

Az individuumok tér- és időfüggésére vonatkozó modellezési kérdések megoldása

A nemzeti névterek fogalma	2
Miért van szükség nemzeti (és globális) névterekre?.....	5
A névterek szintjei és azok egymásra épülése.....	6
Általános modellezési kérdések	9
Többet a KOS-okról	9
Névterek modellezése	12
Földrajzi névtér	13
A földrajzi entitások modelljének kérdései	13
A földrajzi névtér modellje	15
Geoindividuumok azonosítása	15
Idő- és eseménykezelés	15
Geometriakezelés	17
Névkezelés	17
Relációk.....	18
Földrajzi névtér logikai adatmodellje.....	19
A nyelvi elemek kezelése	20
Személynévtér	23
Személynévtér logikai adatmodellje	24
Testületi névtér.....	25
Testületi névtér logikai adatmodellje	26
Eseménynévtér	26
Mi az az idővonal	27
Alapfogalmak	28
Adatmodel	37
Írásműcímter.....	40
Köznévtér	44
Tudásszervezési rendszerek	44
Terminuslista.....	46
Taxonómia.....	46
Tezaurusz	47
Formális ontológia.....	47
Folkszonómia	47
A köznévtér és a tudásszervezési rendszerek kapcsolata	47
Névtérközi kapcsolatok	48
Archontológia.....	48
Iskolázottság.....	49
Társadalomföldrajz.....	49
Választás.....	50
További névtérközi adatkörök.....	50
Hivatkozások	50



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 megnevezzük 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 megragadhassuk 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 egyed típus konkrét egyedeinek előfordulásaira hivatkozunk. A tulajdonnevek használatának lényegi célja egyedi módon hivatkozni valamilyen konkrét egyedre valamely kontextuson belül. Jól ismert

¹ A tudásszervezési rendszerekről egy későbbi fejezetben bővebben szót ejtünk.

az a tény, hogy ezt a célt a gyakorlatban sokszor nem érik el (legtöbbször a kontextusok nem megfelelő kezelése miatt), de ez a 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 tulajdonnévtérek 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 tulajdonnévtérek 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 névtérek első – még nem teljes funkcionalitású – használatának tekinthetők. A fontosabb szerep, az ebből következő nagyobb érdek, valamint a könnyebb megvalósíthatóság miatt a **személynévtérek** építésében érték el eddig a legnagyobb eredményeket. A **testületi névtérek** 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 tévéműsorszámokról, filmről, VHS-kazettáról, dvd-ről, online videóról, könyvről, periodikáról, cikkről, weboldalról, e-bookról, bakelitlemezzel, cd-lemezzel, cd-rom lemezzel, 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

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

dokumentumtípusnak számít(ott). A különböző médiumok, illetve az ezekhez kapcsolódó különböző intézmények szerinti elkülönítés logikája természetesnek látszott a legtöbb érintett szereplő számára, és az effajta magátólértédek mindig messzemenően tiszteltben 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-lemezzel vagy egy okostelefonon a hálózaton 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 entítások 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 egyedi, individuálisak, tehát technikai értelemben tulajdonneveknek tekinthetők. A domainnevek mint tulajdonnevek kiosztása, felismerése, feloldása a web kezdeteitől 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 közgyűjteményi 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.

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 zártság továbbra is megmarad, noha a hálózati technológiák ugyanúgy lehetővé tennék meghaladni, eltüntetni ezt a fajta zártságot 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 biztostani 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.

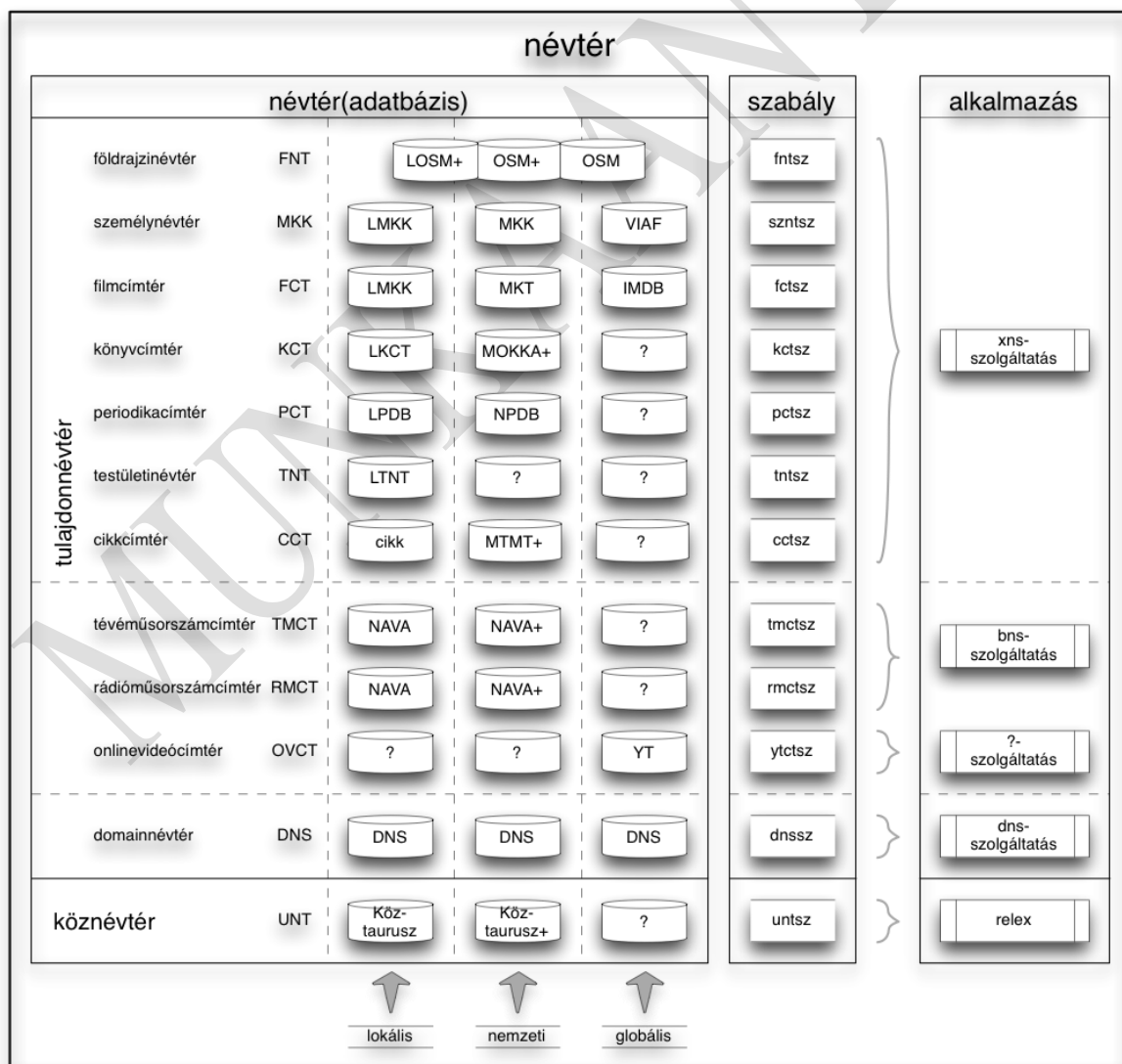
A gyűjtemények közötti átjárhatóságot a közös névterek segítségével lehet biztosítani.

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 – közősek. 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.

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 kezdeti költséges, 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 névteret használó 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 névtereit integráló közös névtér, míg a **globális névtér** valamely névtértípus esetén egy nemzetközi kooperációban működő, több nemzeti és/vagy lokális névtérgazda által fenntartott névteret 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 névtérépítkezési stratégiáját keressük, akkor érdemes minden fontos névteret 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 névtértípusok, 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 (tévés 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 testületi 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 még 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).

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).

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, 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 használnunk az ilyen tudásszervezési rendszereket: mindig valamilyen

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

partikuslárius 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-rendezett – 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 változat 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 egybemosták a hierarchikus jellegű (generikus, partitív, időbeli) relációkat egymással az egyetlen klasszifikációs (igazából csak valamilyen általános alá-főlérendelést 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 köznévvel (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 partikulárnak 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 téridő 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 tárgyalt szempontjából nem tűnik lényegesnek). A kérdés, vajon miként tehetjük ezt meg. Sajnos az ismert tudásszervezési rendszerek (a klasszifikációk, a teauruszok) 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 teaurusz komponensei is, de szükség van további relációk felvételére. Mielőtt kifejténénk, 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.

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 **névhordozókat** (vagy mindkettőt). A válasz az, hogy elvileg és értelme szerint névhordozókat (személyeket, testületeket, földrajzi helyeket stb.), viszont a gyakorlatban mindkettőt vagyis névhordozókat és neveket egyszerre. A tulajdonnevek használatának igazi célja mindig az, hogy velük egyértelműen rá tudjunk mutatni valamely egyedi dologra (névhordozóra). 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 Helységnevtárában – ezzel szemben – a nevekhez rendelnek azonosítókat (noha a rendszer lényege szerint a névhordozókra, konkrétan a településekre koncentrálnak).

Van tehát két entitástípusunk, a nevek és a névhordozók. Ha mind a névhordozókat, mind a neveket kezelni szeretnénk a névterünkben, márpedig erre szükség van, akkor:

mindkét entitástípus (a névhordozók 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 névhordozók 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 **lokálizálására** 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 névhordozók 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.

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.

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 akartól 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 entítások leírásakor. Mindkét objektumtípus előfordulhat a földrajzi objektumok között, tehát beszélhetünk fiat és bona fide földrajzi entításokról (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 entítások legfontosabb jellemzőit – konstitutív jellegük miatt – mindig az időben elhelyezve, időbeli paraméterekkel ellátva kell leírni.

A földrajzi entítások 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 valamilyen időbeli adat milyen pontossággal rendelkezik (perc-, óra-, nap-, hónap-, évpontos stb.).

A földrajzi hely legfontosabb paramétere a **geometriája**, aminek mindig van típusa (pont-, vonal- vagy poligonyszerű), és ez a geometria változhat az időben (például egy település határvonala változik). A földrajzi entításoknak 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 valamilyen 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 tívialisán meg tudjuk adni, hiszen minden diskurzusban így hivatkozunk rájuk.

A földrajzi névtér modellje

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

Idő- és eseménykezelés

Az időkezelés modellezési kérdéseinek áttekintését kezdhethetjük 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 valamely 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/perc-pontos 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 hivatkozunk. Ennek a megoldásnak további előnye még, hogy nem kell felvenni egy újabb

táblában azt az információt, amely összekapcsolja a két eseményt (az egyik név "megszüntét" és a másik név "keletkezését") egymással. Még egyértelműbben látszik e megoldás előnye, ha egy olyan esetre gondolunk, amikor egy városba egyidőben 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.

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.

Névkezelés

Az individuumokat nevekkel 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. Modellezési 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.

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 példány (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 kapcsolatrendszeret, 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 praktikusán 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 geoindivídium között partitív alá-, fölérendeltség létezik (például Szentendre része Pest megyének, illetve a Duna egy része határos Szentendrével), akkor az összekapcsolt geoindivídiumok 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 defináljunk. 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 tulajdonnévtér legfontosabb relációja a példány reláció, hiszen a névtérben valamely entitáshoz kapcsolható elemeket, konkrét névhordozókat akarunk egyértelműen azonosítani, amihez konkrét előfordulásokat, példányokat kell kezelnünk. A példány reláció központi jelentősége más formában is megfogalmazható:

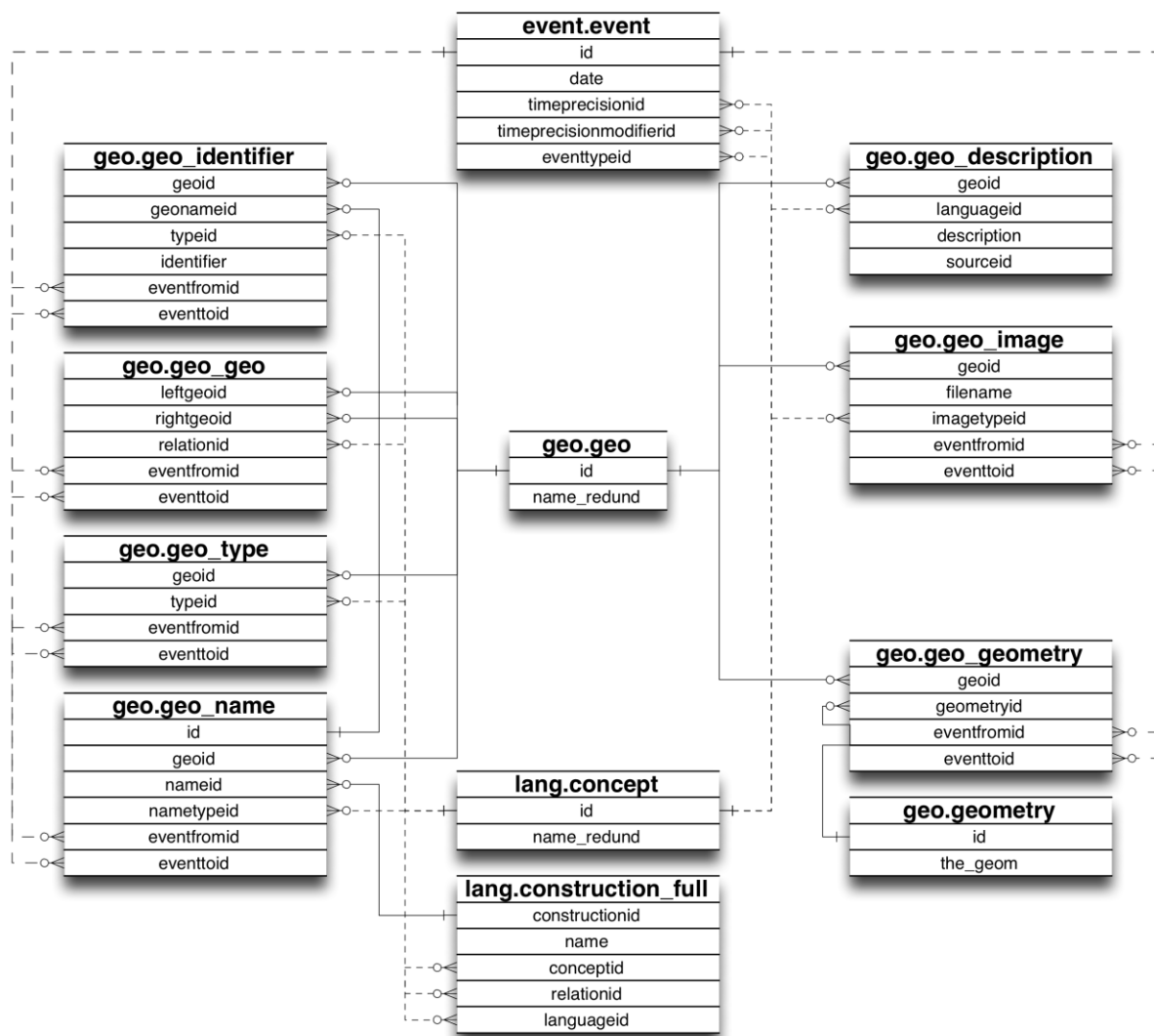
az individuumokat mindig valamilyen általános fogalom alá kell rendelni (vagyis mindig meg kell adni a geoontológiai típusukat).

A tezaszokban 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 megengednünk, mert pontatlan meghatározásával több kárt okozhatunk, mint amennyi előny származik a – látszólagos – egyszerűségéből.

A tezuszban használt deszkriptor, nem-deszkriptor kategóriákra itt nincs szükség. Ez a két kategória a szinonimitás kezelésére alkalmas. A tulajdonnévtérben a szinonimitást nem kell az átkános fogalmak szintjén kezelnünk, 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.

Földrajzi névtér logikai adatmodellje

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



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 reálisan, 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 szegmenseibe 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 fogalomként 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 felvesszünk 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 geoindividuumait, 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 adatsere-, részleges vagy teljes export/import-folyamatokat lehet menedzselni.

Egy nyelvi konstrukció mindig valamilyen nyelven kifejezett megnyilatkozás, amely az adott nyelv morfémakészletéből 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 morfémaiból 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 – vizuális-írásos 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 morfémat állít elő. Az ugyanazon stringet használó morfémaikat úgy tudjuk egyértelműen elkülöníteni egymástól, hogy a stringhez való kapcsolatukban megadjuk meg azt, hogy milyen nyelvű morfémaról van szó, illetve milyen alapvető nyelvi típusbesorolása (legtöbb esetben: milyen szófaja) van a morfémanak. 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 morphtype



fog	fog	angol	főnév
fog	fog	magyar	főnév
fog	fog	magyar	ige

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

Decsi-Nagy-Holt-Duna = Decsi | - | Nagy | - | Holt | - | Duna

megtalált = meg | talál | t

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

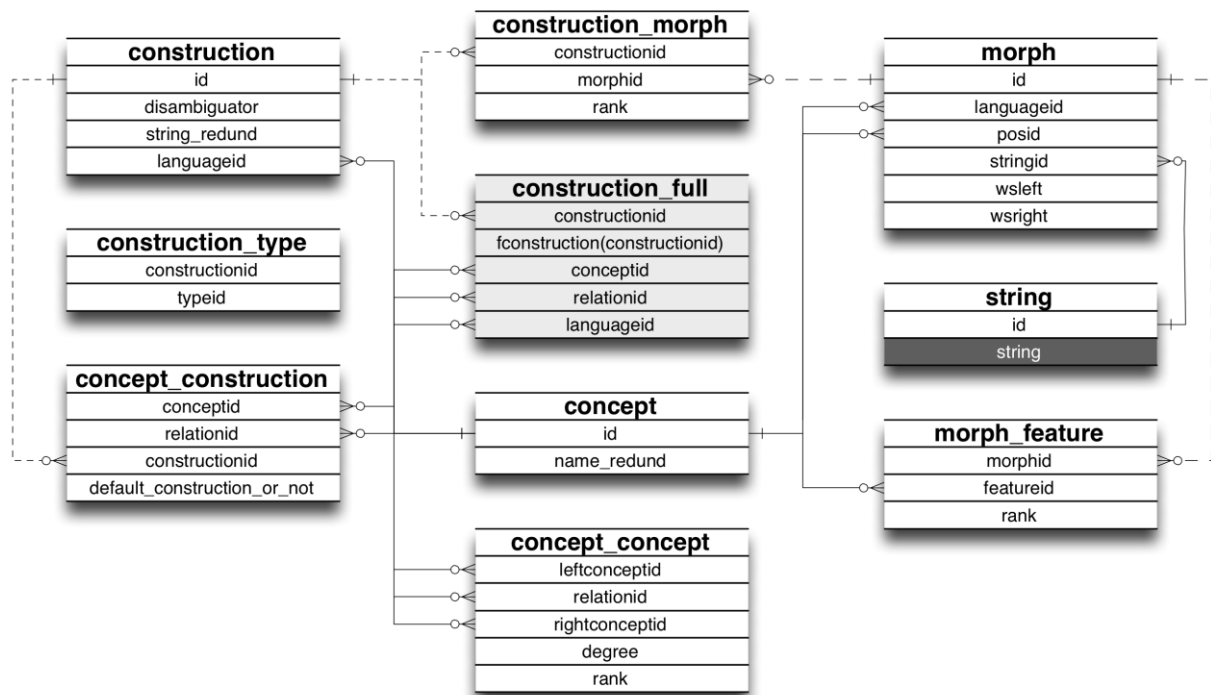
construction	component	rank
Decsi-Nagy-Holt-Duna	decsi	1
Decsi-Nagy-Holt-Duna	-	2
Decsi-Nagy-Holt-Duna	nagy	3
Decsi-Nagy-Holt-Duna	-	4
Decsi-Nagy-Holt-Duna	holt	5
Decsi-Nagy-Holt-Duna	-	6
Decsi-Nagy-Holt-Duna	duna	7

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-főlérendeltje' 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.



A modellben egyetlen helyen "megy be" nyelvi/karakteres információ: a string entitás fogad be karaktereket. A kompozíciós elv miatt azonban fontos feltétel, hogy a string táblában nem lehet 'space' karakter. Kivételt jelentenek azok a mezők, amelyekbe az emberi feldolgozás számára hasznos szöveges információt lehet elhelyezni, de ezek a mezők redundánsak a nyelvi elemek gépi kezelése szempontjából. A modell alapján felépíthető 'construction_full' nézettábla nyújtja a névterek számára szükséges nyelvi adatokat (természetesen ez a nézettábla nem alkotórésze a modellnek).

Személynévtér

A személynévtér modellezésével kapcsolatban megfogalmazhatunk egy kiinduló tézist, miszerint mind a személy, mind a testület kategóriája besorolható az ágens fogalma alá. Az ágens az az entitás, amely valamilyen szerepben hozzájárul a kulturális dokumentumok létrejöttéhez, fennmaradásához, befogadásához stb. Az ágens alaptípusa a természetes személy, az ember, akit döntően természeti tények segítségével írhatunk le. A testület természetes személyek adott csoportjának, valamint szerepek, szabályok, értékek stb. összetartozó halmazának társadalmilag konstruált együttese. A társadalmi konstruáltság miatt a testület leírásához sok esetben társadalmi tények rögzítésére van szükség. A fogalmi különbségek ellenére, az ágens-jelleg miatt a személy és testület entitásainak modellje nagyon hasonló.

A személynévtér valós, történelmi személyek egyértelmű azonosítására szolgál, valós személyek adatait tartalmazza. A személyek azonosítására a születési, halálozási adatait, a

hozzájuk rendelt személyneveket, valamint foglalkozásneveket használhatjuk. A személynevek személyhez rendelése ugyanolyan módon történik, mint a földrajzi helyek esetében, vagyis a neveket a nyelvi rétegből hívhatjuk meg, a személynévtér modelljében nincsenek karakteres mezők. A személynevek önmagukban nem alkalmasak arra, hogy azonosítsák a nevek hordozóit, hiszen minden név esetében meg kell engednünk azt a lehetőséget, hogy több személyhez legyen rendelve. Ezért kell a nevek mellé további adatokat felvennünk, hogy a személyek azonosítása egyértelművé válhasson.

A születési és halálozási adatokkal a személyeket és a földrajzi helyeket kötjük össze – a születési és halálozási – események segítségével. Noha a személyek is individuumok, ezért rájuk igaz is az a tétel, hogy létezésük során sosem foglalhatják el egyszerre ugyanazt a pontot a térben és időben, ezt a feltételt mégsem olyan könnyű a személyek azonosítására használni, mert praktikusán szinte végtelen helyen és időpillanatban lehetünk a téridőn belül. A megoldás az lehet erre a problémára, hogy nevezetes eseményeket emelünk ki a személyek életéből, és azok mentén próbáljuk meg egyértelműen azonosítani 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 egypetűjű 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.

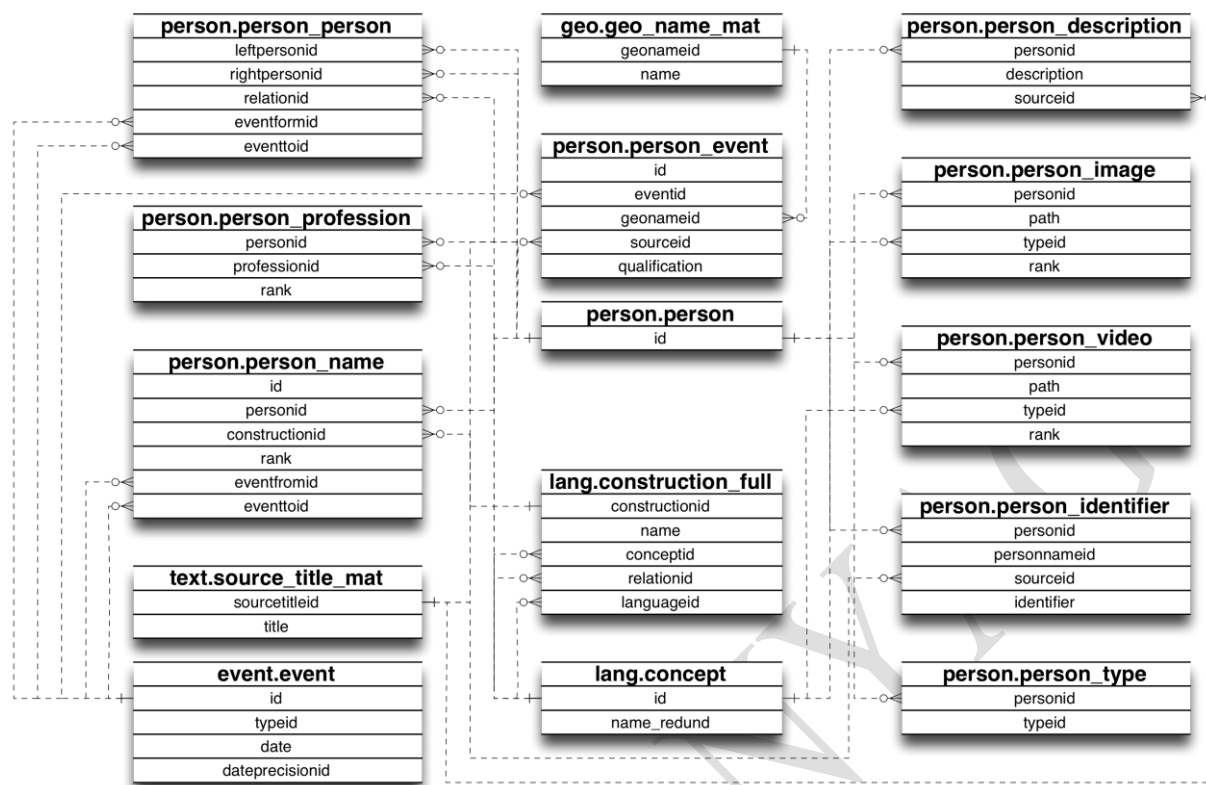
A személyek tipizálását többféle szempont szerint el lehet végezni. A közgyűjteményi adatbázisok közti átjárhatóság megteremtésére szolgáló személynévtérben a személyek nemének, nemzetiségének, valamint az élete során művelt foglalkozások listájának megadása elégséges alapot nyújthat ahhoz, hogy a személyt (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.

Személynévtér logikai adatmodellje

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



Testületi névtér

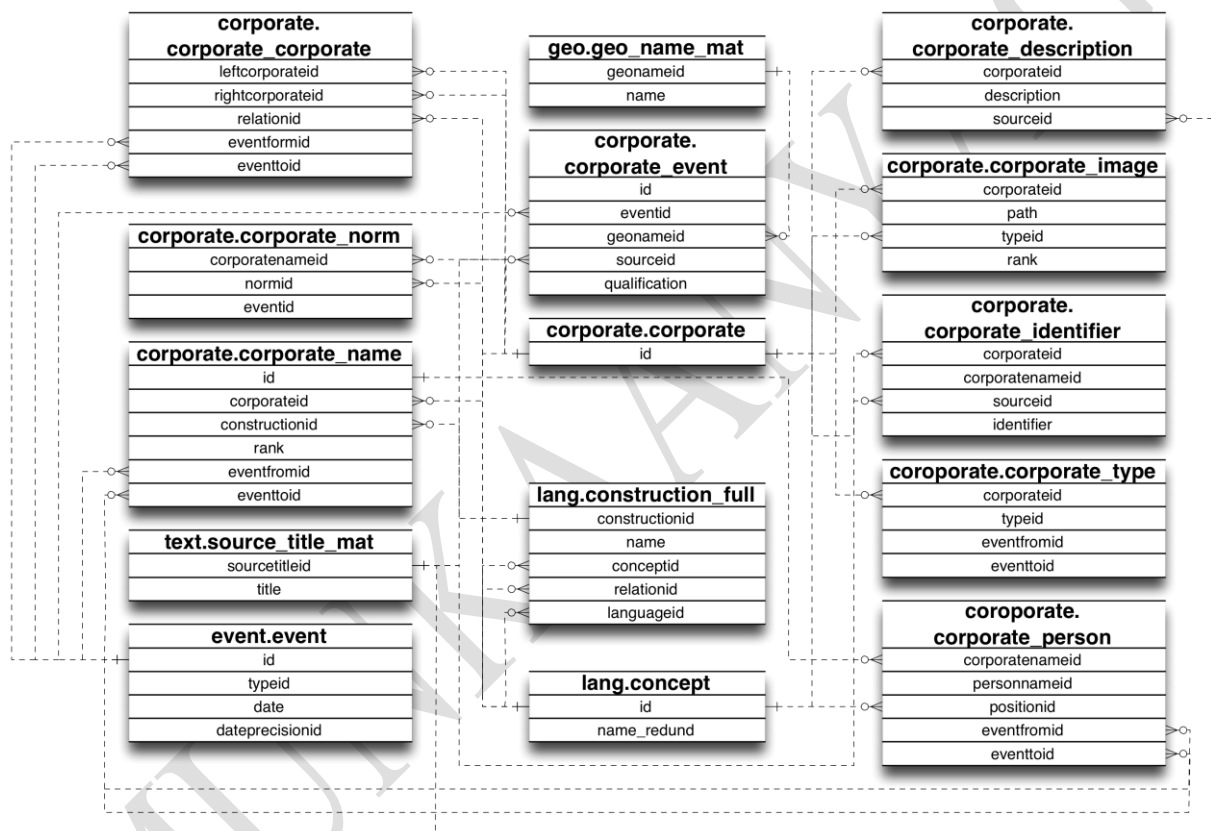
Ahogy azt már korábban jeleztük, a testület fogalma az ágens alá sorolható be. Testületről beszélünk akkor, ha emberek valamilyen csoportját összekapcsoljuk a testületi létüket meghatározó szabályok, értékek, szerepek együttesével. Testületekről van szó akkor, amikor a hétköznapi életben szervezetek, 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 tér-időben, hanem a konstruált, szimbolikus, társadalmi tér-időben 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 a 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 újrakezdeni 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.

Testületi névtér logikai adatmodellje

A testületi névtér adatmodellje az alábbi ábrában foglalható össze.



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 az 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 névtérszegmensekhez 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 idővonalszolgáltatás logikájának megfelelően kell kezelni. A továbbiakban összeszedjük azokat a szempontokat, definíciókat, amelyekre szükség lehet az idővonal-szolgáltatás megtervezéséhez és fejlesztéséhez.

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 főtípusa 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, még ha 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 ismételhető.
- Egy idővonal (adatbázis és alkalmazás) számára gyakorlati szempontból az időfüggőség 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 TÉRPONTOSSÁGÁT (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.

Alapfogalmak

A kronológiaszerver modellezéséhez, felépítéséhez szükséges alapfogalmakat a következőképpen definiálhatjuk.

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.

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)

idővonalszerver (timeline server)

- Az elemi események kezelésére alkalmas alkalmazás.

kalendárszerver (calendar server)

- A naptáreseemények kezelésére alkalmas alkalmazás.

kronológia-adatbázis (chronology database)

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

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.

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 idővonalon is (akár külön időrétegekben, akár popupablakokban, akár máshogy).

- Az eseményeket időbeli értékük alapján idővonalba (kronológiai rendbe, sorba) rendezhetjük.
- Egy esemény lehet:
 - történet (instanteseemény, 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 koordináta-rendszerben 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.

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.

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

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

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 rövidebb 'esemény' kifejezést is.

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 preperiodikus jellegűt a poláris koordináta-rendszerben lehet a legkönnyebben, legadekvátabb módon megragadni.
- A naptáresemények speciális tulajdonságát kezelni képes alkalmazást nevezhetjük kalendárszervernek (vagy naptárszervernek).

idővászon (time canvas)

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

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 használhatóak. Ez az elvárás nem azt jelenti, hogy minden időrétegben feltétlenül 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.

időszakdoboz (period box)

- Az időszakdoboz egy események valamilyen szempont szerint megszürt halmazát tartalmazó, négyszögletes 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).

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 koordináta-rendszeren alapul.
- Az időtengely másik lehetséges használati módja poláris koordináta-rendszeren alapul, és adekvát módon a naptáresemények kezelésére alkalmas, de elképzelhető olyan megoldás is, hogy – valamilyen spirális interfészen keresztül – végtelenített időtengelyként funkcionáljon.

időfelbontás (time granularity)

- Könnyen előfordulhat, hogy egy idővonal-adatbázis annyi eseményt tartalmaz, hogy azok egyszerre már nem jeleníthetők meg az idővásznon 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 idővásznon 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.

időfelbontási szint (granularity level)

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

időközelítés (time zoom-in)

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

időtávolítás (time zoom-out)

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

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.

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.

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 életeseményei, 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 fókuszot, figyelmet irányítanak egy eseményvonalra, 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 képességet kell kialakítani a felhasználók számára (rétegválasztás, rétegekikapcsolás, rétegbekapcsolás, rétegsorrend-kialakítás)
- Az eseményekre kényszereket lehet alkalmazni a dátumértékek alapján (pl. kezdődátum < záródátum vagy befoglaló esemény kezdődátuma <= részesemény kezdődátuma)
- finomabb adatmodell esetén saját sémával rendelkező eseménytípusokat lehet definálni (személyes életesemény, archontológiai esemény, publikálási esemény stb.)
- saját sémával rendelkező eseménytípusok esetén lehetséges típuskényszereket definálni
- saját sémával rendelkező eseménytípusok esetén lehetséges önálló névtereket 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

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.

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 aktuálishoz 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 egyedként kell felvenni az események minden összetevőjét, kapcsolatát.

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.

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 idővonalszervernek biztosítani kell egy

adminisztratív felületet, amelyen keresztül a felhasználók, szerkesztők elvégezhetik az effajta minősítéseket.

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 szinterein 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, megéreztessek 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.

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:

$\exists e(\text{megken}(\text{János}, \text{kenyér}, e) \wedge \text{kenőanyag}(\text{vaj}, e) \wedge \text{kenési_idő}(2013.03.30, e) \wedge \text{kenés_módja}(\text{lassan}, e))$

- o 2. megoldás:

$\exists e(\text{kenő_ágens}(\text{János},e) \wedge \text{megkent}(\text{kenyér},e) \wedge \text{kenőanyag}(\text{vaj},e) \wedge \text{kenési_idő}(2013.03.30,e) \wedge \text{kenés_módja}(\text{lassan},e))$

- 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:
 - o az eseménytípusok egymástól független (különböző) sémákkal rendelkezhetnek
 - o az eseménytípus-modelleket menet közben is lehet bővíteni
 - o 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:
 - o $\text{date}(e)=1989.10.13. 8:23 \rightarrow e_year(1989,e) \wedge e_month(10,e) \wedge e_day(13,e) \wedge e_hour(e,8) \wedge e_minute(23,e)$
 - o $\text{date}(e)=1920\text{-as évek közepén} \rightarrow e_decade(1920,e) \wedge e_timemodifier(\text{közepén},e)$
 - o $\text{date}(e)=\text{tizenkilencedik század elején} \rightarrow e_century(19,e) \wedge e_timemodifier(\text{elején},e)$
- 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 individuum-kapcsolatait. 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 lehesen 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).

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 névtérszegmens 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 milyeneként 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**.

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 kiválogatódnak azok a **művek**, amelyekhez szerzőket, kreatív alkotókat rendelhetünk, és szerzői jogokat tulajdonítunk nekik, illetve a művek azonosítása végett címeket 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 példányaikat 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

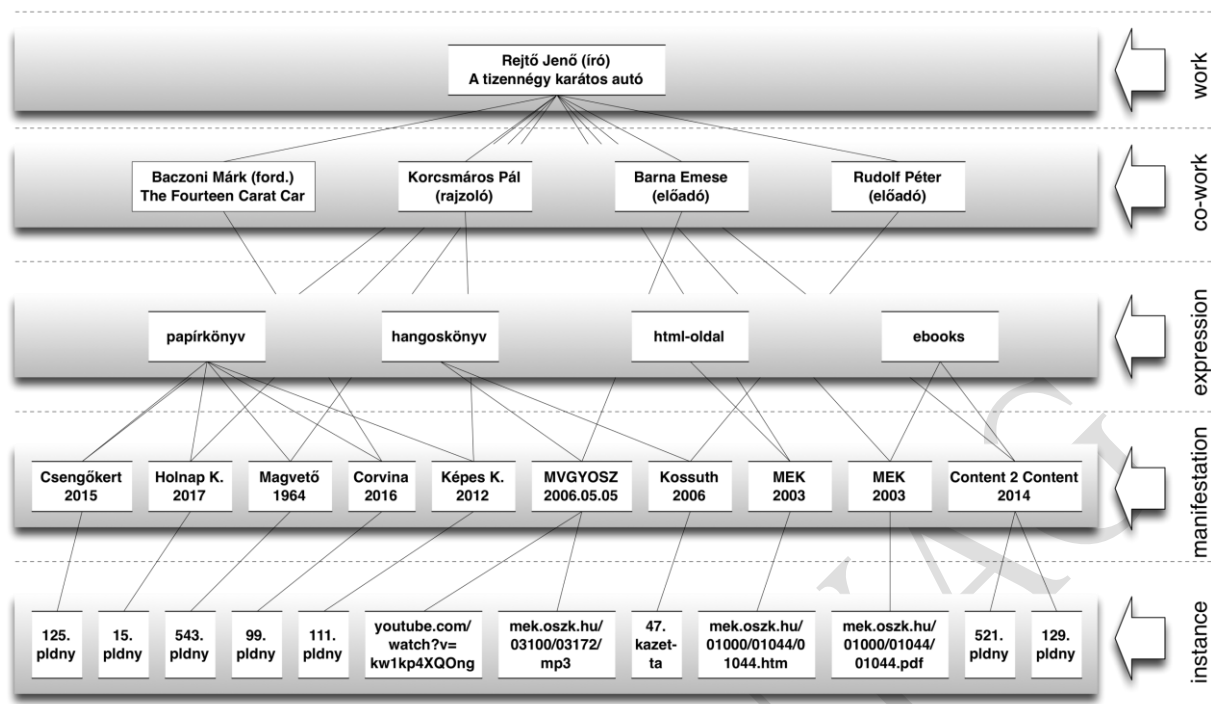
kiadványok (manifestation) mellett a műveket (work) és a művek különböző kifejeződési formáit (expression). Utóbbi szint behozatalával lehet remélni azt, hogy össze tudjuk kapcsolni egy írásos mű (mondjuk: egy regény) hagyományos könyvkiadványait a mű más formátumú, más médiumokon keresztüli (például 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 címekkel) 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 szólalnak meg. Nem elég tehát azt mondani, hogy ezeket hangoskönyvként elváltatjuk a papírkönyvektől (az expression szinten), és megengedjük, hogy a hangoskönyveknek legyenek egymást követő kiadásai, mert így nem tudjuk elváltatni 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 **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 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 írásműcímter-prototípust érdemes felhúzni.

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.

Tudásszervezési rendszerek

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:

$KOS = \langle D, R_1, R_2, \dots, R_i, R_n \rangle$, 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

o 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 gyelembevételére. 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 feltételegyüttestet 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} dokumentumcímkézési gyakorisági függvény
R_8 rokona (egyéb) reláció	R_{17} tetszőleges relációfogalom
R_9 lásd és helyette szinonimareláció-pár	R_{18} tetszőleges osztályfogalom
	R_{19} tetszőleges attribútumfogalom

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

Terminuslista

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

R_1 lexikografikus rendezési reláció

R_2 ekvivalenciareláció

Taxonómia

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

R_1 lexikografikus rendezési reláció

R_3 hierarchikus alárendeltje (tartalmazási) reláció

R_4 különbözőségi reláció

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

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

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

R_1 lexikografikus rendezési reláció

R_5 generikus alá- és fölérendeltje relációpár

R_{17} tetszőleges relációfogalom

R_{18} tetszőleges osztályfogalom

R_{19} tetszőleges attribútumfogalom

Folkszonómia

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

R_1 lexikografikus rendezési reláció

R_{10} ekvivalenciareláció

R_{11} címkegyakorisági reláció

R_{12} címke-együttjárési reláció

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 teaurusz 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. Praktikusan tehát könnyen megoldható a probléma, a Nemzeti Névtér projekt számára pedig nem életbevágóan fontos az elméleti kérdés végleges tisztázása.

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.

Archontológia

Az archontológia a történettudomány egyik fontos segédtudománya. 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 névtérszegmens, a személynévtér és a testületi névtér közti kapcsolatokat írja le úgy, hogy közben az eseménytér segítségével helyezi el az időben ezt a kapcsolatrendszert.

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).

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 névtéroldalain 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.

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.

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.

Hivatkozások

- 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
- BIBFRAME, *Bibliographic Framework Initiative*, at: <https://www.loc.gov/bibframe/>
- 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, Donnelly, Maureen, Smith, Barry (2004). 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)*. IOS Press, 37–48.
- 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

- Cruse, D.A. (1986). *Lexical Semantics*, Cambridge University Press
- 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.
- 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>
- *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>
- Gómez-Pérez, A., Fernández-López, M., Corcho, O. (2004). *Ontological Engineering*, Springer
- Green, Rebecca, Bean, C.A., Myaeng, S. Hyon, *The Semantics of Relationships: An Interdisciplinary Perspective*, Dordrecht: Kluwer, 2001.
- Grenon, P., Smith, B. (2004). SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition and Computation*, 4(1), pp.69-103.
- Guarino, N. (1998). Formal ontology in information systems. In Nicola Guarino (szerk.): *Formal Ontology in Information Systems. Proceedings of FOIS'98*. Trento, Italy, IOS Press, Amsterdam, 3–15. p.
- Guarino, N., Welty, C. (2000). Supporting ontological analysis of taxonomic relationships. *Data & Knowledge Engineering* 39, pp 51-74
- Guarino, N., Welty, C. (2002). A Formal Ontology of Properties. In. *Proceedings of 12th Int. Conf. on Knowledge Engineering and Knowledge Management Lecture Notes of Computer Science*, Springer Verlag
- Heuvelmann, R. (2012). *From Authority Control to Linked Authority Data*, at: http://connect.ala.org/files/ALA_MFIG_Heuvelmann.pdf
- Hubay Miklós Péter (2016). *A BIBFRAME és a könyvtári feldolgozás új keretei*, at: <http://mek.oszk.hu/15600/15678/15678.pdf>
- *ISO 25964 – the international standard for thesauri and interoperability with other vocabularies*, at: <http://www.niso.org/schemas/iso25964>
- 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.nyud.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.
- Lyons, J. (1977). *Semantics*, Cambridge University Press
- Magyar Egységes Ontológia (MEO) projekt dokumentumai, at: <http://ontologia.hu>

- MSZ 3418-87 Magyar nyelvű információkereső teauruszok szerkezete, részei és formái. Budapest, Magyar Szabványügyi Hivatal, 1987 [Hungarian Standard of Thesauri]
- Ryle, G. (1949), *The Concept of Mind*, London: Hutchinson.
- Searle, J. (1995). *The Construction of Social Reality*. Simon and Schuster
- Smith, B., Varzi, A. (2000), Fiat and Bona Fide Boundaries. *Philosophy and Phenomenological Research*, 60, pp. 401–420.
- Sowa, J.F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co.
- Staab, S. Studer, R. (2009). *Handbook of Ontologies*, Springer Verlag
- Strawson, P. F. (1959). *Individuals. An Essay in Descriptive Metaphysics*. London and New York, Routledge.
- Syi (2011). Navigáció. In: Kangyal András, Laufer László (szerk.) *Gépézet: Interfész, interakció, navigáció*, Budapest: L'Harmattan Kiadó, pp. 196-294.
- Szakadát István (2005). Réteges struktúra, alaprelációk, in: Alexin Z., Csenedes D., *III. Magyar Számítógépes Nyelvészeti Konferencia*
- Szakadát István (2005). Réteges struktúra, alaprelációk. In: *III. Számítógépes Nyelvészeti Konferencia*. Szeged, pp. 20-28.
- Szóts Miklós, Lévy Ákos (2005) Szerepfogalmak az ontológiákban - az OntoClean metodológia továbbfejlesztése, in: Alexin Z., Csenedes D., *III. Magyar Számítógépes Nyelvészeti Konferencia*
- 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