1:500.000-1:2.000.000). Ezeket a térképeket gyakran áttekinthető céllal készítik, azzal a céllal, hogy

globális jelenségeket vizuálisan ábrázolni és értelmezni lehessen. Azok a jelenségek, amelyeket általában térképen szeretnénk ábrázolni nagy területekre vonatkoznak, így ezek megjelenítésére nagyméretarányú térképek használata nem indokolt. Ezen kívül egy adott tematika csak egy szűk felhasználói kör részére készül, így ehhez nagyméretarányú alaptérkép létrehozása nem rentábilis. Ennek a problémának a megoldására az a gyakorlat alakult ki, hogy a nagyméretarányú földmérési alaptérképet felhasználva vagy topográfiai térképeket felhasználva saját fedvényeket létrehozva jeleníthetők meg a tematikus adatok.

A térképezés és a térbeli adatok gyűjtése korábban hosszadalmas folyamat volt, a technológia fejlődése azonban ezen a területen is sokat gyorsított a folyamatokon. A GPS, a különböző légifelvételési technológiák elterjedése és a nagyfelbontású műholdképek megjelenítése új adatforrásokat biztosított a térképészet számára digitális formában. Ezekre alapozva a különböző térbeli megjelenítési szintekhez megfelelő alaptérkép létrehozása egyszerűbbé válik.

A korszerű adatbázis koncepcióban szereplő adatok alfanumerikus adatok voltak. Természetesen ezek az adatok jelölhettek földfelszínen elhelyezkedő pontokat, vonalakat vagy területeket (házszám, kerület stb.). De ezek térbeli paraméterei illetve az ezek közötti térbeli összefüggések leírása nem volt megoldott (pl. két utca közötti távolság meghatározása az utcanevekből nem volt lehetséges, vagy egy adott tulajdonsággal rendelkező objektum területe nem volt kiszámolható). Erre a problémára adnak megoldást a földrajzi információs rendszerek.

4.1.1.3 Relációs adatbázisok geometria kezelésének lehetőségei

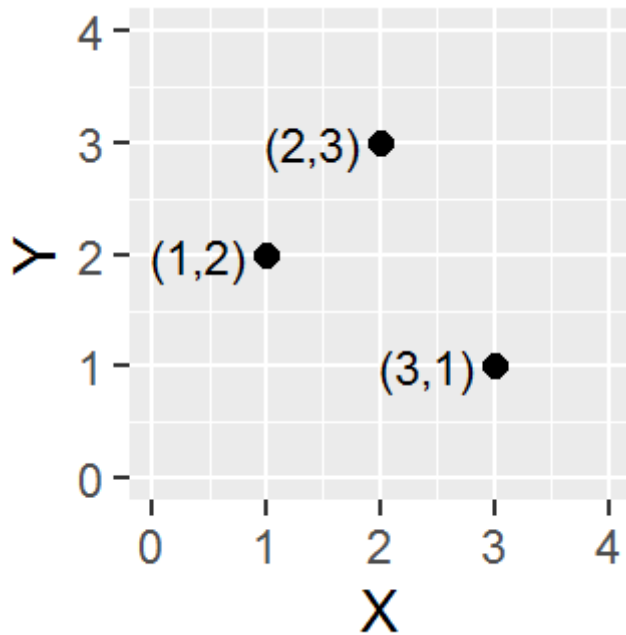
A térbeli adatok tárolását úgy kell megoldani az alfanumerikus adatbázisban, hogy a szükséges információkat a meglévő eszközrendszer felhasználásával valósítsuk meg. Ehhez új adattípusok bevezetésére volt szükség az adott rendszeren belül. A legfontosabb ilyen típus a geometria, amely a relációs adatbázis egy táblájában definiált mező, ez tartalmazza a geometria fizikai leírását (bináris formátumban). Az értelmezhetőség érdekében természetesen szükségesek olyan funkciók, amely az emberi szem számára is értelmezhető formában jeleníti meg a geometriát leíró objektumok sorozatát (pontok és vonalak). A geometria típuson belül a földrajzi objektumok alábbi szupertípusait (supertype) érdemes megkülönböztetni, úgymint:

- pontszerű (csomópont, point, node – 0 dimenziós)
- vonalszerű (görbe, ív, él, arc, line – 1 dimenziós)
- felületszerű (poligon, surface, polygon, area – 2 dimenziós)

4.1.1.3.1.1 Pontszerű (point)

Egy lokáció a térképen, amely definiálja egy objektum pozícióját (pl. egy jármű helye).

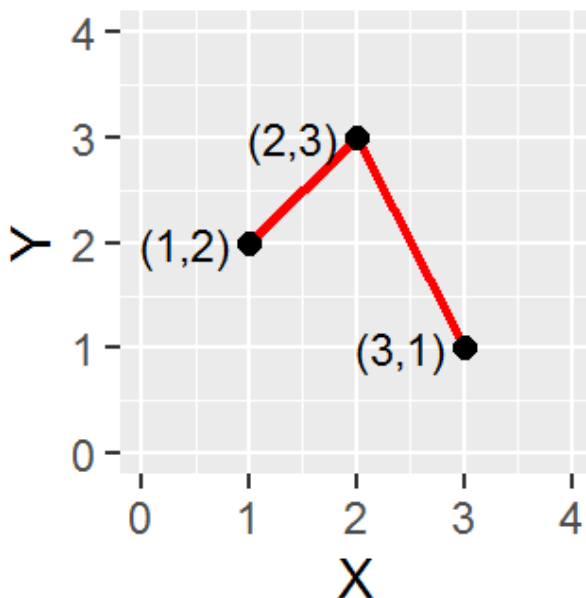
Az önálló pont a valóság modellezése során kapott olyan objektumok leírására szolgál, melyek területi kiterjedése elhanyagolható. Ilyen objektumok lehetnek a kutak, források, geodéziai alappontok, tv-, vagy rádióadó antennák, hidroglóbuszok, kémények, vezetéktartó oszlopok stb.



Három pont objektum az X, Y koordinátáikkal definiálva Euklidészi koordináta rendszerben

4.1.1.3.1.2 Vonalszerű (polyline)

A terepi objektumok egy másik csoportja vonalakkal modellezhető, ilyenek a vízfolyások, utak, vasutak, csővezetékek, föld alatti és föld feletti kábelek stb. A vonalas objektumok töréspontokat tartalmaznak, melyek valamely (általában lineáris) törvényszerűségeen alapuló összekötése szolgáltatja a vonalas objektumot. Pontok sorozatát definiálja, melyeket sorrendben összekötve egy vonalat reprezentálnak (pl. út vagy folyó).

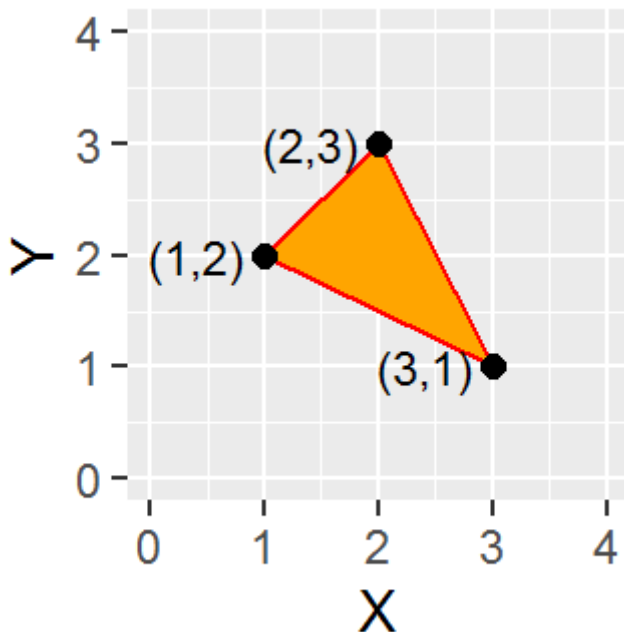


Vonallánc a csomópontok összekötésével definiálva

4.1.1.3.1.3 Felületszerű (polygon)

A poligonok általában zárt, egybefüggő felületeket reprezentálnak (pl. országhatár, parkok, épületek, egy tó körvonala), de lehetséges olyan felületek ábrázolása is, melyek lukakat tartalmaznak, vagy több

különálló egységből állnak. A poligonok pontjait koordináta-párokkal definiáljuk, az első és az utolsó pont megegyezik.



Egy poligon a kapcsolódó élek által közrefogott területtel definiálva

4.1.1.3.2 Téradatok kezelése

Ahhoz, hogy a fent definiált geometria típusokat egy adatbázisban kezelni tudjuk, olyan eszközök bevezetésére van szükség, amelyekkel az alapvető menedzsment (geometria típus létrehozása bemenetből, átnevezés, törlés, másolás, stb.) funkciókon kívül tényleges adatmanipulációt (kivonás, összeadás, területszámítás, stb.) is végrehajthatunk.

A lehetséges műveleteket, operátorokat és azok működését definiálni kell, azonban jelen dokumentumban csak azokra a műveletekre térünk ki, amelyek általában felhasználásra kerülnek a kutatáshoz kapcsolódó térképi adatok manipulációja során. Meg kell vizsgálni, hogy melyek azok az alapesetek, amelyeken keresztül téradatokat állítunk elő, téradatokhoz férünk hozzá vagy akár csak keresünk rá. Nézzünk ezekre néhány példát.

4.1.1.3.2.1 Alaptérkép betöltése

Ebben a használati esetben rendelkezésünkre áll egy vektoros térkép valamilyen vetületi rendszerben, melyet úgy szeretnénk egy adatbázisban tárolni, hogy a térképi adatot széles körben lehessen elérni és hivatkozni. Az alaptérkép megjelenítési szintje tematikus térképek esetében általában közigazgatási szintek definiálásával határozható meg (pl. megyetérkép, településtérkép, városrészek, országhatárok, kistérségek).

4.1.1.3.2.2 Alaptérkép létrehozása aggregálással

A fenti felsorolásból kitűnik, hogy ezek olyan térbeli objektumok, amelyek között valamilyen kapcsolat létesíthető (pl. tartalmazás, része, egésze reláció). És a definiált kapcsolat mentén egy "részletes" térképből előállítható egy általánosított megjelenítési szintje az adott térképnek. Például egy településhatáros térképből a megyéket alkotó településlista alapján létre hozhatunk olyan poligonokat, melyek a tartalmazó megyéket alkotják.

4.1.1.3.2.3 Objektum keresése egy lokáció tartalmazása szerint

Egy GPS pozíció ismeretében szeretnénk megkeresni azokat a térbeli objektumokat, melyek a definiált pontot tartalmazzák.

4.1.1.3.3 Geometriák egymáshoz való viszonyának meghatározása

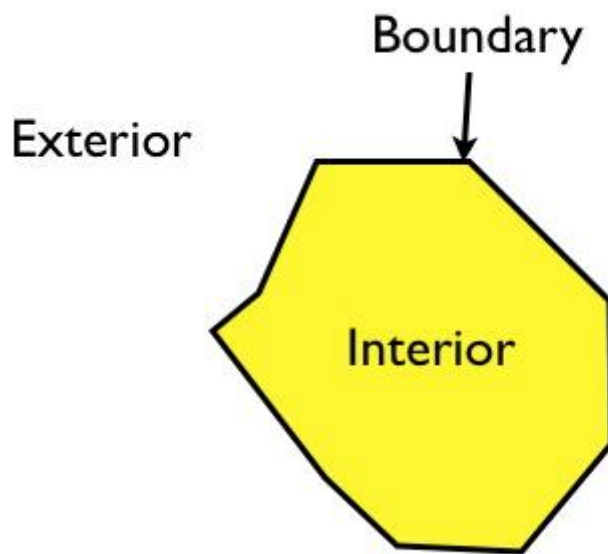
Két térbeli objektum közötti metszet leírására egy általánosan elfogadott keretrendszer a Dimensionally Extended nine-Intersection Model (DE-9IM).

Minden térbeli objektumhoz meghatározható:

Először is, minden térbeli objektumnak:

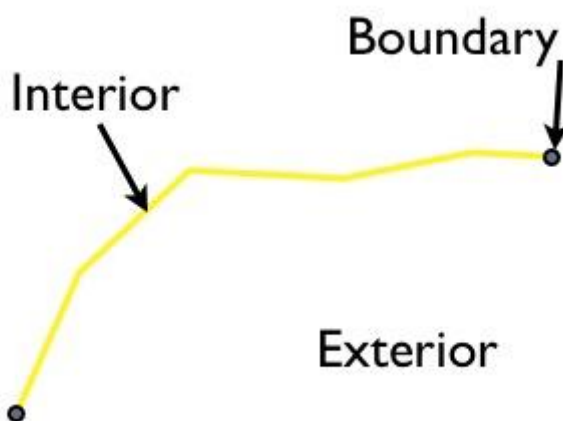
- a belső része (a határvonalon belül eső terület, interior)
- a határa (Boundary)
- a külső része (a határvonalon belül eső terület, exterior).

Poligon esetében ez a következő képpen ábrázolható:



A belső tér a gyűrűk által határolt rész; a határokon maguk a gyűrűk; a külső minden más a síkon.

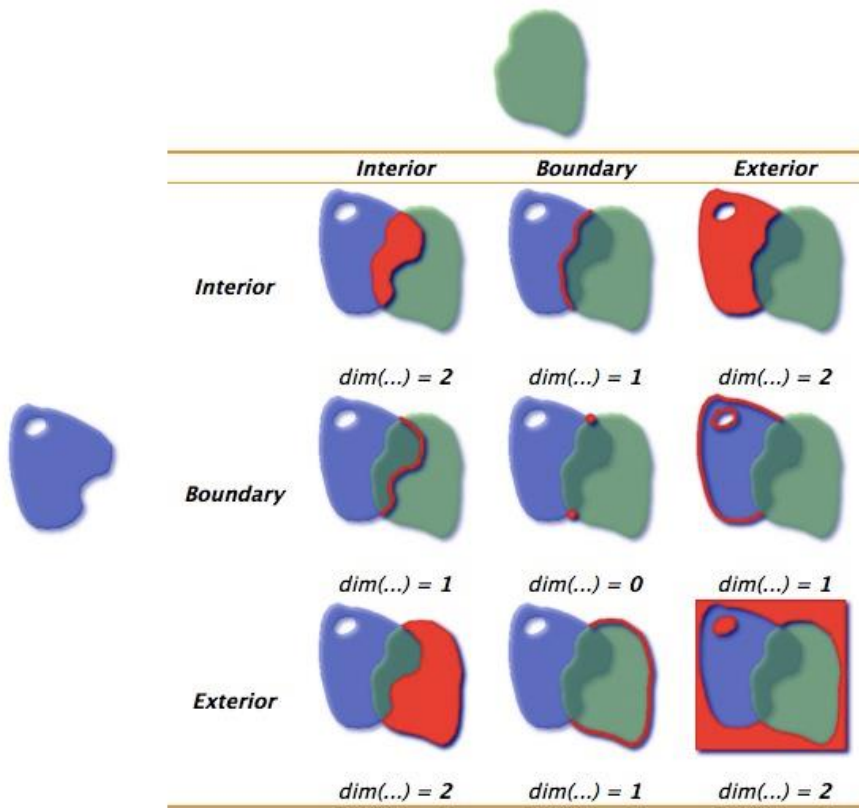
Vonalas objektum esetében a belső, a határoló és a külső nem ismertek:



A belső tér a vonal által határolt vonal része; a határ a vonal végei, és a külső minden más a síkon.

A pontok esetében: a pont maga a belső terület; a határ üres halmaz a külső rész minden más a síkon.

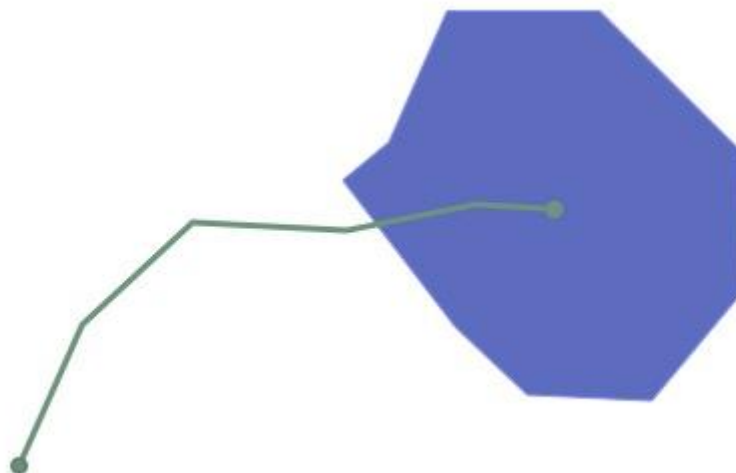
A belső rész, külső rész és a határ definícióját felhasználva, két térbeli objektum közötti kapcsolatot páronként az adott elemek kölcsönhatásának dimenziójával jellemezhetjük, ami kilenc lehetséges metszetet jelent a következők szerint.



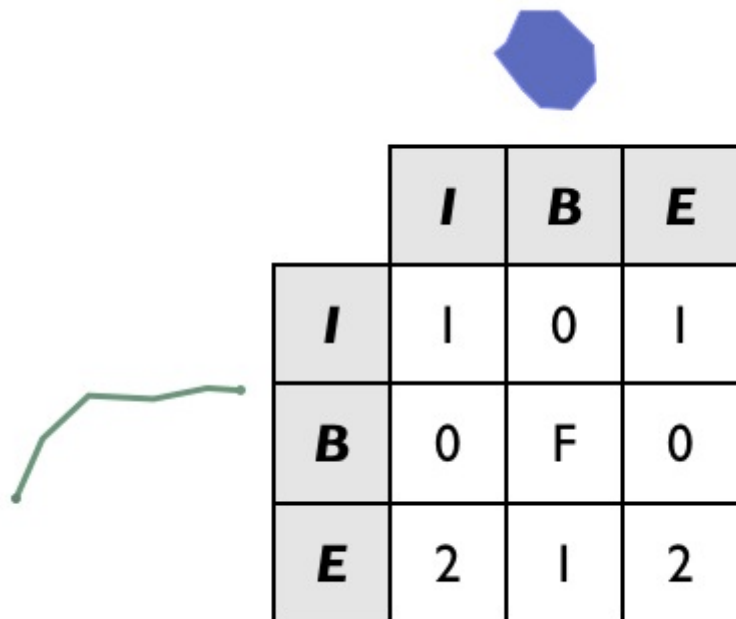
A fenti példában szereplő sokszögek esetében a belső részek metszéspontja 2-dimenziós terület, így a mátrix megfelelő pozíciójában "2" szerepel. A határvonalak egy pontban metszik egymást, ami 0 dimenziójú, így a mátrix megfelelő részét 0-val tölti ki.

Ha nincs kapcsolat az elemek között, ezt a mátrixban "F" jelöli.

Egy példa egy olyan vonalról, amely részlegesen belemetsz egy sokszögbe:



A kapcsolat DE9IM mátrixa a következő:



Vegyük észre, hogy a két objektum határai egyáltalán nem metszik egymást (a vonal végpontja kölcsönhatásban van a sokszög belsejével, nem pedig a határral, és fordítva), így a B/B cella értéke "F".

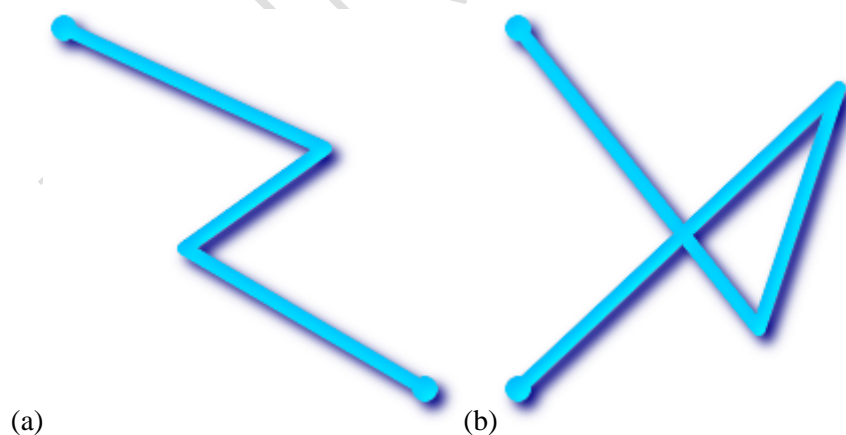
4.1.1.3.3.1 A geometriák OGC-kompatibilitásának biztosítása

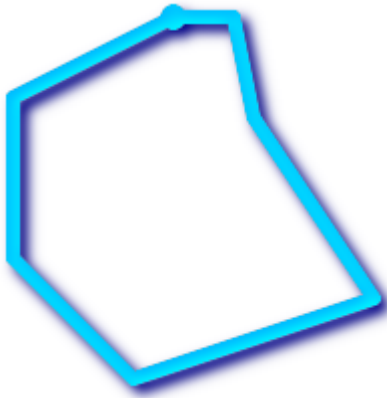
Az OGC-specifikációk szerint egy egyszerű geometria olyan, amely nem rendelkezik szabálytalan geometriai pontokkal, például saját kereszteződéssel vagy öntenzitással, és elsősorban 0 vagy 1-dimenziós geometriára utal (pl. [MULTI]POINT, [MULTI]LINESTRING). A geometria érvényessége viszont elsősorban kétdimenziós geometriára utal (pl. [MULTI]POLYGON) és meghatározza az érvényes poligonra jellemző állításokat.

A pont típus öröklődően egyszerű, mivel 0-dimenziós objektum.

A MULTIPOINT objektumok is egyszerűek, ha nincs két olyan pont koordináta, amelyek egyenlők (azonos koordinátaértékekkel).

A vonallánc (LINESTRING) objektum egyszerű, ha nem halad át kétszer ugyanazon a ponton (kivéve a végpontokat, mely esetben lineáris gyűrűnek nevezik az objektumot és zártnak is tekinthető).





(c)



(d)

(a) és **(c)** egyszerű LINESTRING s, **(b)** és **(d)** nem.

A többelemű vonallánc (MULTILINESTRING) csak akkor egyszerű, ha az összes eleme egyszerű, és két elem közötti metszéspont az elemek határvonalán helyezkedik el.



(e)



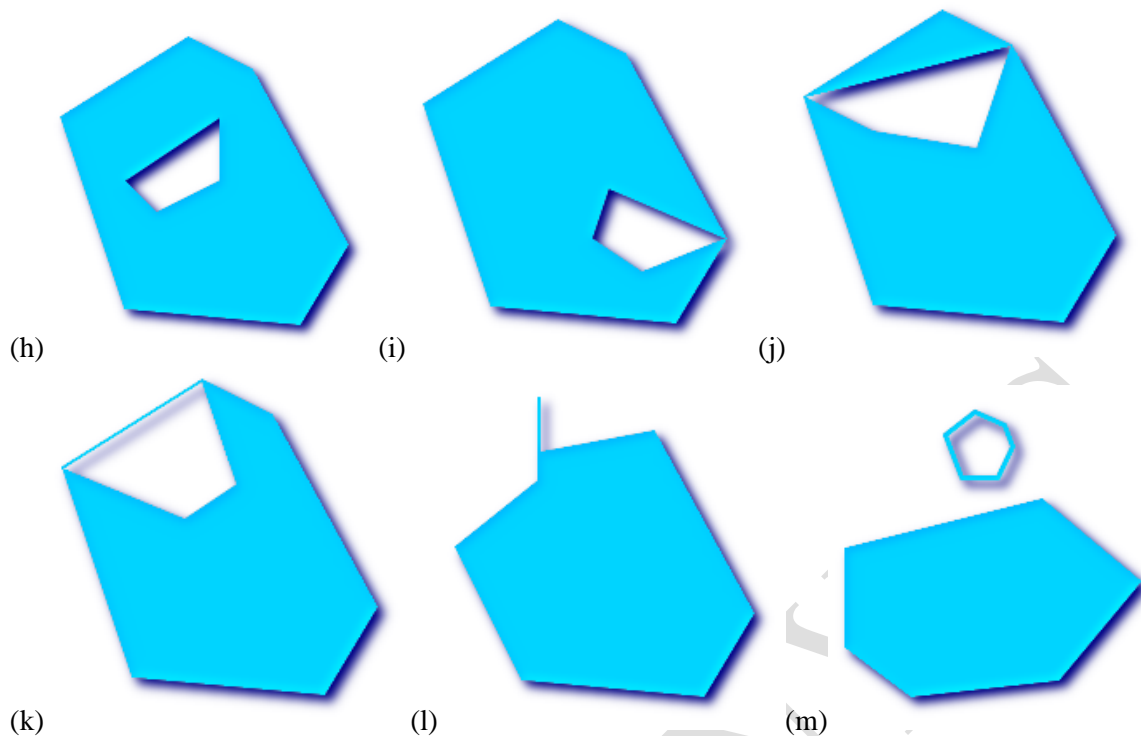
(f)



(g)

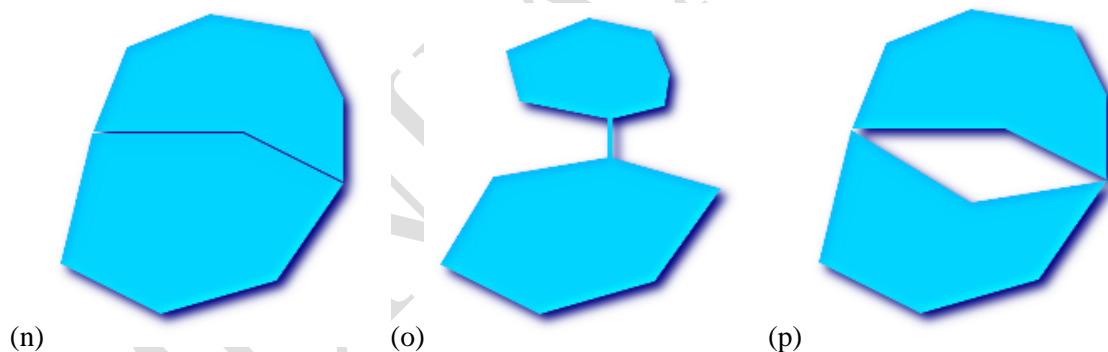
(e) és **(f)** egyszerű MULTILINESTRING s, **(g)** nem.

Definíció szerint a POLYGON mindig egyszerű. Akkor érvényes, ha a határvonalon egy gyűrűnek nincs két metsző pontja (külső gyűrűből és belső gyűrűből áll). A POLYGON határai nem lehetnek megszakítva nem tartalmazhat tüskéket és a belső gyűrűknek teljesen a külső gyűrűn belül kell lennie.



(h) és **(i)** érvényespoligonok, **(j-m)** nem lehet reprezentálni egyszerű poligonnal, de **(j)** és **(m)** multipolygonként érvényesen reprezentálható.

Egy MULTIPOLYGON csak akkor érvényes, ha és csak akkor, ha az összes eleme érvényes. A két elem határai érinthetik egymást, de csak véges számú pontban.



(n) és **(o)** nem érvényes MULTIPOLYGON-ok, **(p)** azonban érvényes.

4.1.1.3.4 Térbeli adatok műveletei

Két geometria egymáshoz való viszonyát tudjuk meghatározni az adott geometriák között értelmezett műveletek bevezetésével. Azokat a legfontosabb műveleteket tekintjük át, melyeket általában alkalmazunk a térbeli adatok kezelésénél.

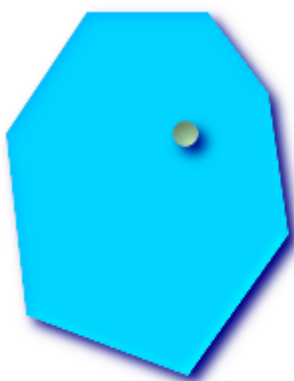
4.1.1.3.4.1 Tartalmazás (Contains)

Értéke akkor és csak akkor igaz, ha B geometriának nincs olyan pontja, amely az A geometrián kívül helyezkedik el, azaz, ha A geometria teljes egészében tartalmazza B geometriát-.

Példák "Igaz" értékre:



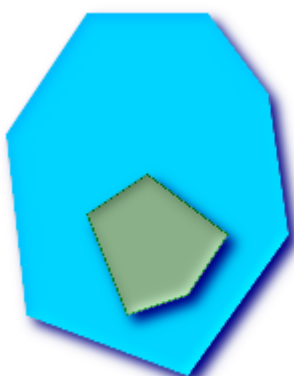
Vonallánc/ Multipont



Poligon/ Pont



Poligon / Vonallánc



Poligon / Poligon

Példák "Hamis" értékre:



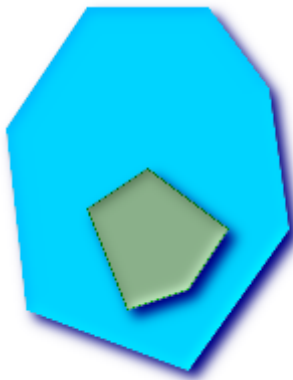
Poligon / Multipont



Poligon/ Vonallánc

4.1.1.3.4.2 Metszet (Intersect)

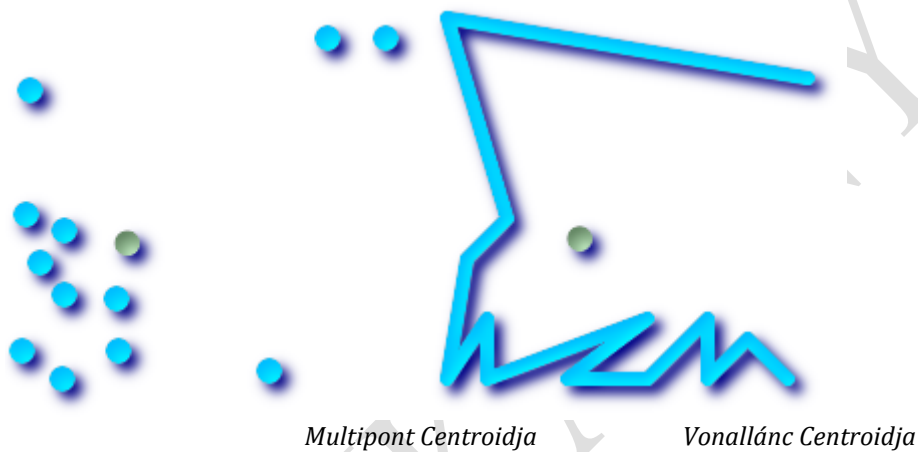
Visszatérési értéke "Igaz", ha geometriák egy része térben ugyanazon a helyen helyezkedik el. Minden geometria típuson elvégezhető a művelet. Az alábbi példa poligonokra vonatkozik.



Poligon / Poligon

4.1.1.3.4.3 Középpont (Centroid)

A mértani középpontját határozza meg az adott geometriának.



Multipont Centroidja

Vonallánc Centroidja



Poligon Centroidja



Geometria gyűjtemény Centroidja

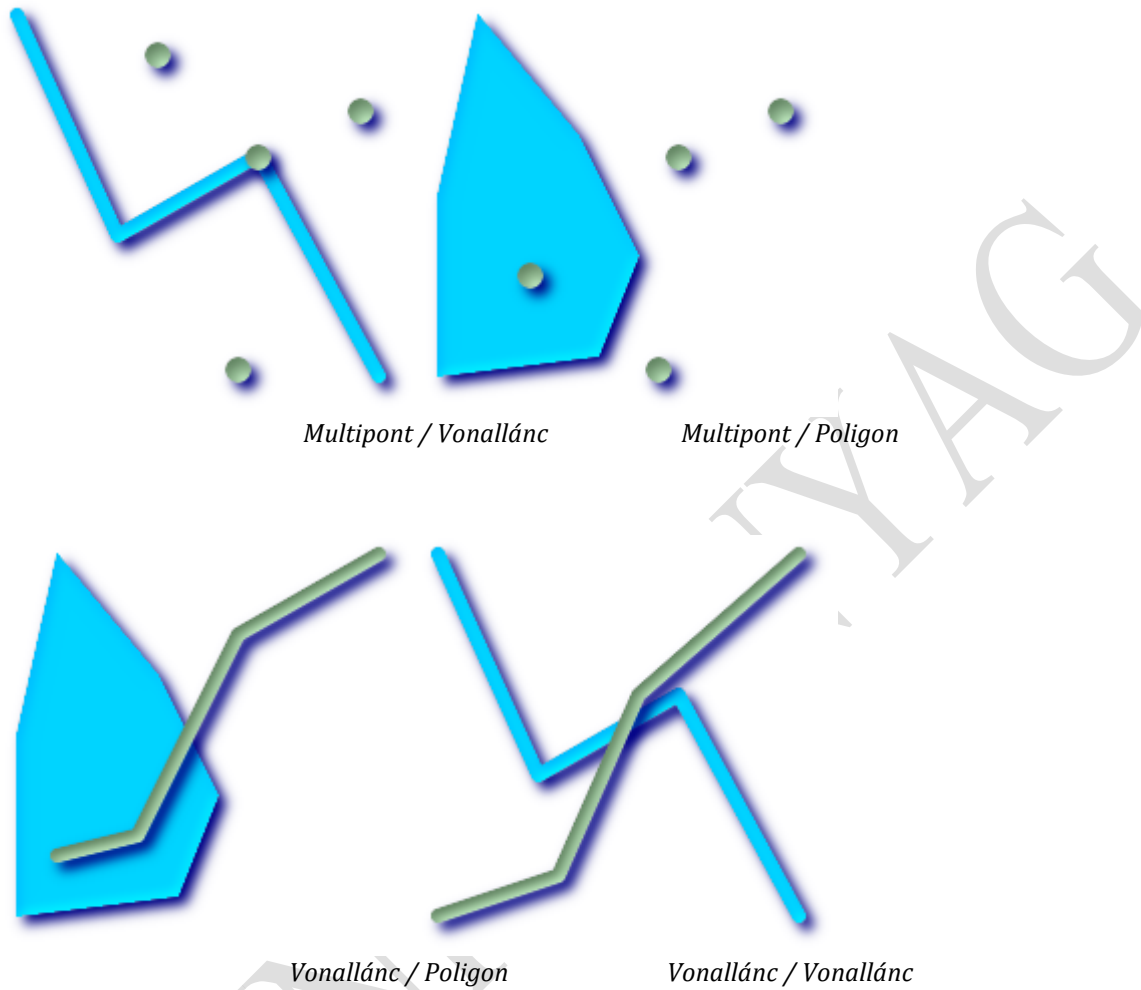
4.1.1.3.4.4 A felület egy pontja (PointOnSurface)

A Centroid művelethez hasonlóan egy pontot ad vissza, de ebben az esetben garantált, hogy az adott geometria felületén helyezkedik el az eredményben kapott pont. Alkalmazható pl. egy település területén lévő pont meghatározására.

4.1.1.3.4.5 Keresztesz (Crosses)

Visszatérési értéke "Igaz", ha a bemenetként megadott poligonok belső pontjai között van olyan (de nem az összes), ami egybeesik.

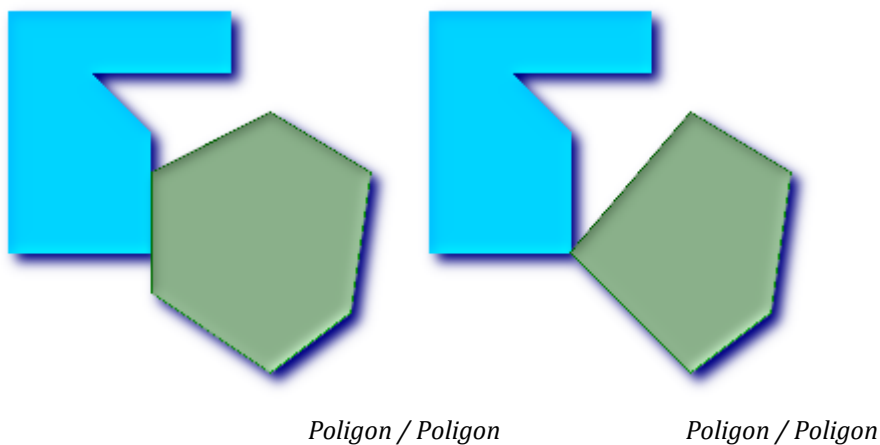
Példák "Igaz" visszatérési értékre:



4.1.1.3.4.6 Érintkezés (Touches)

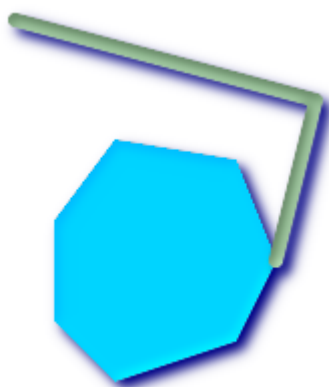
Visszatérési értéke "Igaz", ha a geometriák legalább egy pontja közös, de a belső pontjaiknak nincs közös része.

Példák "Igaz" visszatérési értékre:





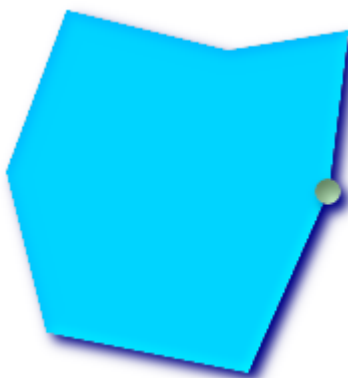
Vonallánc / Vonallánc



Poligon / Vonallánc



Vonallánc / Vonallánc



Poligon / Pont

4.1.1.3.4.7 Befoglaló téglalap (Envelope)

A paraméterben bemenetként megadott geometria befoglaló téglalapját határozza meg a geometriát alkotó pontok koordinátái alapján, a következő négy koordinátapár megadásával: ((MINX, MINY), (MINX, MAXY), (MAXX, MAXY), (MAXX, MINY), (MINX, MINY)).

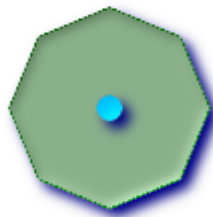
4.1.1.3.5 Geometriák feldolgozása

4.1.1.3.5.1 Pufferzóna generálás (Buffer)

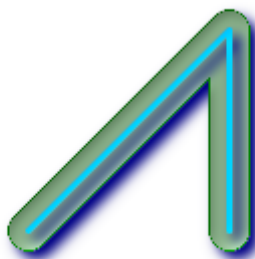
Eredménye egy geometria, amely lefedi a bemeneti geometriától adott távolságra lévő összes pontot. A bemenete bármely geometria típus lehet (pont, multi-pont, vonallánc, multi-vonallánc, poligon, multi-poligon).



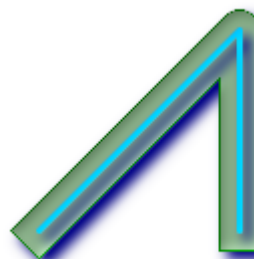
quad_segs=8 (default)



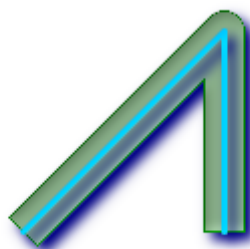
quad_segs=2 (lame)



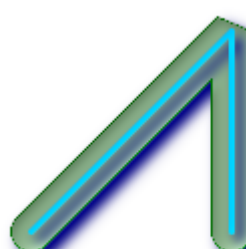
endcap=round join=round (default)



endcap=square



endcap=flat



join=bevel



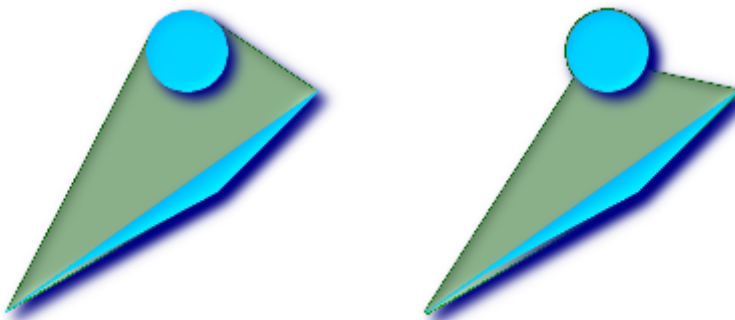
join=mitre mitre_limit=5.0 (default mitre limit) join=mitre mitre_limit=1

4.1.1.3.5.2 Unió (Union)

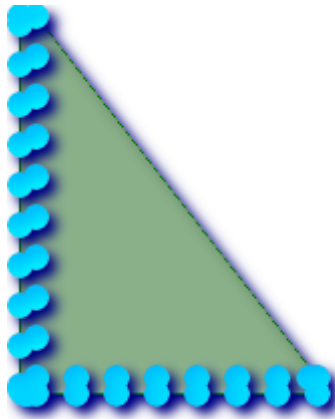
Egy geometriát ad eredményül, amely az összes poligon pont-halmazának összességét reprezentálja. Egy alternatív változata az eljárásnak a memóriakímélő megoldás, amely működése során kevesebb memóriát használ (MemUnion).

4.1.1.3.5.3 Konkáv befoglaló (ConcaveHull) / Konvex befoglaló (ConvexHull)

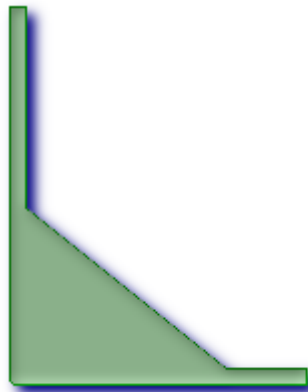
Egy geometria gyűjtemény konkáv/konvex befoglaló poligonja egy olyan konkáv/konvex geometriát (egy lehetséges megoldás) hoz létre, amely a geometria gyűjtemény összes geometriáját tartalmazza. Segítségével lehetővé válik egy geometriákat tartalmazó halmaz (pl. több település) egyetlen poligonnal történő közelítése. Adott esetben a megoldás gyorsabb lehet, mint az alkotó geometriák egyesítése (Union), ezért nagymennyiségű adat feldolgozásánál alkalmazható.



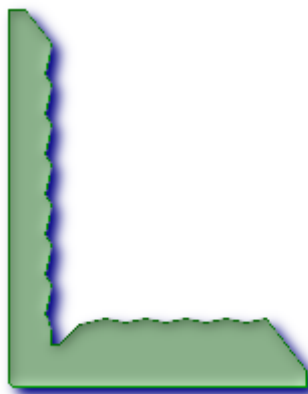
Konkáv befoglaló 2 poligonraencased és 90% szűkítéssel konkáv befoglaló



L alakzatú pontokra illesztett konvex befoglaló



L pontokra 99%-os konvex befoglaló



Konkáv befoglaló az L pontokra 80%-os konvex befoglaló terület



multilinesring illesztett konvex befoglaló



multilinesring illesztett konkáv befoglaló



Konvex befoglaló MultiLinestring és MultiPoint objektumokon

4.1.1.3.5.4 Különbségképzés (Difference)

Két geometria különbsége egy olyan geometria, mely az első geometria azon részét reprezentálja, amelynek a kivonandó második geometriával nincs metszete.

Ha A teljesen tartalmazza B geometriát, akkor az eredmény egy üres halmaz. A következő módon is fel lehet írni a problémát: $\text{GeometriaA} - \text{Metszet}(\text{GeometriaA}, \text{GeometriaB})$



A két eredeti vonal

A két vonal különbsége

4.1.1.3.5.5 Szimmetrikus Különbségképzés (SymDifference)

Az eredményhalmaz azokat a pontokat tartalmazza a két geometriából, ahol a két geometria nem metszi egymást. Azért szimmetrikus, mert az operandusok felcserélhetők, azaz

$\text{SymDifference}(A,B) = \text{SymDifference}(B,A)$. A következő módon is fel lehet írni a problémát: $\text{Union}(\text{GeometriaA}, \text{GeometriaB}) - \text{Metszet}(\text{GeometriaA}, \text{GeometriaB})$.



A két eredeti vonal



A két vonal szimmetrikus különbsége

4.1.1.3.6 Topológia téradatbázisban

A topológiákat eredetileg térbeli adatszerkezetekként kezelték, amelyet elsősorban annak biztosítására használnak, hogy a kapcsolódó adatok konzisztens és tiszta topológiai kapcsolatokat képezzenek.

A térinformatikai rendszerek fejlődésével a topológia alternatív nézetei alakultak ki. A földrajzi adatbázisok ma már támogatják a földrajzi modellezés megközelítését, amely egyesíti a különböző szolgáltatás típusok viselkedését és támogatja a különböző típusú kapcsolatokat is (objektumorientált téradatbázisok). Ebben az összefüggésben a topológia olyan szabályok és kapcsolatok gyűjteménye, amelyek – a szerkesztési eszközök és technikák egy csoportjával együtt – lehetővé teszik a modellezett objektumok/geometriák valóságnak is jobban megfelelő, pontosabb leképezését a geoadatbázisban.

A topológiát leginkább az adatminőség biztosítására használhatjuk, lehetővé teszi, hogy a földrajzi adatbázis valóságként reprezentálja a földrajzi objektumokat.

A geoadatbázis olyan keretrendszert biztosít, amelyen belül az objektumok rendelkezhetnek altípussal, alapértelmezett értékekkel, attribútumtartományokkal, érvényesítési szabályokkal és strukturált kapcsolatokkal adattáblákhoz vagy más objektumokhoz. Ezek a tulajdonságot lehetővé teszik a geoadatbázisban található objektumok közötti hivatkozási integritás megőrzését. A topológia a viselkedési keret kiterjesztésének tekinthető, amely lehetővé teszi a funkciók közötti geometriai kapcsolatok ellenőrzését és a geometriai integritás fenntartását.

Egy topológia létrehozásához a geoadatbázis tervezése szükséges. A topológia rendszerezi a geometriák között lévő térbeli kapcsolatokat. A tervezőnek elemeznie kell a szervezet adatmodellezési igényeit, azonosítja a geoadatbázisban szükséges legfontosabb topológiai kapcsolatokat és meghatározza azokat a szabályokat, amelyek definiálják a különböző geometriák topológiai kapcsolatát.

Egy geoadatbázist folyamatosan használnak és karbantartanak, új objektumokat adnak hozzá, és a meglévő objektumok módosulnak. Ezeket a feladatokat általában adatszerkesztők végzik a geoadatbázisban frissítik az objektumokat és a topológiát használják az adatbázis-tervező által megszabott korlátok közötti kapcsolatok kialakítására és fenntartására. A szervezet munkafolyamatától függően a topológia minden módosítás után ellenőrizhető és érvényesíthető.

A topológia létrehozásakor meg kell adni a topológia szabályban részt vevő objektumosztályokat. Ezek az objektumosztályok tartalmazhatnak pont-, vonal- vagy sokszög-objektumokat. A topológiában a geometriai kapcsolatok nem a sajátosságok, hanem a funkciók részei.

A topológiában a sokszögek a következő objektumokból állhatnak:

- Élek, amelyek a poligonok határait definiálják,
- Metszéspontok, két él keresztezésénél,
- Csomópontok, amelyek meghatározzák a poligon alakját.

Hasonlóképpen, egy vonal objektum egy élből, és legalább két csomópontból áll, amelyek meghatározzák az él végpontjait és az él alakját meghatározó csúcsokat. A pont-objektumok csomópontként viselkednek.

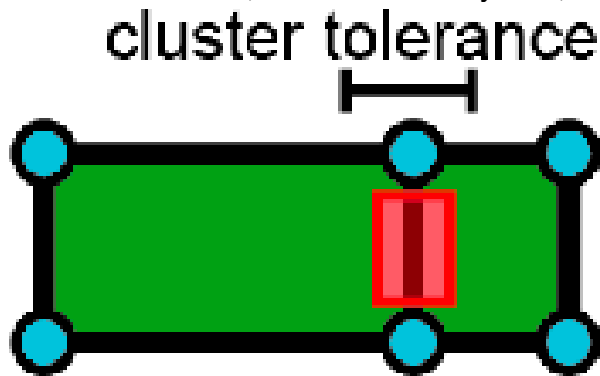
4.1.1.3.6.1 Topológia szabályok

Számos topológia szabályt lehet előírni a geoadatbázisban található objektumokra. Egy jól megtervezett geoadatbázis csak azokat a topológiai szabályokat fogja tartalmazni, amelyek meghatározzák a kulcsfontosságú térbeli kapcsolatokat. A továbbiakban áttekintjük az alkalmazható fő szabály típusokat.

4.1.1.3.6.1.1 Poligonokra vonatkozó szabályok

4.1.1.3.6.1.1.1 Adott értéknél nagyobb terület/oldalhosszúság

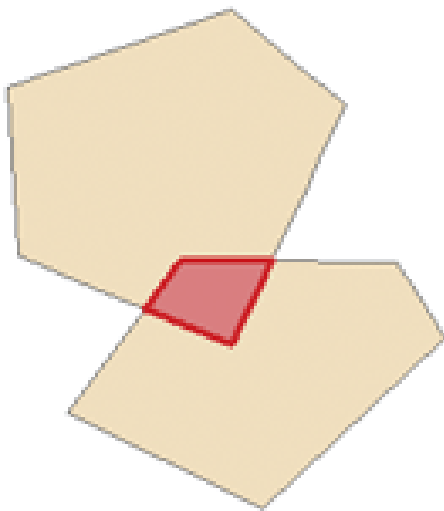
Ez a szabály kötelező a topológiában, és minden vonal- és poligon-szolgáltatási osztályra érvényes. Azokban az esetekben, amikor ez a szabály sérül, az eredeti geometria változatlan marad.



Ha a beállított tolerancia értéknél kisebb adódna, akkor hibajelzés keletkezik. Az ilyen jellegű hiba javítása törléssel lehetséges.

4.1.1.3.6.1.1.2 Nem lehet átfedés két objektum között

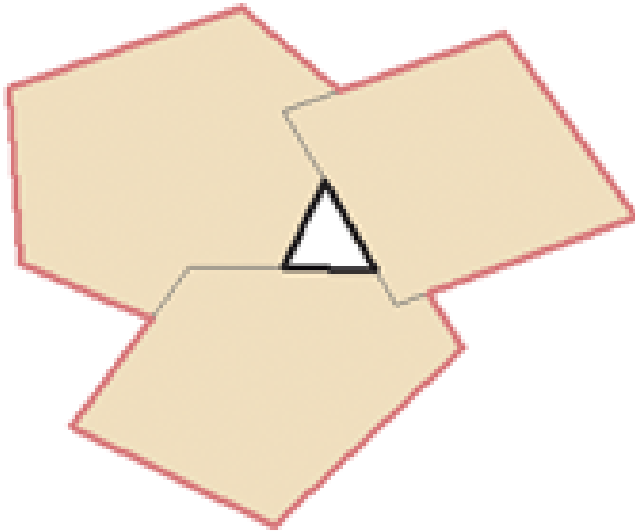
Megköveteli, hogy az egyes objektumok poligonjai nem lapolódhatnak át. A poligonoknak lehetnek közös éleik vagy csúcsaik. Ezt a szabályt akkor használják, ha egy terület nem tartozhat egyszerre kettő vagy több poligonhoz. Hasznos az adminisztratív határvonalak, például a irányítószámkörzetek vagy szavazóköri és a kölcsönösen kizáró területek reprezentálásához, például a földhasználat vagy a felszínborítottság modellezéséhez.



A piros terület mindkét poligonnak része (az objektumok típusa miatt ez nem megengedett). Az ilyen jellegű hiba javítása a következő módon lehetséges: kivonás, egyesítés, új objektum létrehozás.

4.1.1.3.6.1.1.3 Nem lehetnek rések

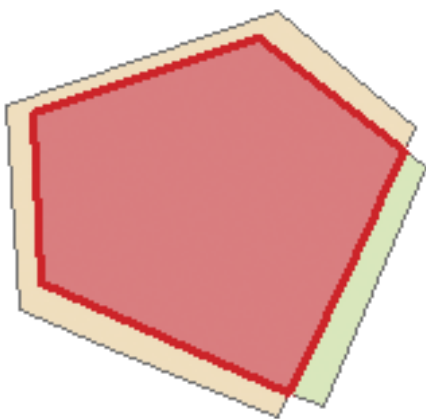
Ez a szabály megköveteli, hogy egyetlen poligonban vagy szomszédos poligonok között ne legyen üres, nem lefedett rész. Minden poligonnak folyamatos felületet kell alkotnia. A felület kerületén mindig hiba fog megjelenni, de ezt a hibát figyelmen kívül lehet hagyni, vagy kivételként megjelölhető. Ez a szabályt olyan adatokra használható, amelyeknek teljesen le kell fedniük a területet. Például a talajtípusok nem tartalmazhatnak hiányosságokat vagy üregeket, hanem a teljes területet le kell fedniük.



A poligonok közötti üres terület nem megengedett. Az ilyen jellegű hiba javítása új objektum létrehozásával lehetséges az üres területen.

4.1.1.3.6.1.1.4 Nem lehet átfedés

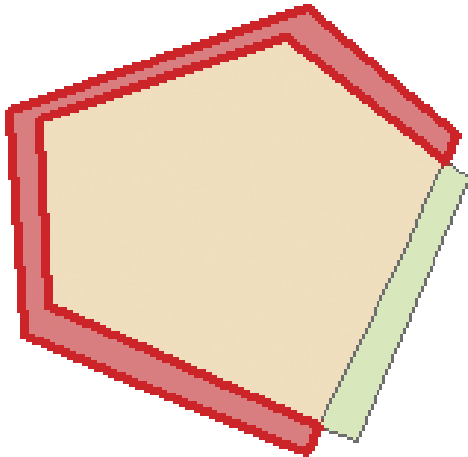
A szabály megköveteli, hogy adott objektumtípushoz tartozó poligonok belseje és egy másik objektumtípus poligonjának belseje között ne legyen átfedés. A két objektumtípus poligonjainak lehetnek közös élei vagy csomópontjai de teljesen különállóak is lehetnek. Ezt a szabályt akkor használják, ha egy terület nem tarthat egyszerre két külön objektumtípusba. Hasznos két egymást kölcsönösen kizáró területosztályozási rendszert kombinálni, mint pl. A lakott terület és a víztesttípus, ahol a lakott területek nem definiálhatók a víztest-osztályban és fordítva.



Az ilyen jellegű hiba javítása az objektumok kivonásával, egyesítésével lehetséges.

4.1.1.3.6.1.1.5 Le kell fednie

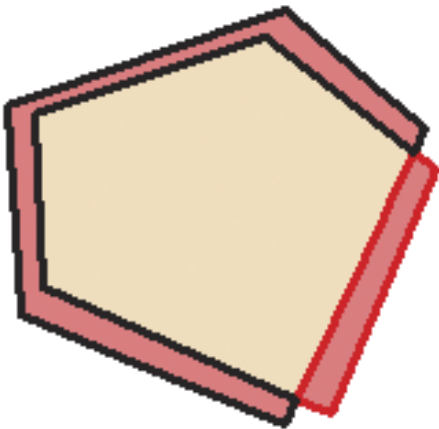
A szabály megköveteli, hogy az egyik objektumtípusú poligonnak meg kell osztania a teljes területét egy másik objektumtípusú poligonjaival. Az első olyan terület, amely kívül esik a másik objektumtípusú poligonjainak területén, hibát eredményez. Ezt a szabályt akkor alkalmazzák, ha egy adott típusú területet, például egy országot, más típusú közigazgatási területekre kell felosztani, de ezek az objektumok kötelezően a felettes területén belül helyezkednek el.



Az ilyen jellegű hiba javítása az objektumok kivonásával, új objektum létre hozásával lehetséges.

4.1.1.3.6.1.1.6 Fedjék egymást

A szabály megköveteli, hogy egy objektumtípusú poligonjainak területét teljesen le kell fednie egy másik objektumtípusú poligonjainak területével. A poligonok élei vagy csúcsai lehetnek azonosak. Ha van nem lefedett terület, az hibát eredményez. Ezt a szabályt akkor alkalmazható, ha két besorolási rendszert használnak ugyanazon földrajzi területre, és az egyik rendszerben definiált bármely pontot meg kell határozni a másik szerint is. Ilyen eset lehet hierarchikusadatrendszer esetén például a kis vízgyűjtők és a nagy vízvezető medencék egymásnak megfelelőit. A szabály alkalmazható nem hierarchikusan kapcsolódó poligon-jellemzőkre is, mint pl. A talajtípus és a meredekségi osztály.

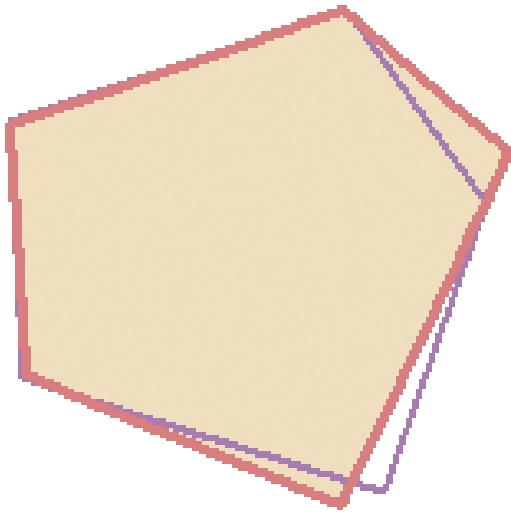


Az ilyen jellegű hiba javítása az objektumok kivonásával, új objektum létre hozásával lehetséges.

4.1.1.3.6.1.1.7 A határvonalaknak egybe kell esnie

A szabály megköveteli, hogy a poligon objektumok határaival egy másik osztályhoz tartozó vonalaknak egybe kell esniük. Ezt a szabályt akkor használják, ha a poligonokhoz olyan vonalas objektumokra van szükség, amelyek a területek határait jelölik. Ez általában akkor lehet szükséges, ha a poligonokhoz egy

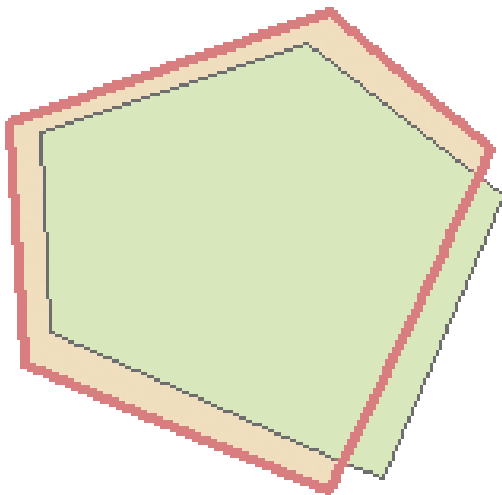
tulajdonságcsoporthat tárolunk, de a határvonalakhoz más attribútumok megadása szükséges. Például a kataszteri egységek egy földrajzi adatbázisban területükkel és a határaikkal is tárolhatók.



Az ilyen jellegű hiba javítása új objektum létre hozásával lehetséges.

4.1.1.3.6.1.1.8 Adott terület határát le kell fednie

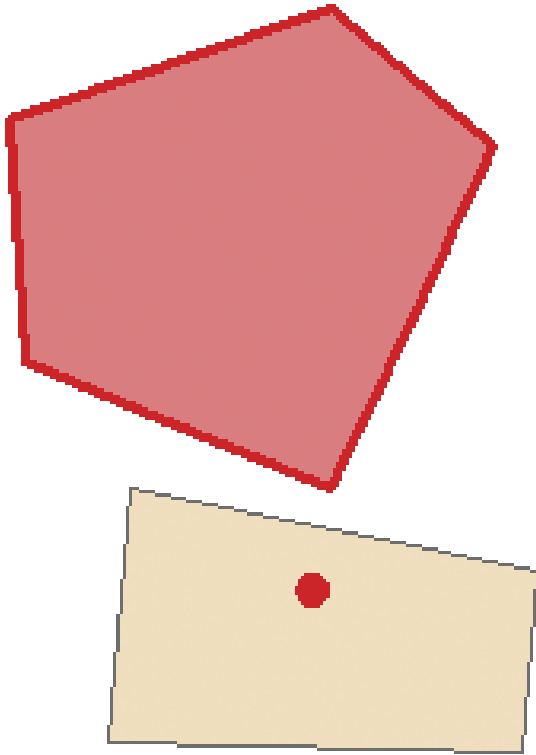
A szabály megköveteli, hogy a poligonok határait egy másik osztály poligonjainak határai lefedjék. Ez akkor hasznos, ha az egyik tulajdonságosztályban lévő poligonok, mint például a részosztályok, egy másik osztály több poligonjából, például települések és városrészek kapcsolata, és a megosztott határokat egymáshoz kell igazítani.



Az ilyen jellegű hiba javítása az objektumok módosításával lehetséges.

4.1.1.3.6.1.1.9 Pont tartalmazása

A szabály megköveteli, hogy egy poligon legalább egy pontot tartalmazzon egy másik objektumtípusból. A pontoknak a poligonon belül kell lennie, nem pedig a határon. Ez akkor hasznos, ha minden poligonnak legalább egy kapcsolódó pontnak rendelkeznie kell, például amikor a csomagoknak címzési ponttal kell rendelkezniük.

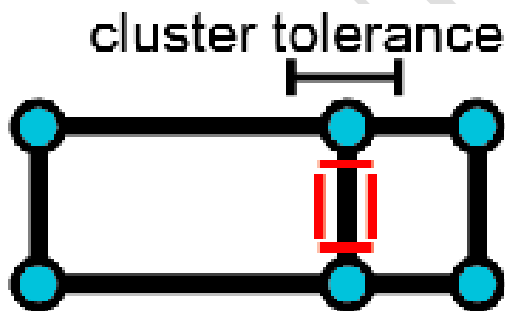


Az ilyen jellegű hiba javítása új objektum létre hozásával lehetséges.

4.1.1.3.6.1.2 Vonalakra vonatkozó szabályok

4.1.1.3.6.1.2.1 Adott értéknél nagyobb terület/oldalhosszúság

Ez a szabály kötelező a topológiában, és minden vonal- és poligon-szolgáltatási osztályra érvényes. Azokban az esetekben, amikor ez a szabály sérül, az eredeti geometria változatlan marad.

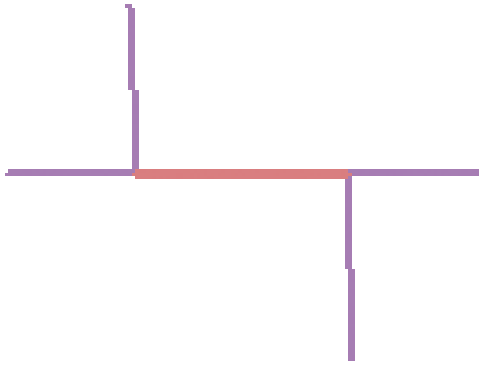


Ha a beállított tolerancia értéknél kisebb érték adódna, akkor hibajelzés keletkezik. Az ilyen jellegű hiba javítása törléssel lehetséges.

4.1.1.3.6.1.2.2

4.1.1.3.6.1.2.3 Nem lehet átfedés két objektum között

A szabály megköveteli, hogy a vonalak ne legyenek átfedve az ugyanazon típusba tartozó másik vonalakkal. Ezt a szabályt kell alkalmazni, ha a vonalszakaszokat nem szabad duplikálni; például vízfolyások esetében. A vonalak keresztezhetik, metszhetik egymást, de nem lehet közös szakaszuk.

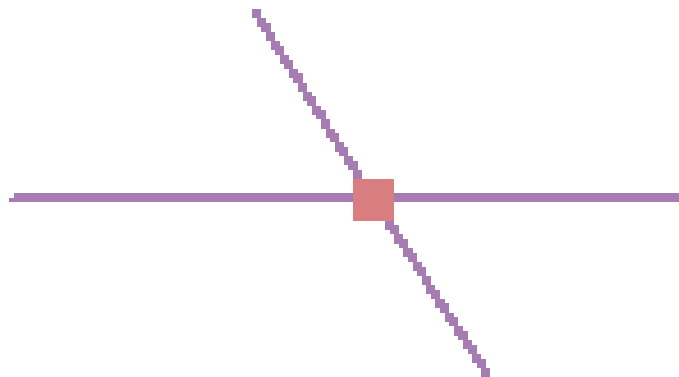


Az ilyen jellegű hiba javítása kivonással lehetséges.

4.1.1.3.6.1.2.4 Két objektum nem metszheti egymást

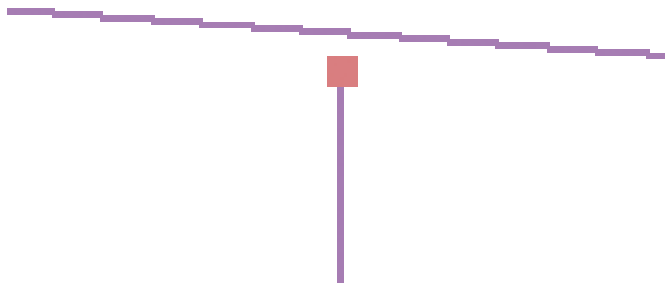
A szabály megköveteli, hogy ugyanabból az objektumtípusból származó vonalak fedjének át vagy metszhessék egymást. A vonalak tartalmazhatnak közös végpontokat. Ezt a szabályt olyan kontúrvonalakra használják, amelyeknek soha nem szabad egymáson áthaladniuk, vagy olyan esetekben, amikor a vonalak metszéspontja csak a végpontokban, például utcaszegmensekben és keresztezésekben fordulhat elő.

Az ilyen jellegű hiba javítása vágással, kivonással lehetséges.



4.1.1.3.6.1.2.5 Nem lehet nem kapcsolódó pont

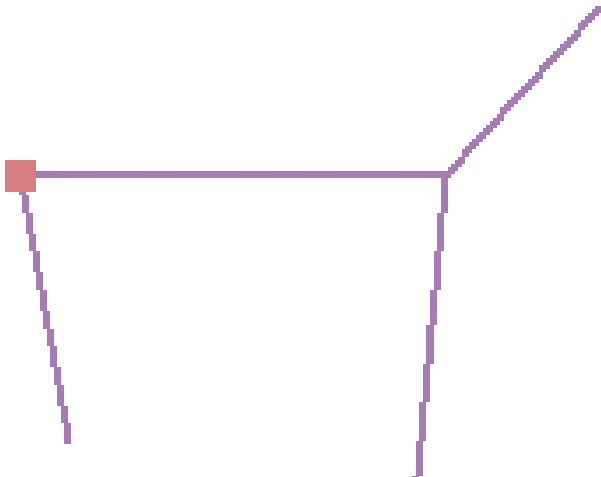
A szabály megköveteli, hogy az adott vonal mindkét végpontja ugyanazon objektumtípusban lévő vonalakra illeszkedjen. Ezt a szabályt akkor használják, ha a vonalkövetelményeknek zárt hurkot kell képezniük, például amikor meghatározzák egy poligon határait. Használható olyan esetekben is, amikor a vonalak általában más vonalához kapcsolódnak, mint az utcák esetében. Ebben az esetben kivételeket lehet alkalmazni, ha a szabályt időnként megsértik, például egy zsákutca esetében.



Az ilyen jellegű hiba javítása kiterjesztéssel, levágással vagy a pont "ragasztott" illesztésével lehetséges.

4.1.1.3.6.1.2.6 Nem lehet ál-csomópont

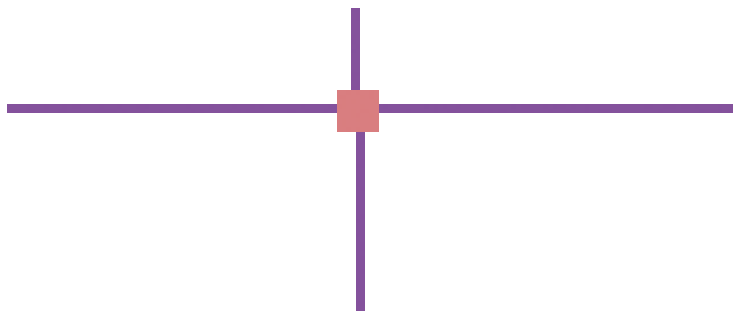
A szabály megköveteli, hogy egy vonal legalább két másik vonalhoz csatlakozzon mindegyik végpontnál. Az olyan vonalak, amelyek egy másik vonalhoz (vagy magukhoz) kapcsolódnak ál-csomópontokathoznak létre. Ezt a szabályt kell alkalmazni, ha a vonalas objektumnak zárt hurkot kell képezniük, például amikor meghatározzák a poligonok határait.



Az ilyen jellegű hiba javítása az objektum legnagyobbal való összeolvasztásával lehetséges.

4.1.1.3.6.1.2.7 Nem metszheti, vagy érintheti az objektum belsejét

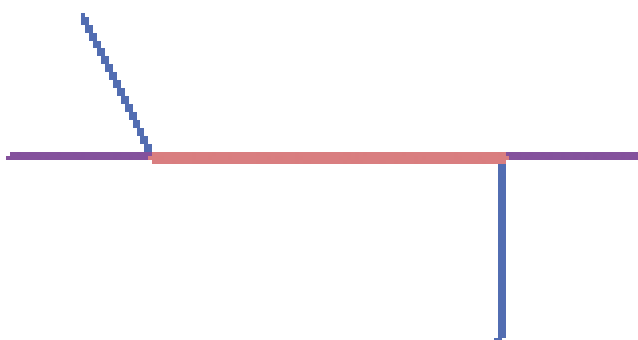
A szabály megköveteli, hogy az egyik objektumtípushoz tartozó vonal csak az azonos típusba tartozó másik vonalhoz kapcsolódhat. Bármely olyan vonalszakasz, amelyben az objektumok keresztezik egymást, vagy bármely metszéspont nem egy végpont, az hibát eredményez. Ez a szabály akkor hasznos, ha a vonalakat csak a végpontokba lehet csatlakoztatni, például földrészletvonalak esetében, amelyeknek osztottnak kell lenniük (csak a végpontok végén), és amelyek nem fedhetik át egymást.



Az ilyen jellegű hiba javítása kivonással és elvágással lehetséges.

4.1.1.3.6.1.2.8 Nem lehet átfedés

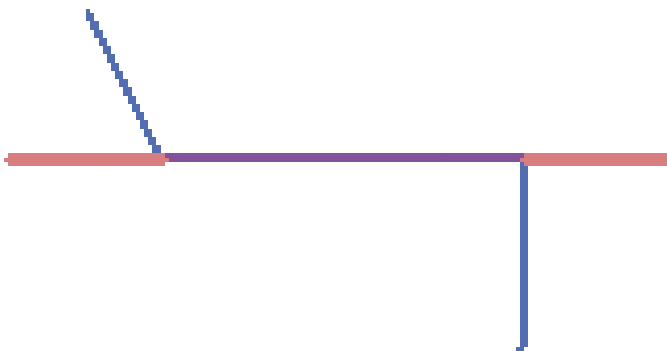
A szabály megköveteli, hogy adott objektumtípushoz tartozó poligonok belseje és egy másik objektumtípus poligonjának belseje között ne legyen átfedés. A két objektumtípus poligonjainak lehetnekközöségei vagy csomópontjai de teljesen különálóakis lehetnek. Ezt a szabályt akkor használják, ha egy terület nem tarozhat egyszerre két külön objektumtípusba. Hasznos két egymást kölcsönösen kizáró területosztályozási rendszert kombinálni, mint pl. A lakott terület és a víztesttípus, ahol a lakott területek nem definiálhatók a víztest-osztályban és fordítva.



Hibát okoz a lila vonalak átlapolódása. Az ilyen jellegű hiba javítása az objektumok kivonásával lehetséges.

4.1.1.3.6.1.2.9 Át kell fednie adott objektumtípussal

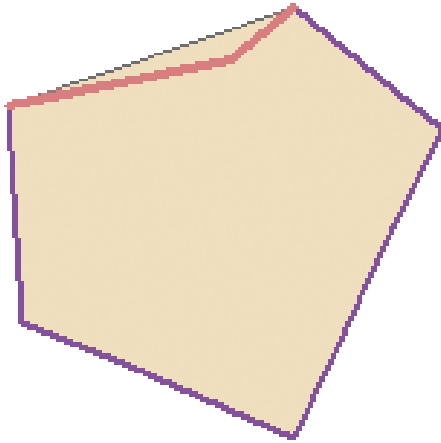
A szabály megköveteli, hogy az egyik objektumtípusból származó vonalakat egy másik típusban lévő vonalnak le kell fedniük. Ez hasznos logikailag különböző, de térben egybeeső vonalakat, például útvonalakat és utcákat modellezéséhez. A buszjárat jellemzői nem térhetnek el az utcai jellemzők osztályában meghatározott utaktól.



Ha a lila vonalak nem fedik egymást, hiba.

4.1.1.3.6.1.2.10 Egybe kell esnie a határvonallal

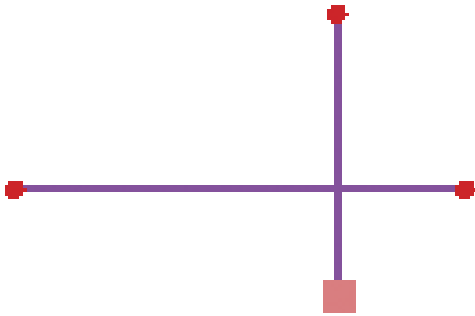
A szabály megköveteli, hogy a vonalakat a terület határai lefedjék. Ez hasznos vonalak modellezéséhez, például földrészlet vonalakhoz, amelyeknek egybeesőnek kell lenniük a területet reprezentáló poligonnal.



Az ilyen jellegű hiba javítása kivonással lehetséges.

4.1.1.3.6.1.2.11 A végpontot le kell fednie

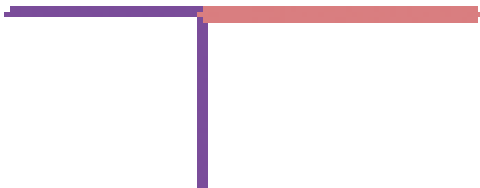
A szabály megköveteli, hogy a vonal végpontjait egy másik objektumtípus pontjaival le kell fedni. Ez olyan esetek modellezésénél hasznos, amikor a szerelvénynek két csövet kell összekötnie, vagy két utca találkozásánál utcai kereszteződést kell találnia.



Az ilyen jellegű hiba javítása új objektum létrehozásával lehetséges.

4.1.1.3.6.1.2.12 Nem fedheti át saját magát

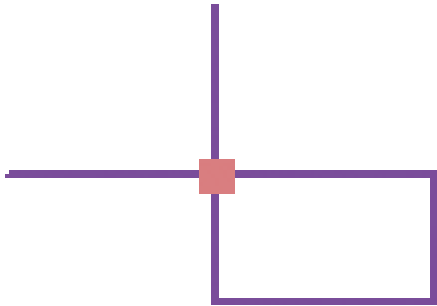
A szabály megköveteli, hogy a vonalszakaszok nem fedik egymást. Metszheti vagy érintheti saját magát, de nem lehetnek egybeeső szakaszai. Ez a szabály használható például olyan utcák esetében, ahol a szakaszok hurokba kerülhetnek.



Az egyenes vonal jellemzői átfedik magát, a korallvonal által jelzett hibával. Az ilyen jellegű hiba javítása az objektum egyszerűsítésével lehetséges.

4.1.1.3.6.1.2.13 Nem metszheti saját magát

A szabály megköveteli, hogy egy vonal ne metsze vagy fedje át saját magát. Ez a szabály használható kontúrvonalak, amelyek nem metszhetik magukat.



Az ilyen jellegű hiba javítása az objektum egyszerűsítésével lehetséges.

4.1.1.3.6.1.2.14 Nem lehet többemű

Szükséges, hogy a vonalnak csak egy része legyen. Ez a szabály akkor hasznos, ha a vonal jellemzői, például az autópályák, nem tartalmazhatnak több részt.

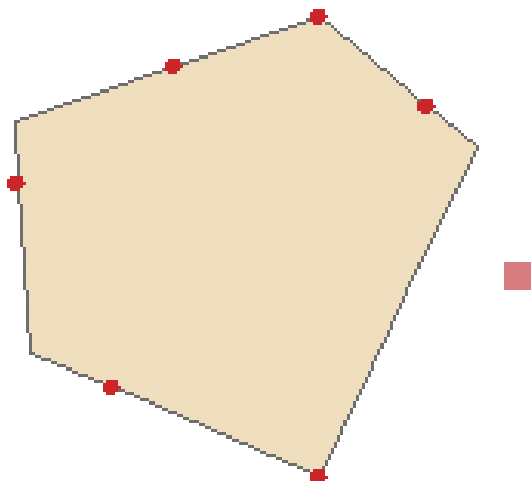


A több alkotóelemből álló vonalas objektumok egyetlen vázlatból készülnek. Az ilyen jellegű hiba javítása az objektum szétbontásával, "felrobbantásával" lehetséges.

4.1.1.3.6.1.3 Pontokra vonatkozó szabályok

4.1.1.3.6.1.3.1 Át kell fednie adott objektumtípussal

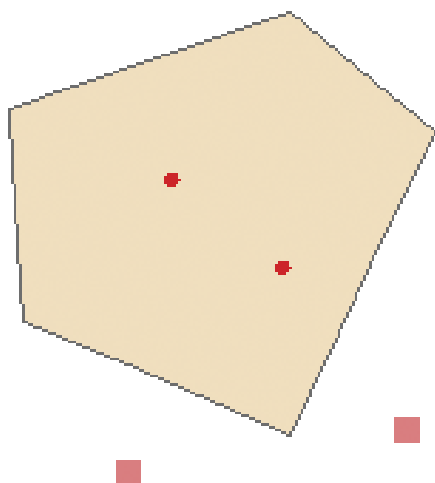
A szabály megköveteli, hogy a pontok a poligon határával egybeessenek. Ez akkor hasznos, ha a pont határjelzők, amelyeknek bizonyos területek szélein kell lenniük.



A jobb oldali négyzet hibát jelez, mert olyan pont, amely nem a poligon határán van.

4.1.1.3.6.1.3.2 A poligon belsejében kell lennie

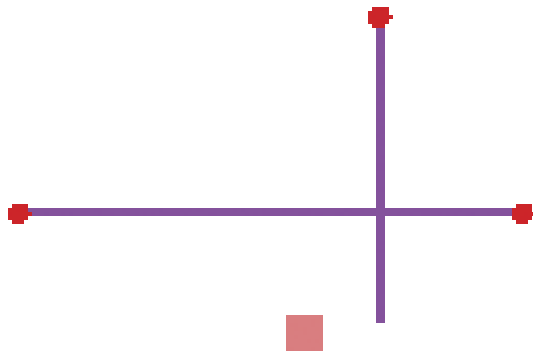
Megköveteli, hogy a pontok a terület jellemzői alá essenek. Ez akkor hasznos, ha a pont funkciók poligonokhoz kapcsolódnak, például kutak és kútbetétek vagy címpontok és parcellák.



A négyzetek olyan hibák, ahol olyan pontok vannak, amelyek nem a sokszög belsejében vannak. Az ilyen jellegű hiba javítása az objektum törlésével lehetséges.

4.1.1.3.6.1.3.3 Át kell fednie a végponttal

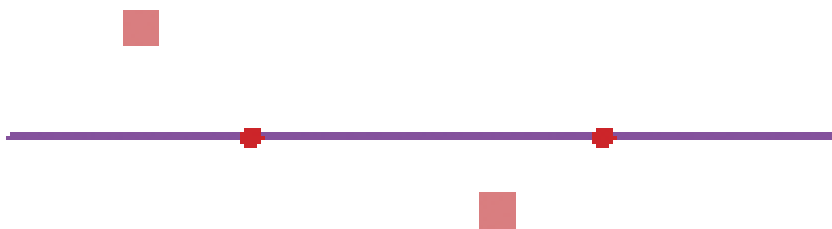
Szükséges, hogy az egyik funkcióosztályba tartozó pontokat le kell fednie a másik funkcióosztályba tartozó sorok végpontjainak. Ez a szabály hasonló a vonalszabályhoz, a "végpontnak le kell fednie", kivéve, hogy azokban az esetekben, ahol a szabály megsértése történt, a csúcstornak inkább a hibaként jelölt pontjellemzője. A határoló sarokjelzőket korlátozhatják a határvonalak végpontjai.



A négyzet olyan hibát jelez, ahol a pont nem esik egybe a vonal végpontjával. Az ilyen jellegű hiba javítása az objektum törlésével lehetséges.

4.1.1.3.6.1.3.4 Egybe kell esnie a vonallal

A szabály megköveteli, hogy az egyik objektumtípusban lévő pontok egy másik típusban lévő vonalakra illeszkedjenek. Ez a szabály olyan pontok esetében hasznos, amelyek egy vonal mentén fekszenek, például az autópályákon elhelyezett jelzőtáblák.



A négyzetek olyan pontok, amelyekre a vonal nem illeszkedik.

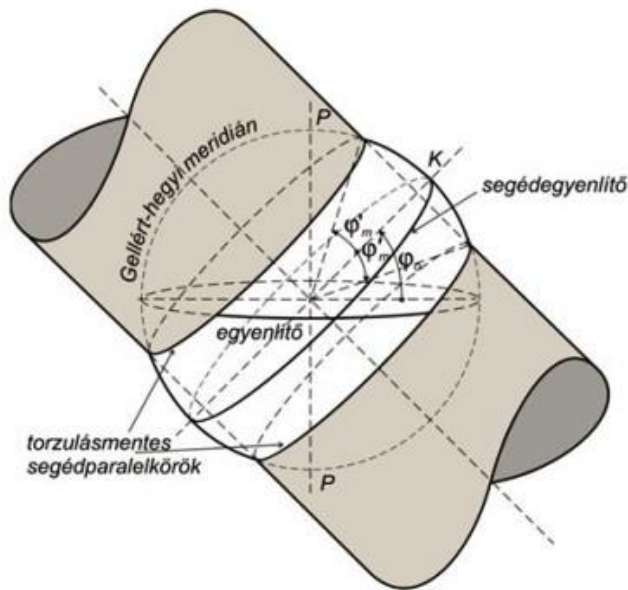
4.1.1.3.7 Térbeli referencia rendszerek alkalmazhatóság szempontjából

Egy pont helyzetének meghatározása azt jelenti, hogy megadjuk koordinátáit valamely alapul választott vonatkozási (koordináta) rendszerben. Magyarországon a két leggyakrabban alkalmazott koordináta rendszer a földrajzi elhelyezkedés meghatározására a GPS mérések koordináta rendszere a WGS84, valamint az EOV (Egységes Országos Vetületi) rendszer. Mindkét koordináta rendszerben egy adott pontot két koordináta x és y koordináta definiál, de a koordináta értékek különböző rendszerben épülnek fel. A szoftverek és eszközök döntő része a nemzetközi az elterjedt néven GPS koordináta rendszeren alapul.

A tárolt térképi adatok feldolgozásának egyszerűsítése és kompatibilitásának elérése szempontjából fontos egy olyan egyezményes vonatkoztatási rendszer kiválasztása, amely nem csak tematikus adatok térképi megjelenítésére, de akár térinformatikai célokra is alkalmazhatóvá teszi az tárolt téradatokat. A lehetséges alkalmazható vetületi rendszereket a továbbiakban tekintjük át.

4.1.1.3.7.1 EOV - Egységes Országos Vetület

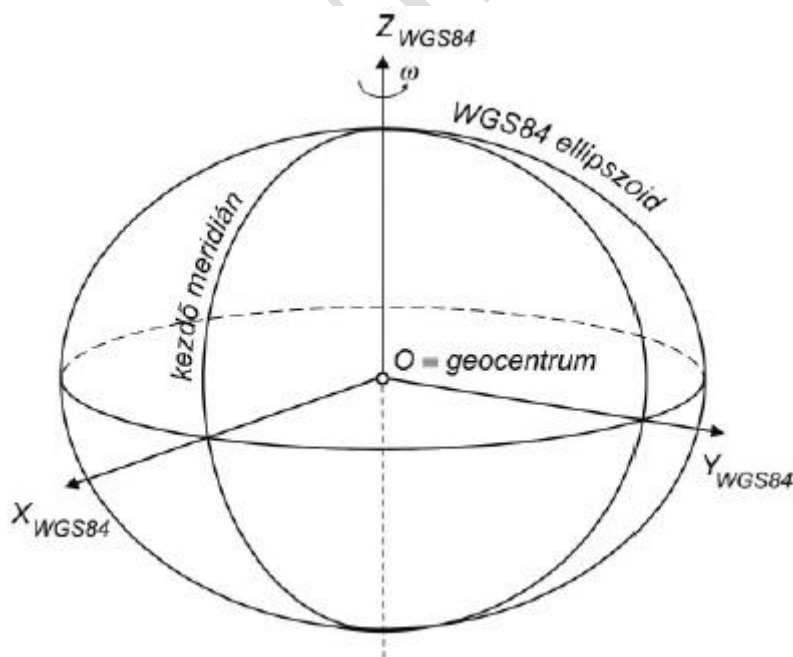
Az Egységes Országos Vetület (röviden: EOV) rendszere Magyarországon 1975-ben került bevezetésre, összhangban az egységes országos térképrendszerrel (EOTR-rel). Ferdetengelyű, szög tartó, ún. redukált hengervetület. A koordináta-tengelyek tájékozása ÉK-i, vagyis a pozitív X iránya északra, a pozitív Y iránya keletre mutat.



Segítségével az ország egész területét egyetlen vetületi síkon ábrázolhatjuk. A síkvetületi koordináták előjeles mennyiségek. A koordináta-rendszer kezdőpontját eltolták úgy, hogy a negatív előjelű koordinátákat kiküszöböljék. A koordináta-tengelyek önmagukkal párhuzamosan X irányban 200 000, Y irányban 650 000 méterrel vannak eltolva, így a koordináták értékéből könnyen eldönthető, hogy X vagy Y koordinátával van dolgunk. A számítások egyszerűsítése érdekében és az X koordináta 400 000 m-nél mindig kisebb, az Y koordináta 400 000 m-nél mindig nagyobb legyen, a koordináták felcserélése nem lehetséges.

4.1.1.3.7.2 WGS84 - vonatkoztatási rendszer

A GPS-műholdak által sugárzott fedélzeti pályaadatok vonatkoztatási rendszere WGS84 (World Geodetic System 1984) néven ismert, az Egyesült Államok Védelmi Minisztériumának Katonai Térképészeti Szolgálatára globális méretű katonai térképészeti célokra hozta létre. A Földet ellipszoid alakú sávokkal „fedi le”, így a magassági koordináták is kezelhetők vele. A szabvány neve: EPSG:4326.



Arra tervezték, hogy regionális és globális igényeket is kielégítsen. Például különböző nemzeti rendszerek közötti kapcsolat megteremtése. Egy geodéziai világrendszer (World Geodetic System),

amely geocentrikus elhelyezésű (kezdőpontja a föld tömegközéppontja), biztosít egy geometriai alapfelületet és modellezi a Föld nehézségi erőterét.

A WGS84 adja a GPS műholdas helymeghatározó rendszer vonatkoztatási rendszere is.

4.1.1.3.7.3 Web Mercator

Google Mercator, vagy más néven spherical mercator a Google Maps, Microsoft Virtual Earth, a Yahoo Maps, Open Street Map és más kereskedelmi API szolgáltatók által használt vetület. A vetület a földet gömbként és nem elipszoidként kezeli. Az egysége nem fok, hanem méter.

A kis méretarányokban jelentkező nagyfokú torzításai ellenére a vetület jól használható interaktív, tetszőlegesen nagyítható világtérképekhez, mivel belenagyítva egy kisebb területeken relatíve kicsi lesznek a torzulások annak köszönhetően, hogy ez a vetületváltozat is közel szögtartó. A Web Mercator vetületnél egy, a WGS84 ellipszoid fél nagytengelyével egyenlő sugarú gömbről vetítünk a (szögtartó) gömbi Mercator-vetület segítségével a síkba.

A Google Maps az úgynevezett Mercator “vetítés” egyik változatát alkalmazza, amelyen az északi és a déli pólusok egy pont helyett egy vonalat alkotnak. A szabvány neve: EPSG:3857.

4.1.1.3.7.4 Konklúzió

A fentiek alapján a WGS84 vetületi rendszer kiválasztása indokolt a téradatak közös tárolásának vetületi rendszereként, mivel széles körben elterjedt és általánosságban elmondható, hogy a lokálisan alkalmazott vetületi rendszerekből való szabatos áttérés megoldott. Így akár - a betöltött adatok pontosságának függvényében - térinformatikai elemzésekre is alkalmas lehet a létrehozott adatrendszer.

4.2 Gyakorlati következmények

4.2.1.1 A térérzékeny keresés gyakorlati lehetőségei a névterekben

A térérzékeny keresést a felhasználó igényeihez kell igazítani. A felhasználó megadhatja a szűrési szempontjait akkor, amikor beírja saját keresőkifejezését, tehát az első keresés során, de fel lehet kínálni számára azt is, hogy az első keresés után szűkíthessen a visszaadott találati listába foglalt adatok alapján. Az első keresés során fel kell kínálni azt a lehetőséget, hogy a felhasználó megadhassa azt, hogy valamilyen földrajzi térszegmensen belüli találatokat adjon vissza a keresőrendszer. Ennek többféle technikája lehetséges.

- adott földrajzi ponthoz való közelség megadása távolság megjelölésével
- adott földrajzi térszegmens megadása befoglaló téglalap segítségével
- adott földrajzi térszegmens megadása objektum segítségével (pl. egy ország területén)
- bizonyos topológiai feltételek teljesülésével (pl. a Duna mentén fekvő települések)

A térérzékeny keresés további lehetőségeként fel lehet kínálni azt, hogy a felhasználó finomítson a rendszer által elsőre visszaadott találati listán. Ehhez azt kell biztosítani, hogy a felhasználó a találati listában mindig feltüntetett partitív tartalmazási reláció segítségével egyetlen kattintás segítségével megadhassa azt, hogy az általa kiválasztott keresőparaméter alapján a keresőrendszer adott földrajzi térszegmensre szűkítve adja vissza a következő találati listát. Ezt azért lehet biztosítani, mert a névtérben olyan földrajzi helyek kerülhetnek be, amelyeknek – valamilyen pontossággal – mindig ismerjük a földrajzi térben való elhelyezkedését. Ez alapján minden földrajzi helyre fel lehet építeni egy olyan útvonalat, amely azt mutatja, hogy a legnagyobb földrajzi egység (jelen esetben a ‘Föld’) felől elindulva, milyen sorozatos szűkítésekkel lehet eljutni az adott földrajzi entitáshoz. Ezt a partitív alárendeltje tartalmazási reláció iteratív egymásba ágyazásával lehet elvégezni.

Mivel a földrajzi névtér találati listájában mindig meg kell adni az adott földrajzi hely partitív fölötteseit, az így felkínált adatokat lehet térbeli szűrésre is használni. Vegyük a következő példát.

Bánfi-szivattyútelep ⇒ Föld | Európa | Magyarország | Hajdú-Bihar megye | Püspökladányi járás
| Püspökladány | Bánfi-szivattyútelep

A partitív fölrendelt földrajzi helyek egyetlen útvonalba szervezett listájának mindegyik elemén link van, amelynek elsődleges funkciója az, hogy egyetlen kattintással a kiválasztott földrajzi hely oldalára lehessen ugrani. A teljes útvonal feltüntetése arra is alkalmas, hogy még ha a felhasználó nem is akarna

valamelyik partitív fölöttes földrajzi hely oldalára ugrani, akkor is láthatja, hogy a találati lista adott eleme milyen “útvonalon” érhető el (mindig a Föld mint mindent magába foglaló földrajzi hely felől elindulva). Ezek az útvonal-adatok sokszor elégségesek ahhoz, hogy segítségükkel a felhasználó dönthessen arról, melyik elemet válassza ki a találati listából.

Sokszor előfordulhat olyan eset, amikor a találati listában nagyon sok azonos nevű földrajzi helyet kap vissza a felhasználó.

Keresés → őreghegy

1800 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010

| | |
|--|-------|
| Alsóőreghegy → Föld Európa Magyarország Tolna megye Paksi járás Dunaföldvár Dunaföldvár Alsóőreghegy Mgl. | 14107 |
| Alsóőreghegy vasúti megállóhely → Föld Európa Magyarország Pest megye Gödöllői járás Gödöllő Gödöllő Alsóőreghegy vasúti megállóhely megszűnt vasúti megálló | 89476 |
| Felsőőreghegy → Föld Európa Magyarország Tolna megye Paksi járás Dunaföldvár Dunaföldvár Felsőőreghegy külsőterület | 38460 |
| Kültelek (Őreghegy) Őreghegy → Föld Európa Magyarország Baranya megye Pécsi járás Abaliget Abaliget Kültelek (Őreghegy) Lh. | 3539 |
| Őreghegy → Föld Európa Magyarország Somogy megye Siófoki járás Balatonendréd Balatonendréd Őreghegy Mgl. | 3970 |
| Őreghegy → Föld Európa Magyarország Baranya megye Pécsi járás Berkesd Berkesd Őreghegy Lh. | 4609 |
| Őreghegy → Föld Európa Magyarország Baranya megye Komlói járás Komló Komló Őreghegy Üd. | 4927 |
| Őreghegy → Föld Európa Magyarország Baranya megye Pécsi járás Orfű Orfű Őreghegy Üd. | 5151 |
| Őreghegy → Föld Európa Magyarország Baranya megye Hegyháti járás Tékes Tékes Őreghegy Mgl. | 5403 |
| Őreghegy → Föld Európa Magyarország Fejér megye Enyingi járás Enying Enying Őreghegy Mgl. | 7694 |


1-10 11-20 21-30 31-40 41-50 51-60 61-70 71-80 →

A fenti képernyőkép azt mutatja, hogy az ‘őreghegy’ keresésre a kereső – a földrajzi névtér jelenlegi adatai közül – 80 találatot ad vissza. Ebben a halmazban csak a “felettes útvonalakban” látható adatok segíthetik a felhasználót a választásban. Az igaz, hogy jobb híján végig nézhető a találati lista minden eleme, de ezen a ponton lehet felkínálni azt a lehetőséget, hogy a felhasználó térbeli szűrést végezhesen el, amivel lecsökkentheti a következő találati lista elemszámát.

A partitív fölöttesekből álló útvonal elemei arra is alkalmasak lehetnek, hogy szűrési feltételként szolgáljanak. Egyszerűen megvalósítható ugyanis az, hogy bármelyik útvonal-elemre kattintva jelezze a felhasználó, hogy a kiválasztott elem földrajzi határain belülre akarja szorítani a következő keresését (természetesen úgy, hogy a rendszer az általa megadott keresőkifejezésre keressen, csak most már a leszűkített földrajzi tartományon belül).

Másfajta módon lehet alkalmazni a térérzékeny keresés technikáját a személynévtéren vagy a testületi névtéren belül. Ezekben a névtérekben a névtérben feldolgozott entitásoknak nincs olyan direkt térbeli kötődése, mint a földrajzi névtér elemei esetében, de néhány információt lehet arra használni, hogy a keresések során szűrni tudjanak a felhasználók az információk tartományban.

Noha a személyek vagy a testületek nincsenek a térhez kötve, ezért nem tudjuk egyértelműen a térben lokalizálni őket, azért léteznek olyan eseményeket az esetükben is, amikor lehet a térhez kötésükről beszélni. A személyek nevezetes életadatai közül meg szokás adni a születés és a halál helyét (és persze időpontját), és ezek az adatok alkalmasak arra, hogy a térbeli szűrésekhez felhasználjuk őket. Vegyük azt a példát, amikor ‘Szabó István’-ra keresünk.

Keresés → szabó istván 

1800 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010

bármilyen foglalkozás személynév:

| | |
|--|-------|
| Szabó István (Debrecen, 1896 - , 1974) országgyűlési képviselő | 12879 |
| Szabó István (Budapest, 1946 - ,) országgyűlési képviselő | 12880 |
| Szabó István (Endrőd, 1896 - , 1958) főispán | 13046 |
| Szabó István (Mátészalka, 1887 - Mátészalka, 1978) országgyűlési képviselő | 14698 |
| Szabó István (Cserseztomaj, 1931 - Budapest, 1976) műfordító, író | 18606 |
| Szabó István (Botpalád, 1920 - Pécs, 1957) földműves, rokkantnyugdíjas | 20791 |
| Szabó István (Pécs, 1916 - Budapest, 1989) tüdőgyógyász, orvos | 26351 |
| Szabó István (Isztimér, 1891 - Budapest, 1957) vegyészmérnök, borász | 29686 |
| Szabó István (Budisava, 1910 - Szabadka, 1992) színész | 32591 |
| Szabó István (Budapest, 1933 - ,) színész | 32592 |

← 1-10 11-20 21-30 31-37 →


A megadott feltétel alapján a személynévtér jelenlegi adatbázisában 37 találatot kapunk.

A listában szereplő személyek születési helyét használhatjuk arra, hogy a felhasználó arra szűrhesse, hogy milyen földrajzi térben született személyeket szeretne látni. Mivel ismerjük a személyek születési helyét, és azokhoz mindig hozzá tudjuk rendelni a partitív fölérendelt földrajzi helyekből álló teljes útvonalat, ezért ez utóbbi segítségével ugyanazt a szűrést végre lehet hajtani, mint amit korábban már felvázoltunk. Ennek megvalósításához a keresőfelületen fel kell ajánlani a kiválasztható releváns felettes helyek listáját, amelyből a felhasználó egy kattintással elküldheti a rendszer számára a szűrési információt.

Ezen a felületen a szülőváros kiválasztása mellett egy térképen egy adott nagyobb terület szerkesztői eszközökkel (pl. egy befoglaló téglalap megrajzolásával), történő lehatárolására is lehetőséget lehet biztosítani. Lekérdezés segítségével összetettebb geometriai feltételeket is definiálhatunk (pl. országhatár közelében született személyek szűrése).

4.2.1.2 Időérzékeny keresés

Az időérzékeny keresés során az időbeli közelséget lehet kihasználni. Vegyük a következő példát. Ha a személynévtérben adott névre keresünk (mondjuk 'Nagy István'-ra).

Keresés → nagy istván 

1800 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010

bármilyen foglalkozás személynév: nagy istván ok

| | |
|---|-------|
| Nagy István (Csikmindszent, 1873 - Baja, 1937) festő, festőművész | 21289 |
| Nagy István (Búcs, 1905 - Villány, 1974) református lelkész, amatőr botanikus | 21403 |
| Nagy István (Maroslekenye, 1909 - Budapest, 1976) színész | 21558 |
| Nagy István (Mezőkövesd, 1900 - Mezőkövesd, 1984) matyó szücsbímző | 28262 |
| Nagy István (Kiskunlacháza, 1829 - Budapest, 1896) újságíró | 30818 |
| Nagy István (Kolozsvár, 1914 - ,) színész | 31793 |
| Nagy István (Csíkszereda, 1907 - , 1983) igazgató, karmester, karnagy | 31794 |
| Nagy István (, 1750 - , 1860) színész | 31795 |
| Nagy István (Petrosény, 1959 - ,) színész | 31796 |
| Nagy István (Magyarittabé, 1958 - ,) néprajzkutató | 36694 |

← 1-10 11-20 21-30 31-40 →

A képernyőképen látható, hogy ebben az esetben a névtér jelenlegi adatbázisából 40 elemű találati listát kapunk. Ahogy növekszik a személynévtérben rögzített személyek száma, nyilvánvalóan úgy lesz egyre hosszabb ez a találati lista.

Az ilyen esetekben praktikus szűrési lehetőséget kínálhatunk az időérzékeny szűrés segítségével. A személyek életrajzi adatai alapján ugyanis mindig el tudjuk helyezni őket az időben, ami megadja annak lehetőségét, hogy úgy szűrjük a találati lista elemeit, hogy csak a felhasználó által megadott időszávon belülre került személyeket adja vissza a keresőrendszer.

Ahhoz, hogy a felhasználó meg tudja adni azt az időszávet, amire szűrni akar, szükség van valamilyen adatbeviteli lehetőségre. Erre kiválóan alkalmas az az idővonal-sáv, amely a minden névtér tetején látható. Ezen az idővonalon lehet jelezni az adott oldalon megjelenített adatok közül azokat, amelyek eseményszerűek, tehát időbeli vonatkozásuk van. Ezen az idővonalon egy csúszkát elhelyezve a felhasználó könnyen beállíthatja azt, hogy milyen időszakra akarja szűkíteni a keresés szköpját.

A fent bemutatott időbeli szűkítést lehetővé tevő keresést minden olyan adattípus esetén alkalmazni lehet, ahol az adattípusba tartozó példányok esetén létezik időfüggés. Az ágensek (személyek vagy testületek) vagy a földrajzi helyek esetében ilyen lehet a megnevezési eseményekhez vagy az individuumok típusba sorolásához tartozó időadat, de a geoidividuumok geometriájának időbeli változásai is "felkínálják" az időbeli szűrés lehetőségét. Ha dokumentumokat (könyveket, cikkeket, filmeket, zeneműveket stb.) kezelünk, akkor ugyanúgy kihasználhatjuk azt, hogy a különböző jellemzőkben (létrehozási, megjelenési, forgalmazási stb. eseményekben) van időfüggés, ami időszűrésre használható.

5 Hivatkozások

- Abel, F., The Benefit of Additional Semantics in Folksonomy Systems. in: *Proceedings of the Second Ph.D. Workshop in Seventeenth ACM Conference on Information and Knowledge Management, CIKM 2008*, Napa, USA, October 26-30, 2008., pp. 49-56.
- Abel, Fabian, Nicola Henze, Daniel Krause, and Matthias Kriesell, On the Effect of Group Structures on Ranking Strategies in Folksonomies, in: *Workshop on Social Web Search and Mining at 17th International World Wide Web Conference (WWW 2008)*, 2008
- Adomavicius, G.; Tuzhilin, A. (June 2005), "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", in: *IEEE Transactions on Knowledge and Data Engineering* 17 (6), pp.734-749.
- Alexiev, Vladimir, Isaac, Antoine, Lindenthal, Jutta (2016), On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI). *Int J Digit Libr.* 17:39–48
- Amittai Axelrod, Metacarta: On building a high performance gazetteer database. URL citeseer.ist.psu.edu/639680.html
- Ankerl Géza, *Építészeti és kommunikáció*, Budapest: Műszaki Könyvkiadó, 1991
- ArcGIS (2003). *Working With Geodatabase Topology*. An ESRI® White Paper, May 2003
- Berners-Lee, Tim, James Hendler, and Ora Lassila, The Semantic Web, in: *Scientific American*, May 2001, at: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>
- Bittner, T., Donnelly, M., Smith, B. (2004). Individuals, universals, collections: On the foundational relations of ontology. In Achille C. Varzi, Laure Vieu (eds.): *Formal Ontology in Information Systems. Proceedings of the Third International Conference*. Amsterdam, Berlin, Oxford, Tokyo, Washington DC, IOS Press, 37-48. p.
- Bittner, T., Donnelly, M., Smith, B. (2009). A Spatio-Temporal Ontology for Geographic Information Integration. *International Journal of Geographical Information Science*. 23(6), pp.765-798.
- Bittner, Thomas – Smith, Barry. Granular partitions and vagueness. In *FOIS '01: Proceedings of the international conference on Formal Ontology in Information Systems*. New York, NY, USA, 2001, ACM Press, 309–320. p. URL citeseer.ist.psu.edu/bittner01granular.html .
- Bittner, T., Smith, B. *Granular spatio-temporal ontologies*, 2003. URL citeseer.ist.psu.edu/bittner03granular.html .
- Bittner, Thomas – Donnelly, Maureen – Smith, Barry. Individuals, universals, collections: On the foundational relations of ontology. In Achille C. Varzi – Laure Vieu (szerk.): *Formal Ontology in Information Systems. Proceedings of the Third International Conference (FOIS*

- 2004). Amsterdam, Berlin, Oxford, Tokyo, Washington DC, 2004, IOS Press, 37–48. p.
- Bittner, Thomas – Smith, Barry – Donnelly, Maureen. *Systems of granular partitions*. URL <http://citeseer.ist.psu.edu/612893.html> .
 - Bittner, Thomas – Smith, Barry. A theory of granular partitions. In Matthew Duckham – Michael F. Goodchild – Michael F. Worboys (szerk.): *Foundations of Geographic Information Science* . London, 2003, Taylor and Francis, 117–151. p. URL <http://wings.buffalo.edu/philosophy/faculty/smith/articles/partitions.%20pdf>>.
 - Bittner, Thomas –Stell, John G. Approximate qualitative spatial reasoning. 2. évf. (2001) 4. sz., *Spatial Cognition and Computation* , 435–466. p. ISSN 1387-5868.
 - Bittner, Thomas –Stell, John G. Vagueness and rough location. 6. évf. (2002) 2. sz., *Geoinformatica* , 99–121. p.
 - Bittner, Thomas. Approximate qualitative temporal reasoning. 36. évf. (2002) 1-2. sz., *Annals of Mathematics and Artificial Intelligence* , 39–80. p. ISSN 1012-2443.
 - Borgo, Stefano – Masolo, Claudio. *Mereogeometries: a semantic comparison*, 2001, LADESB-CNR. URL citeseer.ist.psu.edu/borgo01mereogeometries.html .
 - Breese, J.S., Heckerman, D. and Kadie, C.M. Empirical analysis of predictive algorithms for collaborative filtering. Microsoft Research Technical Report, (MSR-TR-98-12), October 1998.
 - Broder, A. A taxonomy of web search. SIGIR Forum, 36, 2 (2002), 3-10.
 - Brookfield, Karen. *Az írás*. Budapest: Park Kiadó, 1995.
 - Casati, Roberto – Varzi, Achille C. Spatial entities. In Oliviero Stock (szerk.): *Spatial and Temporal Reasoning*. Dordrecht, 1997, Kluwer Academic Publishers, 73–96. p. ISBN 0792346440 (paper).
 - Casati, Roberto, Varzi, , Achille C. (eds.). *Events*. Ashgate, 1996
 - Chalmers, Matthew, Andreas Dieberger, Kristina Höök, Åsa Rudström, Social Navigation and Seamless Design, *Cognitive Studies*, 11(3), Sept. 2004, pp. 1-11.
 - Clarke, Stella G Dextre, Smedt, Johan De (2011). *ISO 25964 ISO 25964-1: a new standard for 1: a new standard for development of thesauri and exchange of thesaurus data*, at: https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2011/presentations/DextreClarke_DeSmedt_ISO25964.pdf
 - Clarke, Stella G. Dextre, Smedt, Johan De (2011). *ISO 25964 ISO 25964-1: a new standard for 1: a new standard for development of thesauri and exchange of thesaurus data*, at: https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2011/presentations/DextreClarke_DeSmedt_ISO25964.pdf
 - Cohn, A G. – Varzi, A. Mereotopological connection. 32. évf. (2003), *Journal of Philosophical Logic* , 357–390. p.

- Cohn, A. G. –Varzi, A. Connection relations in mereotopology. In H. Prade (szerk.): *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI-98)*. 1998. August, Wiley, 150–154. p.
- Corcho, O., Fernández-López, M., Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point?, in: *Data & Knowledge Engineering*, Vol. 46, 2003, pp.41-64.
- Cruse, D.A. (1986). *Lexical Semantics*, Cambridge University Press
- Csigó Péter, A konvergens televíziózás - Web, tv, közösség, L'Harmattan, 2009
- Dancs Szabolcs (2017). Bemutatkozik az ISNI, a nemzetközi szabványos névazonosító. *TMT*, vol. 64, No. 4, 199-2007.o.
- Darwen, What a Database Really Is: Predicates and Propositions. In: Date, C. J., Darwen, Hugh, McGoveran, David. *Relational Database Writings 1994-1997*. Reading, Mass.: Addison-Wesley, 1998
- Date, C. J., Darwen, Hugh, McGoveran, David. *Relational Database Writings 1994-1997*. Reading, Mass.: Addison-Wesley, 1998.
- Date, Chris. *An Introduction to Database Systems*. Addison-Wesley, 2000
- Davidson, D. (1967), 'The Logical Form of Action Sentences', in N. Rescher (ed.), *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press, pp. 81–95; reprinted in *Events*, pp. 3–17, and in Davidson 1980, pp. 105–122.
- Davidson, D. (1969). 'The Individuation of Events', in N. Rescher (ed.), *Essays in Honor of Carl G. Hempel*, Dordrecht: Reidel, pp. 216–34; reprinted in *Events*, pp. 265–283, and in Davidson 1980, pp. 163–180.
- Dieberger, A. (1999). Social connotations of space in the design for virtual communities and social navigation. In A. Munro, K. Höök and D. Benyon (Eds.), *Social Navigation of Information Space*, pages 35-52. London: Springer (<http://tecfa.unige.ch/~nova/rss4you.html>)
- Dieberger, A., Supporting Social Navigation on the World-Wide Web. *International Journal of Human-Computer Studies*, 46(6), pp. 805-826.
- Dobrowolski, Z. K.. Osztályozási rendszerek szerkesztése. In: Ungváry Rudolf, Orbán Éva (szerk.). *Osztályozás és információkeresés, I-II*. Budapest: OSZK, 2001.
- Doctorow, C. (2001). *Metacrap: Putting the torch to seven straw-men of the meta-utopia*, at: 2001.08.26. at: <http://www.well.com/~doctorow/metacrap.htm>
- Donnelly, Maureen – Smith, Barry. Layers: A new approach to locating objects in space. In Werner Kuhn – Michael F. Worboys – Sabine Timpf (szerk.): *Spatial Information Theory. Foundations of Geographic Information Science, International Conference, COSIT 2003*, Springer, 2003, 46–60. p.

- Dourish, P. and Chalmers, M. (1994). Running out of Space: Models of Information Navigation, *Proceedings of the Conference Human Computer Interaction'94*
- Dourish, P., Adler, A., Belloti, V., Henderson, A. Your Place or Mine? Learning from Long-Term Use of Audio-Video Communication. *Computer-Supported Cooperative Work*, 5(1), 33-62, 1996.
- Egenhofer, M. J. – Mark, D. M. Naive geography. In A. U. Frank –W. Kuhn (szerk.): *Spatial Information Theory - A Theoretical Basis for GIS (COSIT'95)*. Berlin, Heidelberg, 1995, Springer, 1–15. p., URL <http://citeseer.ist.psu.edu/egenhofer95naive.html>
- Engeström, Jyri. Why some social network services work and others don't — Or: the case for object-centered sociality, 2005.04.13. at: http://zengstrom.com/blog/2005/04/why_some_social.html
- Feilmayr, Ch., Wöß, W. (2016). An analysis of ontologies and their success factors for application to business. *Data & Knowledge Engineering*, Vol. 101, pp.1-23.
- Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. (1987). The vocabulary problem in human-system communication. in: *Commun. ACM* 30, 11
- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. (2001). Understanding top-level ontological distinctions. In *Proc. of IJCAI 2001 workshop on Ontologies and Information Sharing*.
- Goldberg, David, David Nichols, Brian M. Oki, Douglas Terry, Using collaborative filtering to weave an information tapestry, in: *Communications of the ACM*, 1992, Vol.35, No.12, pp.61-70.
- Golder, Scott A., Bernardo A. Huberman, Usage Patterns of Collaborative Tagging Systems, in: *Journal of Information Science*, 2006, 32(2). 198-208. at: <http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf>
- Gómez-Pérez, A., Fernández-López, M., Corcho, O. (2004). *Ontological Engineering*, Springer
- Goody, Jack – Watt, Ian. Az írásbeliség következményei. In: Nyíri Kristóf – Szécsi Gábor (szerk.): *Szóbeliség és írásbeliség*. Budapest, Áron Kiadó, 1998, 111–128. o.
- Goody, Jack. Nyelv és írás. In Nyíri Kristóf – Szécsi Gábor (szerk.): *Szóbeliség és írásbeliség*. Budapest, Áron Kiadó, 1998, 189–221.o.
- Goody, Jack. *The Domestication of the Savage Mind*. Cambridge, Cambridge University Press, 1977
- Green, Rebecca, Bean, C.A., Myaeng, Hyon, S. (2001). *The Semantics of Relationships: An Interdisciplinary Perspective*, Dordrecht: Kluwer
- Grenon, P., Smith, B. (2004). SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition and Computation*, 4(1), pp.69-103.
- Griffiths, T. A. – Fernandes, A. AA. – Paton, N. W. – Mason, K. – Huang, B. – Worboys, M. – Johnson C. Tripod: A comprehensive system for the management of spatial and aspatial historical objects. In *Proceedings Ninth ACM*

- Hill, Linda L. Core elements of digital gazetteers: Placenames, categories, and footprints. In *ECDL '00: Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries*. London, UK, 2000, Springer-Verlag, 280–290. p. URL citeseer.ist.psu.edu/hill00core.html.
- Hobbs, Jerry R. *Granularity*, 2002. 15 August.
- Hubay Miklós Péter (2016). *A BIBFRAME és a könyvtári feldolgozás új keretei*, at: <http://mek.oszk.hu/15600/15678/15678.pdf>
- Huszák Péter, Szabó Mihály, *Közösségi archívumok*, in: *Magyar Távközlés*, 2006/2.
- Kripke, Saul. *Megnevezés és szükségszerűség*. Budapest: Akadémiai Kiadó, 2010
- Kroski, Ellyssa. The Hive Mind: Folksonomies and User-Based Tagging, 2006.03.07, at: <http://infotangle.blogspot.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/>
- Kuhn, Werner – Worboys, Michael F. – Timpf, Sabine (szerk.). *Spatial Information Theory. Foundations of Geographic Information Science, International Conference, COSIT 2003*, Proceedings, Lecture Notes in Computer Science, 2825. köt. Springer, 2003.
- Kuti Judit, Varasdi Károly, Cziczelszki Judit, Gyarmati Ágnes, Nagy Anikó, Tóth Marianna, Vajda Péter (2006). Igei wordnet és igei eseményszerkezet ábrázolása, *IV. Magyar Számítógépes Nyelvészet Konferencia*, Szeged, 2006, 97–108. at: http://www.nytud.hu/oszt/korpusz/resources/kuti_et_al_WN_2006.pdf
- Lappalainen, Mikko, Frosterus, Matias, Nykyri, Susanna (2014). Reuse of library thesaurus data as ontologies for the public sector. In *IFLA WLIC 2014*, 16-22.p.
- Linden, G., Smith, B., and York, J. (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing*, Jan-Feb 2003.
- Lowe, E. J. (1998). *The possibility of metaphysics*. Oxford: Clarendon Press.
- Lyons, J. (1977). *Semantics*, Cambridge University Press
- Lyons, J. *Semantics I-II*, New York: Cambridge University Press, 1977
- Maddux, Roger D. Relation algebras for reasoning about time and space. In M. Nivat – C Rattray – T. Rus – G. Scollo (szerk.): *Algebraic Methodology and Software Technology (AMAST '93)*. New York, N.Y., 1993, Springer-Verlag, 27–44. p. URL citeseer.ist.psu.edu/maddux93relation.html.
- Magyar Egységes Ontológia (MEO) projekt dokumentumai, at: <http://ontologia.hu>
- Manov, Dimitar – Kiryakov, Atanas – Popov, Borislav – Bontcheva, Kalina – Maynard, Diana – Cunningham, Hamish. *Experiments with geographic knowledge for information extraction*. URL <http://citeseer.ist.psu.edu/688096.html>.
- Manovich, Lev. *The Language of New Media*. 2001

- Marlow, Cameron , Naaman, Mor , Boyd, Danah and Davis, Marc. (2006). Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 31–40, New York, ACM Press
- Marlow, Cameron, Mor Naaman, Danah Boyd, and Marc Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pp 31–40, New York, NY, USA, 2006. ACM Press.
- Martin Szomszor, Ciro Cattuto, Harith Alani, Kieron O'Hara, Andrea Baldassarri, Vittorio Loreto, Vito D.P. Servidio, Folksonomies, the Semantic Web, and Movie Recommendation, In: 4th European Semantic Web Conference, Bridging the Gap between Semantic Web and Web 2.0, 3-7th, June 2007, Innsbruck, Austria. at: <http://eprints.ecs.soton.ac.uk/14007/1/ESWC2007.pdf>
- Mathes, Adam (2004). *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*, UIC Technical Report, at: <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Mathes, Adam. *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*, UIC Technical Report, 2004, at: <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Merholz, P. Clay *Shirky's Viewpoints are Overrated*, 2005.8.7, at: <http://www.peterme.com/archives/000558.html>
- Mika, Peter. Ontologies Are Us: A Unified Model of Social Networks and Semantics, in: Y. Gil et al. (eds.) *ISWC 2005*, LNCS, 2005, pp.522-536.
- *Mobile Web 2009 = Desktop Web 1998*, Jakob Nielsen's Alertbox, February 17, 2009, web: <http://www.useit.com/alertbox/mobile-usability.html>
- O'Brien, Jeffrey M. The race to create a 'smart' Google, in: *CNN Money*, 2006.11.20, at: http://money.cnn.com/magazines/fortune/fortune_archive/2006/11/27/8394347/
- O'Reilly, Tim. *What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software*, <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- Ong, Walter. *Szóbeliség és írásbeliség*. Budapest, AKTI–Gondolat Kiadó, 2010
- Rich, E. "User Modeling via Stereotypes", in: *Cognitive Science*, 1979, vol. 3, no. 4, pp. 329-354.
- Chapman, Robert L. (ed.). *Roget's International Thesaurus*, New York: HarperCollins Publishers, 1992.
- Ryle, G. (1949), *The Concept of Mind*, London: Hutchinson.
- Saussure, Ferdinand de. *Bevezetés az általános nyelvészetbe*. Budapest, Corvina Kiadó, 1998

- Schaefer, Mark W. *The Content Code: Six essential strategies to ignite your content, your marketing, and your business.* Mark W.\Schaefer. 2014
- Scott A. Golder, Bernardo A. Huberman, Usage Patterns of Collaborative Tagging Systems, in: *Journal of Information Science*, 2006, 32(2). 198-208. at: <http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf>
- Searle, J.R. (1995). *The Construction of Social Reality.* The Free Press
- Shirky, Clay. *Ontology is Overrated: Categories, Links, and Tags*, 2005 at: shirky.com/writings/ontology_overrated.html
- Simons, Peter (2000). *Parts: A Study in Ontology.* Oxford University Press
- Smith, b. –Mark, D. M. Do mountains exist? towards an ontology of landforms. 30. évf. (2003) 3. sz., *Environment & Planning B: Planning and Design*, 411–427. p.
- Smith, B. Boundaries: An essay in mereotopology. In L. Hahn (szerk.): *The Philosophy of R. Chisholm.* 1997, Open Court, 534–561. p.
- Smith, B. On drawing lines on a map. In A. U. Frank –W. Kuhn (szerk.): *Spatial Information Theory - A Theoretical Basis for GIS (COSIT'95)*. Berlin, Heidelberg, 1995, Springer, 475–484. p.
- Smith, B., Varzi, A. (2000), Fiat and Bona Fide Boundaries. *Philosophy and Phenomenological Research*, 60, pp. 401–420.
- Smith, Barry –Varzi, Achille C. Fiat and bona fide boundaries: Towards on ontology of spatially extended objects. In *COSIT '97: Proceedings of the International Conference on Spatial Information Theory.* London, UK, 1997, Springer-Verlag, 103–119. p.
- Smith, Barry –Varzi, Achille C. Surrounding space: The ontology of organism environment relations. 121. évf. (2002) 2. sz., *Theory in Biosciences*, 139–162. p. URL http://ontology.buffalo.edu/smith/articles/Surrounding_space.pdf.
- Smith, Gene. *Folksonomy: social classification*, 2004 .08.03., at: http://atomiq.org/archives/2004/08/folksonomy_social_classification.html
- Sowa, J.F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations.* Brooks Cole Publishing Co., 2000.
- Speroni, Pietro (2006). *On Tag Clouds, Metric, Tag Sets and Power Laws*, at: <http://blog.pietrosperoni.it/2005/05/25/tag-clouds-metric/>
- Speroni, Pietro. *On Tag Clouds, Metric, Tag Sets and Power Laws*, 2006.03.07, at: <http://blog.pietrosperoni.it/2005/05/25/tag-clouds-metric/>
- Šrejder, Ju. A. *Egyenlőség, hasonlóság, rendezés.* Budapest: Gondolat, 1975

- Šrejder, Ju. A. Rendszerek és modellek. In: Ungváry Rudolf, Orbán Éva (szerk.) *Osztályozás és információkeresés*. Budapest: OSZK, 2001, II.kötet, 297–325.o.
- Staab, S. Studer, R. (2009). *Handbook of Ontologies*, Springer Verlag
- Star, Susan Leigh. *Slouching toward Infrastructure*, 1996, at: <http://is.gseis.ucla.edu/research/dl/star.html>
- Strawson, P. F. (1959). *Individuals. An Essay in Descriptive Metaphysics*. London and New York, Routledge.
- Strawson, P. F. (1959). *Individuals. An Essay in Descriptive Metaphysics*. London and New York: Routledge
- Syi (2014). *syi.hu/cse*. L'Harmattan – Könyvpont Kiadó
- Syi, Egyben az egész. Egytől egyig, Typotex, 2007
- Syi, Interakció. Rab Árpád (szerk.) *Csomópontok*. Gondolat, Infonia, 2017, 13-32.o,
- Syi, Markovich Réka, Hamp Gábor, Goody-listák jogszabálysövegekben. Három tételben. *Jel-Kép*, (3), 13–24. o.
- Syi, Olvasni az oszlopok közt. Rab Árpád (szerk.) *Csomópontok*. Gondolat, Infonia, 2017, 129-153.o.
- Syi, Navigáció. In: Kangyal András, Laufer László (szerk.) *Gépézet: Interfész, interakció, navigáció*. Budapest: L'Harmattan Kiadó, 2011. pp. 196-294.
- Szabó Dóra. *Felhasználói segédlet a Web of Science adatbázis és a ResearcherID használatához*. Szent István Egyetem Kosáry Domokos Könyvtár és Levéltár. at: http://lib.szie.hu/sites/default/files/files/SZIE_KDKL_WoS_ResearcherID_utmutato.pdf
- Szabó Orsolya, Tóth Eszter (2008). *ISMN útmutató*. Országos Széchényi Könyvtár Magyar ISBN és ISMN Iroda, at: http://www.ismn-international.org/download/Manual_translations/HU_ISMN_utmutato.pdf
- Szakadát István (2005). Réteges struktúra, alapelációk. In: Alexin Z., Csentes D., III. *Számítógépes Nyelvészeti Konferencia*. Szeged, pp. 20-28.
- Szakadát István, Szóts Miklós, Gyepesi György, Varasdi Károly, Ungváry Rudolf, Simonyi András, Gyarmathy Zsófia, Szaszko Sándor, Szeredi Dániel (2006). MEO ontológiamodell. In: Alexin Zoltán, Csentes Dóra (szerk.) *IV. Magyar Számítógépes Nyelvészeti Konferencia*, Szeged, pp. 377-383.
- Szakadát, I., Szóts, M., Gyepesi, Gy. (2006). MEO - Ontology Infrastructure. In Gabor Magyar – Gabor Knapp – Wita Wojtkowski – Gregory Wojtkowski – Joze Zupancic – Stanislaw Wrycza (eds.): *Advances in Information Systems Development : New Methods and Practice for the Networked Society, Proceedings Information Systems Development*. Springer
- Szóts Miklós, Lévay Ákos (2005) Szerepfogalmak az ontológiákban - az OntoClean metodológia továbbfejlesztése, in: Alexin Z., Csentes D., III. *Magyar Számítógépes Nyelvészeti Konferencia*

- The Phone of the Future, in: *Economist*, 2006.12.02, pp.10-12.
- Thomasson, A.L. (2004). Methods of categorization. In Achille C. Varzi – Laure Vieu (eds.): *Formal Ontology in Information Systems. Proceedings of the Third International Conference*. Amsterdam, Berlin, Oxford, Tokyo, Washington DC, IOS Press, 3–16. p.
- Toffler, Alvin. *Harmadik hullám*, Budapest: Typotex, 2001
- Tózsér János (2009). *Metafizika*, Budapest: Akadémiai Kiadó
- Ullman, Jeffrey D., Widom, Jennifer. *Adatbázisrendszerek. Alapvetés*. Budapest: Panem & Prentice-Hall, 1998.
- Ungváry Rudolf, Orbán Éva (2001). *Osztályozás és információkeresés, I-II.* Budapest: OSZK
- Ungváry Rudolf, Orbán Éva (szerk.). *Osztályozás és információkeresés, I-II.* Budapest: OSZK, 2001.
- Vander Wal, Thomas (2005b). *Folksonomy Definition and Wikipedia*, 2005.11.05, at: <http://www.vanderwal.net/random/entrysel.php?blog=1750>
- Vander Wal, Thomas (2007). *Folksonomy*, 2007.02.02, at: <http://vanderwal.net/folksonomy.html>
- Vander Wal, Thomas (2008). *Getting to know collective and collaborative*, 2008.03.29, at: <http://www.personalinfocloud.com/2008/03/getting-to-know.html>
- Vander Wal, Thomas: *Folksonomy Definition and Wikipedia*, 2005.11.05, at: <http://www.vanderwal.net/random/entrysel.php?blog=1750>
- Vander Wal, Thomas: *Off the Top: Explaining and Showing Broad and Narrow Folksonomies*, 2005.02.21, at: <http://www.vanderwal.net/random/category.php?cat=153>
- Varzi, A. Basic problems of mereotopology, 1998. URL citeseer.ist.psu.edu/varzi98basic.html .
- Varzi, A.C. (1996). Parts, wholes, and part-whole relations: The prospects of mereotopology. In *Data and Knowledge Engineering*, 20. évf. 177–198. o.
- Varzi, AC. Mereological commitments. 54. évf. (2000), *Dialectica* , 283–305. p.
- Varzi, Achille C. *Undetached parts and disconnected wholes*. URL citeseer.ist.psu.edu/varzi00undetached.html .
- Varzi, Achille C. Vagueness in geography. *Philosophy and Geography* , 2001, 4. évf. 1. sz., 49–65. p.
- Varzi, Achille C. Vieu, Laure (szerk.): *Formal Ontology in Information Systems. Proceedings of the Third International Conference (FOIS 2004)*. Amsterdam, Berlin, Oxford, Tokyo, Washington DC, 2004, IOS Press
- Vendler, Z. (1957). ‘Verbs and Times’, *Philosophical Review*, 66, pp.143–60.
- Will, Leonard (2012). The ISO 25964 Data Model for the Structure of an Information Retrieval Thesaurus, *Bulletin of the American Society for Information Science and Technology* – April/May 2012 – Volume 38, Number 4, at: https://www.asis.org/Bulletin/Apr-12/AprMay12_Will.pdf
- Zemach, E.M. (1970). Four Ontologies. *The Journal of Philosophy*, 47(8), pp.231-247.

- Zeng, Marcia Lei. *ISO 25964: Thesauri and Interoperability with Other Vocabularies*, at: <https://www.slideshare.net/mzeng/iso-25964-thesauri-and-interoperability-with-other-vocabularies>
- Zeng, Marcia Lei. *Update: The state of KOS in the Linked Data movement*, at: http://www.getty.edu/research/tools/vocabularies/7_itwg_kos_lod_update_zeng.pdf
- Zvolenszky Zsófia (2015). Pofonegszerű, kedves Kripkém? *Műút*, <http://www.muut.hu/?p=15361>

5.1.1 Szabványok, szolgáltatások

- Digitális geo-térbeli metaadat tartalmi szabvány (Content Standard for Digital Geospatial Metadata (version 2.0)) FGDC-STD-001-1998
- Digitális geo-térbeli metaadat tartalmi szabvány, Első Rész: Biológiai adat profil (Content Standard for Digital Geospatial Metadata, Part 1: Biological Data Profile) FGDC-STD-001.1-1999
- Térbeli adatátviteli szabvány (Spatial Data Transfer Standard (SDTS) FGDC-STD-002) (módosított verzióját ANSI NCITS 320:1998 néven fogadták el)
- Térbeli adatátviteli szabvány, Ötödik Rész: Raszter Profil és Kiterjesztések (Spatial Data Transfer Standard (SDTS), Part 5: Raster Profile and Extensions) FGDC-STD-002.5
- Térbeli adatátviteli szabvány Hatodik Rész: Pont Profil (Spatial Data Transfer Standard (SDTS), Part 6: Point Profile) FGDC-STD-002.6
- ISO 25964. Thesauri and interoperability with other vocabularies
Part 1: Thesauri for information retrieval
Part 2: Interoperability with other vocabularies. at: <https://www.niso.org/schemas/iso25964>
- MSZ 3418-87 Magyar nyelvű információkereső teauruszok szerkezete, részei és formái. Budapest, Magyar Szabványügyi Hivatal, 1987
- Egyetemes Tizedes Osztályozás : UDC Publ. No. P057 / [szerk. és bev. Barátné Hajdu Ágnes; közrem. Ackermanné Kelő Kamilla et al.]. – Budapest : Országos Széchényi Könyvtár Könyvtári Intézet, 2005
- SKOS Simple Knowledge Organization System, at: <https://www.w3.org/2004/02/skos/>
- Resource Description Framework (RDF), at: <https://www.w3.org/RDF/>
- Academia.edu, at: <https://www.academia.edu/>
- BIBFRAME, *Bibliographic Framework Initiative*, at: <https://www.loc.gov/bibframe/>
- BNF, at: <http://catalogue.bnf.fr/index.do>
- DOI, at: <https://www.doi.org/>

- FRBR (2008). *Functional Requirements for Bibliographic Records, Final Report*, IFLA, at: https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf
- FRSAD (2010). *Functional Requirements for Subject Authority Data (FRSAD) A Conceptual Model*, at: <https://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>
- GDN, at: <http://www.dnb.de/EN/Standardisierung/GND/gnd.html>
- *Getty Thesaurus of Geographic Names*, at: <http://www.getty.edu/research/tools/vocabularies/tgn/about.html>
- *Getty Vocabulary Program ontology*, at: <http://vocab.getty.edu/ontology>
- Google Maps, at: <https://www.google.hu/maps>
- Google Scholar, at: <https://scholar.google.hu/>
- IMDb, at: <http://www.imdb.com/>
- ISAN, at: <http://www.isan.org/>
- ISBN, at: http://www.oszk.hu/isbn_ismn
- ISMN, at: http://www.oszk.hu/isbn_ismn
- ISNI, at: <http://www.isni.org/>
- ISO 25964 – the international standard for thesauri and interoperability with other vocabularies, at: <http://www.niso.org/schemas/iso25964>
- ISRC, at: <http://isrc.ifpi.org/en/>
- ISSN, at: <http://www.oszk.hu/issn>
- ISTC, at: <http://www.istc-international.org/>
- ISWC, at: <http://www.iswc.org/>
- Magyar Egységes Ontológia (MEO) projekt dokumentumai, at: <http://ontologia.hu>
- MSZ 3418-87 Magyar nyelvű információkereső tezasauruszok szerkezete, részei és formái. Budapest, Magyar Szabványügyi Hivatal, 1987 [Hungarian Standard of Thesauri]
- MTMT, at: <https://www.mtmt.hu/>
- NBN, at: <http://nbn.depositolegale.it/>
- ODT, at: <https://doktori.hu/>
- OpenStreetMap, at: <https://www.openstreetmap.org/>
- ORCID, at: <https://orcid.org/>
- PIM, at: <https://opac-nevter.pim.hu/en/>
- ResearcherID, at: <http://www.researcherid.com/>
- ResearchGate, at: <https://www.researchgate.net/>
- Scopus, at: <https://www.scopus.com/home.uri>
- ULAN, at: <http://www.getty.edu/research/tools/vocabularies/ulan/>
- VIAF, at: <https://viaf.org/>
- Web of Science, at: <http://webofknowledge.com/>
- Wikidata, at: https://www.wikidata.org/wiki/Wikidata:Main_Page

- MSZ 3418-87 Magyar nyelvű információkereső teauruszok szerkezete, részei és formái. Budapest, Magyar Szabványügyi Hivatal, 1987 [Hungarian Standard of Thesauri]
- *National geospatial-intelligence agency (nga) geonet names server.* URL http://earth-info.nga.mil/gns/html/geonames_dd_dms_date_20070302.zip
- Könyvtári és Szakirodalmi Tájékoztatási Szabványosítási Bizottság: Földrajzi nevek, mint adatbázisrekordok tárgyi hozzáférési pontjai. könyvtári és szakirodalmi tájékoztatási szabályzat, 2005. június.
- Open GIS Consortium: Gazetteer service profile of the web feature service implementation specification, Open GIS Consortium Inc.
- FRBR (2008). *Functional Requirements for Bibliographic Records, Final Report*, IFLA, at: https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf
- FRISAD (2010). *Functional Requirements for Subject Authority Data (FRISAD) A Conceptual Model*, at: <https://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frisad-final-report.pdf>
- Getty *Thesaurus of Geographic Names*, at: <http://www.getty.edu/research/tools/vocabularies/tgn/about.html>
- *Getty Vocabulary Program ontology*, at: <http://vocab.getty.edu/ontology>
- BIBFRAME, *Bibliographic Framework Initiative*, at: <https://www.loc.gov/bibframe/>
- *Alexandria digital library gazetteer.* URL [http://www.alexandria.ucsb.edu/downloads/gazdata/adlgaz-namelist-20020% 315](http://www.alexandria.ucsb.edu/downloads/gazdata/adlgaz-namelist-20020%315).