

A Radar adatrendszer modelljének leírása

RADAR alapadatok

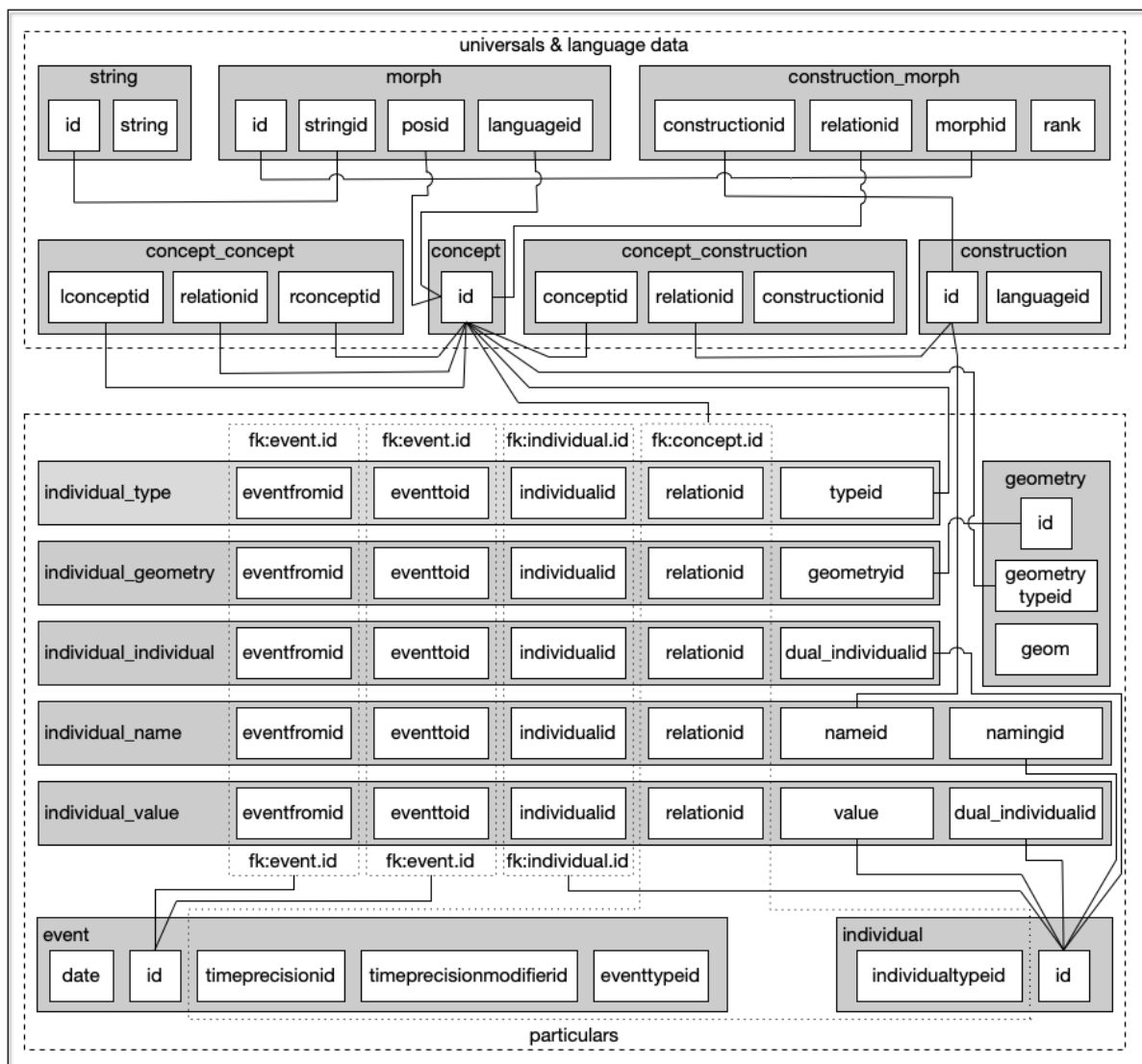
A **RADAR** adatrendszer alapvetően a földrajzi térhez kapcsolható, individuális adatokat tartalmazza, de a rendszer fogadókész bármilyen más individuális adat fogadására is. A rendszer két nagyobb sémában tárolja az adatokat. A *RADAR* sémában vannak az individuális jellegű adatok, a *LANG* sémában az általános fogalmi tudásunkat reprezentáló, nyelvi jellegű adatok.

Mivel az individuális adatok térhez és időhöz vannak kötve, és az őket kezelő rendszernek kezelni akarja ezt a két-irányú függést, akkor minden lehetséges ponton fel kell venni az ilyen jellegű kapcsolati adatokat. A **RADAR** adatmodellje ezt az elvárást igyekszik kielégíteni.

Amilyen mértékben sikeres, olyan mértékben válik képessé arra, hogy a jelenbeli állapotok leírásával azonos módon tudja kezelni a múltbeli állapotokat. Elvileg minden individuális adatrendszer történeti kell legyen (az időfüggés miatt), de a gyakorlatban ritkán szoktak időfüggő rendszereket építeni.

RADAR séma

Az individuális adatokat tartalmazó *RADAR* séma szerkezetét az alábbi ábrán mutatja. Az ábrán felső részében vázlatosan látható a *LANG* séma szerkezete is.



Általános és egyedi fogalmak

A világban vannak dolgok, amelyet az emberi agy valamilyen módon képes reprezentálni, vagyis fogalmakat képezve gondolni rájuk, majd ezeket a reprezentációkat (fogalmakat) az ember képes valamilyen nyelven kifejezni és kommunikálni mások számára.

Filozófiai értelemben kétféle entitásról beszélhetünk. Egy entitás, amelyről adatokat rögzíthetünk, lehet:

- UNIVERZÁLÉ (általános fogalom, fogalom - CONCEPT), illetve lehet
- PARTIKULÁRÉ (individuális fogalom, instancia, előfordulás, individuum - INDIVIDUAL).

A kétféle entitás közti filozófiai különbség a térhez és időhöz való viszonyban van. Az UNIVERZÁLÉ (fogalom) nem lokalizálható a térben és időben, míg a PARTIKULÁRÉ (individuum) mindig a téridő adott pontjához köthető.

Az EMBER általános fogalma olyan entitást reprezentál, amit nem tudunk a téridő valamely pontjához kötni. Ezzel szemben egy konkrét, létező ember (*Kovács János*) mindig lokalizálható a téridőben. Ha egy embert azonosítani akarunk, elvileg mindig adott az a lehetőség, hogy kijelentsük, valaki a földrajzi tér adott helyén tartózkodik adott pillanatban. Más nem foglalhatja el a téridő kontinuum ugyanazon pontját. Praktikus szempontok miatt természetesen másfajta azonosítási technikákat használunk az ember esetében (születési hely és idő, anya neve, de ezekben az azonosításra szolgáló adatokban is ott a tér- és időfüggés).

Az UNIVERZÁLÉ-PARTIKULÁRÉ fogalmi kettős megfeleltethető az informatikában jól ismert osztály-példány dichotómiának.

Az UNIVERZÁLÉ-PARTIKULÁRÉ fogalmi kettősre a **RADAR** modellben a CONCEPT és INDIVIDUAL kategóriapárost használjuk.

Lehet arról gondolkodni, vitatkozni, hogy léteznek-e UNIVERZÁLÉK vagy sem, de azt nem érdemes vitatni, hogy egy reprezentációs rendszerben van értelme általános fogalmakat használni, hiszen ezek által vagyunk képesek gazdaságosan hivatkozni az egyedi, konkrét előfordulások általános minőségeire.

Az általános fogalmakat kétféle módon reprezentálhatjuk. A fogalmak által kiformálható fogalmi struktúrát valamilyen mesterséges, a gép által is értelmezhető, formális nyelven írhatjuk le. Ezt a formális ontológiát eredményező munkát a **RADAR** modell *LANG* nyelvi sémájában tárolt adatok alapján végezhetjük. A rendszerben tárolt tudást az ember számára a fogalmakhoz rendelt, valamilyen természetes nyelven kifejezett nyelvi kifejezések tehetjük értelmezhetővé. A fogalmak ebben az értelemben természetesnyelv-függetlenek.

Individuális fogalmakkal kapcsolatos relációk

Az egyedi individuumokról általános fogalmak segítségével tudunk állításokat tenni. Az első állítás egy konkrét individuumról az lehet, amikor deklaráljuk, hogy létezik, de megadhatjuk további tulajdonságait, más individuumokhoz való kapcsolatait is.

- egy település (*Balatonakarattya*) létezését azáltal deklarálhatjuk, hogy megadjuk, melyik fogalom terjedelme alá sorolható be
Balatonakarattya | IS-INSTANCE-OF | SETTLEMENT
- ezután megadhatjuk finomabb fogalmi besorolását
Balatonakarattya | IS-INSTANCE-OF | VILLAGE
- lokalizálhatjuk a földrajzi térben való helyét, vagyis hozzárendelhetjük a geometriáját
Balatonakarattya | HAS-GEOMETRY | *geom:xyz*
- megköthetjük az időben
Balatonakarattya | EXISTS | *eventfrom:date* | *eventto:date*
- rögzíthetjük a nevét
Balatonakarattya | NAMING | *lang:char*
- hozzárendelhetünk tetszőleges tulajdonságot
(*Balatonakarattya* | POPULATION | *number* | *event:date*)
- leírhatjuk más individuumhoz való viszonyát
(*Balatonakarattya* | IS-PART-OF | *Veszprém megye*)

Az individuumokat leíró adatokat a **RADAR** sémába írjuk be. Mivel itt alapvetően partikuláris adatokat tárolunk, minden predikátum (tábla) rendelkezik az időhöz való kötés biztosító mezővel. Mivel az időhöz kötés eseményeken keresztül valósítjuk meg (lásd az időkezelésről szóló fejezetet), és az események lehet pontszerűek, illetve szakaszszerűek, az egyszerűség és egyértelműség kedvéért minden predikátumban (táblában) két esemény

mező van (`eventfromid`, `eventtoid`), amelyeket értelemszerűen kell kitölteni attól függően, hogy milyen típusú eseményről van szó.

Előfordulhat, hogy egy entitás valamilyen tulajdonsága, kapcsolata kapcsán nem ismerjük az időhöz kötés adatát, ezért a két idővel kapcsolatos mezőre nem írunk elő kötelező kitöltési feltételt.

A *RADAR* sémában van egy **individual** tábla, ami csak azt a funkciót szolgálja, hogy felvehessünk egy új példányt a létező individuumok közé (és ezáltal ID-t adhassunk neki). A séma úgy biztosítja az individuális entitások földrajzi térhez, illetve időhöz való kötésének lehetőségét, hogy önálló entitásként definiálja mind a teret (**geometry**), mind az időt (**event**). Az időt minden individuumokkal kapcsolatos táblában fel kell venni, míg a térhez való viszonyt egy külön predikátumban (táblában) fejezhetjük ki.

A sémában három olyan tábla van (**radar.individual**, **radar.event**, **radar.geometry**), ahol a nem vesszük fel önálló adatként a predikátum lényegét jelentő fogalomra utalást, ti. azt, hogy valami (egy individuum, egy esemény, egy térbeli entitás) létezik (EXITS), mert a három tábla léte kifejezi ezt aényt.

Helykezelés

A **geometry** táblába lehet beírni azokat az adatokat, hogy a földrajzi entitások milyen alakzatot formálnak a térben, vagy másként: milyen földrajzi entitásokra hivatkozhatunk a modellen belül. A modell ezen részében csak azt írjuk le, hogy milyen geometriai minőségeket kezelünk a rendszeren egészen belül. A térhez való viszony önmagában még nem határoz meg egyértelműen egy entitást, hiszen időben változhatnak a dolgok földrajzi kapcsolatai.

Előfordulhat, hogy ugyanaz a geometria több földrajzi entitáshoz is hozzárendelhető: egy adott geometria reprezentálhatja Duna nyomvonalát is, de az országhatár egy részét is, vagy egy település határvonala teljesen megegyezhet a parlamenti választások során definiált szavazóköri valamelyikének határaival.

Az individuumokhoz külön predikátum (tábla) segítségével kapcsolhatjuk a térhez (az **individual_geometry** táblába tett bejegyzéssel), amikor már meg kell (lehet) adni az éppen leírt entitás és a földrajzi tér közti viszony időbeliségét is.

A települések határvonalainak változását úgy kezelhetjük, hogy ahány geometriai alakzat létezett az idők során (és ha van róla adatunk), mindről felveszünk egy-egy rekordot a **geometry** táblába, és az **individual_geometry** táblába írjuk be a geometriákra való hivatkozásokat, valamint a határváltozások időpontjaira mutató eseményazonosítókat.

Időkezelés időpontossággal és eseményeken keresztül

Az individuális létezők időhöz való viszonyát úgy írjuk le, hogy természetben tapasztalt periodicitáshoz igazodva mértékeket képzünk (másodperc, perc, óra, nap, hét, hónap, év, évtized, évszázad, évezred stb.), és ezek segítségével a létező dolgok bizonyos minőségeit az időtengelyhez kötjük. Ehhez természetesen az adatbázis-rendszerek által biztosított külön – date/time típusú – adattípust használjuk.

Az időkezelés jól ismert nehézsége, hogy az egyedi jelenségek időbeliségére vonatkozó információ sokszor nem vagy nem elég pontosan áll rendelkezésre a gyakorlatban.

A kilencvenes évek zenéjével, XVI. századi franciaországi eseményekkel kapcsolatos adatok rögzítését akadályozhatja, hogy ilyen esetekben az időre való hivatkozás nem elég pontos ahhoz, az adatrendszerbe (dátum típusú adatként) fel tudjuk venni.

Mivel a társadalmi folyamatokra vonatkozó adatok esetében szinte mindig, de még a természeti folyamatokkal kapcsolatos adatok esetében is elég gyakran kijelenthetjük, hogy szükség van az időbeli adatok rögzítésére, valamint kezdenünk kell azzal a helyzettel, amikor nem elég pontos időbeli adatokkal rendelkezünk. A problémát úgy oldhatjuk meg, ha a *RADAR* sémában az időpontosságot (időpontatlanságot) önálló változóként kezeljük.

A hétköznapi életben előfordulhatnak olyan időhivatkozások is, amelyek a szokásos időperiodicitásunkhoz (perc, óra, nap, hét, hónap stb.) képest tartalmaznak olyan pontosító jelzéseket, amelyeket nem tudunk mérhetővé tenni (például a hét „közepén”, a hónap „végén”). A rendszerben szükség van az ilyen – ordinális jellegű – időpontosság kezelésére is.

A *RADAR* séma az individuumok időhöz való kötését úgy oldja meg, hogy önálló entitásként definiálja az eseményeket (EVENT). Természetesen lehetne másként is kezelni az időbeli kapcsolódást, de az esemény-alapú modellezésnek vannak olyan előnyei, amelyek meghaladják a megoldás hátrányait (elképzelhető, hogy más célok esetében nem érné meg az itt alkalmazott technikát választani). Ez a megoldás egyetlen helyen kezeli az időbeliséget, ami technikailag azért előnyös, mert az eltérő időpontosság problémájával csak egy helyen kell foglalkozni a rendszer egészen belül.

Az **event** táblában egyedi eseményeket vehetünk fel úgy, hogy megadjuk az esemény idejét (dátumként), majd egy új mezőben (`timeprecisionid`) jelezzük, hogy milyen pontosságú adatról van szó (ehhez létezik egy kötött

adatszótár (nap-pontos, hét-pontos, év-pontos stb.). Ha egy esemény idejének csak az évét tudjuk, akkor megadhatjuk azt, hogy év-pontos eseményről van szó stb. A rendszer ebből az adatból tudja, hogy a megadott datetime-értékből mennyi az értékes adat.

A hétköznapi életben előforduló ordinális jellegű időpontosítások (közepén, végén) megragadása miatt szükség van egy további mező (timeprecisionmodifierid) felvételére is. Ide is kötött adatszótárból lehet felvenni adatokat.

Az időpontosságot kifejező két mezővel biztosíthatjuk, hogy az események időtengelyhez kötését egységes módon valósíthassuk meg.

Az események rendezhetősége, áttekinthetősége miatt érdemes még felvenni egy olyan mezőt (eventtypeid), amelyben megadhatjuk az események típusát.

Ha az időt nem pusztán a jelenségek fizikai időbeli kötöttségén keresztül, hanem események definiálásán keresztül értelmezzük, akkor annak meglesz az az előnye is, hogy összetartozó individuális eseményeket könnyen összekapcsolhatunk, amire csak a fizikai időbeliség kezelésével nem lenne mód.

Ha egy-egy település valamely statisztikai adatát ugyanazon dátum megadásával kötjük az időhöz, vagy egy házasság esetében a feleség és férj esetében is ugyanazt a dátumot adjuk meg, akkor a két-két megegyező dátumérték összekapcsolja ugyan őket, de más individuumok esetében is felvehetjük ugyanazt a dátumértéket, és ez az időegyezés szemantikailag már nem fejez ki értelmes kapcsolatot az individuumok között.

Egy dátumváltozó segítségével nem tudunk szemantikai kapcsolatot teremteni individuumok között, hiszen ugyanabban az időpillanatban végtelen mennyiségű esemény történhet. Ezzel szemben az individuumok közti egyedi időbeli kapcsolatokat megragadhatjuk az eseményeken keresztül úgy, hogy felvesszünk egy egyedi eseményt megfelelő időhöz kötéssel, és az esemény azonosítójával hivatkozunk az időbeliségre.

Ha adott egyetemi évfolyam diákjainak adatait leíró táblában hozzá akarjuk rendelni az egyetemkezés időpontját, és beírjuk minden diák rekordjába a pontos dátumot, és ez a dátum két egyetem esetében megegyezik, akkor egyfelől csak a dátumadatok alapján nem tudjuk megkülönböztetni a két egyetem diákjait, másfelől ez a megoldás nem teljesen normalizált, hiszen ugyanazt az adatot sokszor írjuk be (egy egyetem esetében is), ami növeli a tévesztés esélyét.

Gyűjtemény

Az UNIVERZÁLÉ-PARTIKULÁRÉ fogalmi kettős mellett a **RADAR** modellben szükség van egy harmadik absztrakt entitástípus a **gyűjtemény** (COLLECTION) felvételére is. A hétköznapi életben ugyanis meglepően gyakran előfordul, hogy olyan fogalmakat használunk, amelyek nem felelnek meg a tiszta (szigorú) osztályozási logika elvárásainak, amikor a fogalmak terjedelmébe nem azonos minőségű példányok tartoznak. Ezek azok az esetek, amikor a fogalom terjedelmét nem lehet valamilyen intenzióval meghatározni.

Ez a kettősség megfeleltethető a matematikában ismert halmaz kontra osztály (típus) ellentétnek.

A magyar közigazgatásban a megyerendszer tiszta logikája megtörik azon a ponton, hogy a fővárost egyik megyéhez sem sorolták be, így az országot csak úgy lehet területi értelemben teljesen lefedni, hogy a 19 megye mellé beveszik még a fővárost is (ami nem megye). Van tehát az országnak egy olyan – nagyon gyakran használt – 20 elemű területi felosztása, amely 19 megyéből és a fővárosból (tehát két különböző minőségű fogalom példányaiból) áll.

A gyűjteményi fogalmak természetesen relációba állíthatók mind az általános, mind az individuális fogalmakkal, illetve saját magukkal (Bittner et al.)

LANG séma – a nyelvi adatréteg

A **RADAR** sémán kívül kezeljük a nyelvileg kifejezhető adatokat (a **LANG** sémában). Itt két fontos komponens van elkülönítve. Bármilyen természetes nyelven kifejezhető kifejezést képest befogadni ez a nyelvi modul, amely kezeli a természetes nyelvű kifejezések nyelvét, szófaját, a kifejezésen belül felbontható morféimákat. Közel egymillió természetes nyelvű kifejezés szerepel benne, ezek domináns része személynév, testületi név és földrajzi név, de 150 ezer általános fogalom is van benne. Az alábbi kép mutatja azt a találati listát, amit az 'eszperantó' kifejezésre ad vissza a rendszer.

construction	constrid	relationid	conceptid	concept	morphid	posid	lang	morph in construction
eszperantista	20543	69606	20520	eszperantista	20547	28	1	eszperantista (20547)
eszperantista	828146	69606	131241	eszperantista	295542	20	1	eszperantista (295542)
eszperantista író	354260	69606	113289	eszperantista író	295542	20	1	eszperantista (295542) író (36593)
eszperantó	20545	69606	20522	eszperantó	20549	20	1	eszperantó (20549)
eszperantó	20544	69606	20521	eszperantó	20548	28	1	eszperantó (20548)
Eszperantó	434208	100117	100115	földrajzi név	268738	29	1	eszperantó (268738)
Eszperantó Eszperantó	450741	100117	100115	földrajzi név	268738	29	1	eszperantó (268738) eszperantó (268738)
Eszperantó-forrás tufaüreg	444140	100117	100115	földrajzi név	268738	29	1	eszperantó (268738) - (137281) forrás (26416) tufaüreg (304283)
Eszperantó-hegy	286376	100117	100115	földrajzi név	268738	29	1	hegy (32632) - (137281) eszperantó (268738)
Eszperantó köz	471964	100117	100115	földrajzi név	268738	29	1	köz (46834) eszperantó (268738)
Eszperantó-mésztufabarlang	439237	100117	100115	földrajzi név	268738	29	1	eszperantó (268738) - (137281) mésztufabarlang (303722)
eszperantónyelvész	320611	69606	104191	eszperantónyelvész	282022	28	1	eszperantónyelvész (282022)
eszperantó nyelvész	354222	69606	113251	eszperantó nyelvész	20549	20	1	eszperantó (20549) nyelvész (60318)
Eszperantó park	482395	100117	100115	földrajzi név	268738	29	1	eszperantó (268738) park (64927)
Eszperantó pihenő	492123	100117	100115	földrajzi név	268738	29	1	eszperantó (268738) pihenő (66125)
Eszperantó sétány	474436	100117	100115	földrajzi név	268738	29	1	sétány (72038) eszperantó (268738)
eszperantó tanár	356520	69606	115503	eszperantó tanár	20548	28	1	eszperantó (20548) tanár (79138)
Eszperantó tér	450742	100117	100115	földrajzi név	268738	29	1	tér (80815) eszperantó (268738)
Eszperantó út	450743	100117	100115	földrajzi név	268738	29	1	út (84958) eszperantó (268738)
Eszperantó utca	450744	100117	100115	földrajzi név	268738	29	1	eszperantó (268738) utca (85088)
orosz és eszperantó nyelvtanár	357588	69606	116571	orosz és eszperantó nyelvtanár	62627	28	1	nyelvtanár (60372) orosz (62627) és (107492) eszperantó (20548)
orosz és eszperantó tankönyvíró	358011	69606	116994	orosz és eszperantó tankönyvíró	62628	20	1	eszperantó (20548) orosz (62628) és (107492) tankönyvíró (282590)

A természetes nyelvű kifejezések lehetnek tulajdonnevek vagy köznevek (az egyszerűség kedvéért utóbbiak közé lehet sorolni a mellékneveket és az igéket is). A tulajdonneveket individuumok megnevezésére, referálására használjuk, és ezt a funkciót (merev jelölés) filozófiailag megkülönböztetjük a köznevek szemantikát hordozó, jelentést kifejező képességétől. Az, hogy valakit *Kovács*-nak hívnak, nem fejez ki semmilyen jelentést, csak arra jó, hogy tudjuk vele hivatkozni (rámutatni) az adott személyre. Ellenben, ha azt mondjuk valakire, hogy *KOVÁCS*, akkor azzal már valamilyen jelentést, valamilyen világtudást is hozzákapsolunk. Az a kérdés, hogy ezt a jelentést hogyan kapcsolhatjuk a köznevekhez akkor, amikor tudjuk, hogy a különböző természetes nyelveken különböző kifejezésekkel fejezzük ki ugyanazt a jelentést (kutya – dog). Megoldás lehet erre az, hogy olyan entitást veszünk fel a nyelvi modulba, ami természetesnyelv-független. Ez kezeli a **concept** tábla, ami az ontológia CONCEPT metafogalma alá sorolható fogalmak kezelésére szolgál. Ha szerepel a **concept** táblában a KUTYA fogalma, akkor azt már összeköthetjük két köznévvvel, az angol *dog* és a magyar *kutya* kifejezésekkel a **concept_name** táblában.

Annak érdekében, hogy az általános fogalmainkat minél pontosabban értelmezni tudjuk, ontológiát kell építeni. Erre a célra szolgál a **concept_concept** tábla, hol ontológiai állításokat tehetünk úgy, hogy összekapcsolunk három fogalmat egymással. Ezek az állítások természetesnyelv-függetlenek, tehát ha megfelelő nyelvi kötés van hozzájuk, bármilyen természetes nyelvre „átfordíthatók”. Az ontológiát önmagában is, más célokra is lehet építeni, használni.

A nyelvi kifejezéseket több egymásra épülő rétegben építjük fel. A **lang.string** táblába veszünk fel minden olyan MORFÉMÁT, amely nyelvi kifejezés komponense lehet: szótó, toldalékolt szó, toldalék, írásjel, szám. Ez a tábla nem tartalmazhat szóközt és nagybetűt is csak ritkán. Az ide kerülő elemekről még azt sem tudhatjuk, hogy milyen nyelvhez tartoznak. Egy adott karaktersorozat csak egyszer szerepelhet a táblában, mert definiálva egy egyediséget előíró kényszer a táblán.

A *nap* vagy az *eleven* morféma egyszer szerepel függetlenül attól, hogy angol vagy magyar szóként értelmezzük-e majd a későbbiekben.

A string-ekre többször is hivatkozni lehet a következő (**lang.morph**) táblából. Egyrészt itt kell megadni, hogy egy adott string (*stringid*) milyen nyelvhez tartozik (*languageid*), másrészt meg kell adni a string nyelvtani funkcióját, szófaját (*posid*). Ezek együttese már meghatároz egy morfémát, ami már nyelvi kifejezéseket alkotóeleme lehet. Ebben a táblában meg lehet még adni a morféma nyelvi viselkedését jellemző adatot. Természetesen a morfémák nagyobb része szó, és csak kis részük toldalék. A morfémákból képzett szavakat, több szóból álló kifejezéseket, vagyis a nyelvi konstrukciókat írjuk be a harmadik táblába (**lang.construction**), amely tábla csak arra szolgál, hogy azonosítót adjunk a konstrukciónak, illetve jelezhessük azt, hogy a konstrukció milyen nyelvű (*languageid*). Ez utóbbi jelzésre azért van szükség, mert előfordulhat olyan konstrukció, amelynek a nyelve nem egyezik a konstrukciót alkotó morfémák, szavak nyelvével.

A magyar nyelvben mondhatjuk, hogy valaki *Shakespeare-kutató*, és ezt a kifejezést nyilván magyarként értelmezhetjük, de a kifejezés egyik összetevője (*Shakespeare*) ettől még nem lesz magyar, hanem angol marad.

Azt, hogy a konstrukciók hogyan épülnek fel a morfémákból, egy következő táblában adhatjuk meg (**lang.construction_morph**), ahol a konstrukció azonosítóját (*constructionid*) kapcsoljuk össze a komponens morfémák azonosítóival (*morphid*), illetve megadjuk a morfémák sorrendjét (*rank*).

Mivel a konstrukciók gyakran összetettek, vagyis több morféma összekapcsolásával jönnek létre, ezért van szükség egy új kapcsolótáblára a morfémák és konstrukciók között.

A regisztrált munkanélküliség konstrukciót (940237) két morféma összekapcsolásával kapjuk meg: *regisztrált* (366407) + *munkanélküliség* (58091).

A nyelvi konstrukciók szemantikai tartalmukat azáltal nyerhetik el, ha összekötjük őket a fogalmi építkezés alapelemeivel, a fogalmakkal (amelyek természetesnyelv-függetlenek). A szemantikai tartalommal rendelkező fogalmakat tároljuk a fogalom táblában (**lang.concept**), és ezeket kötjük össze a konstrukciókkal egy kapcsolótáblában (**lang.concept_construction**). Ha egy adott nyelven két kifejezést egy fogalomhoz kapcsolunk, ez azt jelenti, hogy a két kifejezésnek azonos a jelentése, de eltér az alakjuk és nyilván más a nyelvi viselkedésük is.

Feltételezhetjük, hogy amikor egy angol és egy magyar ember a kutyájára gondol, akkor – durván – ugyanazt a jelentéstartalmat építi fel a fejében. Ezt a közös minőséget ragadhatjuk meg fogalmi szinten a 'kutyaság' fogalmával. Amikor nyelvi megnyilatkozásokat bocsátanak ki a 'kutyaságról', akkor eltérő konstrukciókat használnak (*dog*, illetve *kutya*), de tartalmilag ugyanarról beszélnek. Ugyanezt mondhatjuk el akkor is, ha egy nyelven eltérő szóalakokat használnak ugyanazon jelentéstartalommal (magyar nyelven: *eb* és *kutya*).

A szemantikai tudást a fogalmi struktúrán vagyis a fogalmak egymáshoz való viszonyán keresztül fejezzük ki (**lang.concept_concept**). Szükség van arra is, hogy fogalmainkat olykor konstansokkal kapcsoljuk össze (**lang.concept_literal**).

Név vs. megnevezés

Ahhoz, hogy a gondolati tartalmainkat közölni tudjuk egymással, nyelvi megnyilatkozásokat formálunk. A dolgokat leegyszerűsítve azt mondhatjuk, hogy neveket adunk a létezőknek, hogy ezekkel tudjunk hivatkozni rájuk. A világban léteznek dolgok, és ha ezekhez neveket rendelünk, akkor utóbbiakkal hivatkozhatunk az előbbiekre. A nevekkal való hivatkozás azonban felveti azt a kérdés, hogy hogyan kell értelmeznünk a magukat neveket, univerzáléknak vagy partikuláréknak kell-e minősíteniük őket. Amikor egy emberre utalunk a *Kovács János* név segítségével, akkor egy individuumhoz akarunk kötni egy nyelvi megnyilatkozást, és ez a tény azt sugallhatná, hogy ez a (tulajdon)név is rendelkezik valamilyen individuális minőséggel. Ez azonban nincs így, hiszen a névhez nem tudunk semmilyen (tér- és időbeli) megkötést rendelni. A nevek nem lehetnek individuumok, mert nem tudjuk az időben megkötni őket. A nevek ugyanis egy időben több helyen is létezhetnek. Egy személyre tudjuk azt mondani, hogy éppen most hol tartózkodik, de a személy nevére nem mondhatjuk azt, hogy "most itt vagy ott van". Pont ebből fakad a homonimitás problémája is. Amikor egy névvel azonosítani akarunk egy embert, azzal szembesülhetünk, hogy ugyanazt a nevet kiadhatjuk több személy számára is. Akkor viszont adott a kérdés: hogyan lehet a névttereket azonosításra használni, ha a tulajdonnevek – önmagukban – nem alkalmasak erre.

A kérdést feltehetjük kicsit másként is: hogyan lehet a (tulajdon)neveket úgy egyedivé tenni, hogy alkalmassá váljanak az individuális névhordozók azonosítására. Nos, a megoldás az lehet, hogy az azonosításra nem a (tulajdon)neveket, hanem **megnevezéseket** (megnevezési eseményeket vagy megnevezési állapotokat) használunk. Ez a megoldás Kripke tulajdonnév-értelmezésére vezethető vissza (Kripke 2010). Kripke szerint egy tulajdonnév referenciáját nem valamilyen hozzá társított leírás alapján határozhatjuk meg, hanem a névhez kapcsolódó, történeti láncba szerveződött megnevezések révén:

"... mikor kimondom a „Rubik Ernő” nevet, annak referenciája az a személy, akit a megnyilatkozásomhoz kapcsolódó névhasználatok történeti lánc (vagy inkább hálójá) meghatároz. A lánc kezdeténél a név bevezető használata áll, és ezt követik a név továbbadó használatai, amelyek révén a név végül eljutott hozzám. Nagyjából így foglalhatjuk össze a tulajdonnevek történetilánc-referenciaelméletét." (Zvolenszky 2015)

A (tulajdon)névadás és -használat tehát valamely megnevezői közösség praxisában kialakuló merev jelölő kialakítását és tartós fennmaradását jelenti. Az új elem itt a megnevezői és névhasználati gyakorlat mozzanata. "Rámutatások", névadási gesztusok révén egy közösség gyakorlatában egy névhordozóhoz tapad egy (tulajdon)név, és ez a megnevezési aktusok összefüggő hálózatán keresztül fenntartja a név és a referált dolog közti állandó kapcsolatot. Nem a (tulajdon)név azonosít tehát, hanem a névhasználó közösség megnevezési gyakorlata, a megnevezés(lánc), ami viszont mindig egyedi események láncolatát jelenti. Ez azért fontos, mert az egyedi névhordozókat az individuális megnevezéseken keresztül azonosíthatjuk (ezzel eltűnik a nevek és névhordozók közti filozófiai különbség, hiszen a megnevezések mint események és a névhordozók egyaránt individuális entitások). Természetesen ugyanazt a (tulajdon)nevet más és más megnevezéseken keresztül különböző névhordozókra lehet alkalmazni – akár egy, akár több időben. Ez az értelmezés így kezelni képes azt a homonimitási jelenséget, amiről korábban említést tettünk. A megnevezést nem egyszeri aktusként, hanem egy folyamatos megnevezési eseménysorozatként (vagy épp egy megnevezési állapotként) érdemes felfogni. Ez a megoldás azért képes megoldani a konkrét névhordozók azonosítását, mert a megnevezést (mint eseményt vagy állapotot) már hozzáköthetjük valahogyan a tér és idő koordinátaíhoz. A megnevezésnek ez a térben és időben

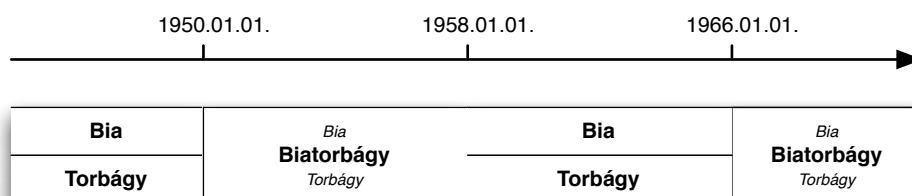
történő lokalizálása úgy is értelmezhető, hogy ez teremti meg azt a kontextust az adott (tulajdon)név számára, amely révén utóbbi alkalmassá válhat az egyedi azonosításra.

Eddig még nem volt arról szó, de most már szükséges utalni rá, hogy nevekből és megnevezésekből többféle is lehetséges, vagyis mind a neveket, mind a megnevezéseket tipizálhatjuk. A személynevek esetében beszélhetünk becenevekről vagy álnevekről, hivatalos, születési, felvett, írói stb. nevekről. Ezek mind eltérő megnevezési gyakorlatokhoz és közösségekhez kapcsolódhatnak, és persze az is előfordulhat, hogy többféle névhasználói közösség ugyanazt a nevet tartja fent valakivel kapcsolatban. A megnevezési gyakorlat mint kontextus felfogás tudja azt a jelenséget is kezelni, amikor két megnevezés esetében a névalak ugyan teljesen megegyezik, de az névhasználói kontextuson belül más típusú nevekről van szó.¹

A hosszabb elméleti felvezetés után térjünk rá arra a kérdésre, mi következik mindebből a névterek építésének konkrét gyakorlatára vonatkozóan. Az eddig bemutatott elméleti megfontolásoknak több fontos következménye van.

Abból a tényből, hogy a névtérnek két, egyenrangú komponense a név és a névhordozó (a jelölő és a jelölet), az következik, hogy a névtérben mindkét komponenst kezelni kell tudni. Praktikusán ez annyit jelent, hogy mind a névhordozókat, mind a neveket azonosítani kell valamilyen módon, és – talán ez a legfontosabb – harmadik komponensként fel kell venni a névtérbe a köztük levő kapcsolatot is. Annál is inkább szükség van erre, mert szinte mindegyik névhordozótípus (a személyek, a testületek, a földrajzi helyek stb.) névhasználati gyakorlatában nagyon gyakori a szinonimitás jelensége, vagyis az, hogy egy névhordozóhoz több név is tartozik. Ezt úgy kell kezelni, hogy mind a névhordozót, mind a hozzá kapcsolható összes nevet önmagában azonosítjuk,² és a megnevezési esemény (gyakorlat) leképezéseként ezek összekapcsolását is annyiszor elvégezzük, ahány névadási esemény (praxis) előfordul az adott névhordozó esetében (hogy erre miért van szükség, rövidesen bemutatjuk Biatorbágy példáján keresztül).

A megnevezést mint kontextust elég csak egyértelműen azonosítani, nem kell alaposabban feltárni (bár szükség esetén lehetséges). Azzal, hogy úgy kapcsoljuk össze a névhordozó (mondjuk a személy) egyedi azonosítóját (personid) és a hozzárendelt névelem egyedi azonosítóját (nameid), hogy ezt az összerendelést is azonosítjuk, magát a megnevezési kontextust ragadhatjuk meg. Nézzük meg Biatorbány példáját! Bia és Torbágy községeket 1958-ban egyesítették, az új településnek 1950 és 1958 között Biatorbágy volt a neve. 1958-ban újra szétvált a két település, és hivatalosan is visszakapták a Bia és Torbágy neveket. Végül 1966-ban újra egyesítették őket, és innentől kezdve megint a Biatorbágy nevet viseli a település. A két korábban önálló település, Bia és Torbágy az idők során kétszer is hivatalosan viselte a saját nevét, kétszer pedig a Biatorbágy hivatalos névvel jelölték az összevont települést. Ez három nevet, de hat megnevezést (névhasználati gyakorlatot) jelent. A nevek szintjén a Bia, Torbágy és Biatorbágy nevek keletkezésük óta léteztek, de a megnevezések/névhasználatok szintjén nem mindig volt így. Ez az eset arra is példát mutat, hogy amikor a megnevezésekkel hozzákapszolunk egy névhordozóhoz egy adott nevet, akkor megtehetjük, hogy – több megnevezés felvételével – ugyanazt a nevet többször is odakötjük ugyanahhoz névhordozóhoz.



Természetesen az emberek számára a nevek az elsődlegesek, és számukra a fenti nevek (Bia, Torbágy, Biatorbágy) többszörös használata nem kell, hogy elkülönüljön egymástól, de precíz, gépi és szaktudományi feldolgozás számára ezeket a különbségeket kezelni kell.

Ontológia

A (formális) ontológia a világról szóló tudásunkat reprezentáló fogalmak strukturált rendszere. Az ontológia azt értelmezi, hogy milyen dolgok léteznek a világban, ezeknek milyen tulajdonságaik vannak, illetve ezek milyen viszonyban vannak egymással. Lényege szerint univerzális állításokból áll, bár tartalmazhat individuális állításokat is.

¹ Egy település hivatalos nevének megváltoztatása után a régi névhez két – olykor radikálisan eltérő, egymással versengő – névhasználati gyakorlat tartozhat: a hivatalos gyakorlat akadályozhatja, tilthatja a régi név használatát, míg a "laikus", hétköznapi praxis támogathatja azt.

² Az azonosítás itt mesterséges azonosító hozzárendelését jelenti.

A filozófiában nincs mindenki által elfogadott, közös ontológia, különböző ontológiai elköteleződések (ontological commitments) mentén nagyon eltérő ontológiákat lehet építeni.

Mindig fontos kérdés az ontológia granularitása, hogy ti. milyen pontos fogalmakat kell alkotni valamely tudásterület reprezentálásához.

Amíg sok tudásterület számára elegendő lehet 6-8 alapszint felvenni az ontológiába, addig egy festékipari ontológia számára sokkal finomabb felbontású színfogalmakra lehet szükség. A **RADARBAN** ezer színfogalom van.

A hétköznapi világban az emberi testrészekre hivatkozni elég párszáz fogalom, míg az orvosi praxisban nagyságrenddel több fogalom szükséges.

Az ontológiák úgy építjük, hogy első lépésben fogalmakat teremtünk arra nézve, hogy

- mi létezik a világban (EMBER, KUTYA, ÁLLAT, BOSZORKÁNY, ARANY, VÍZ, ÉTER), majd
- az újonnan felvett fogalmakról elmondjuk, hogy az általuk leírt létezőknek milyen tulajdonságaik vannak (SWAN | HAS-COLOR | WHITE, WATER | HAS-PROPERTY | LIQUID), illetve
- a fogalmak hogyan viszonyulnak a többi fogalmakhoz (DOG | IS-KIND-OF | ANIMAL, GRANDFATHER | IS-SAME-AS | FATHER(FATHER)).

A formális ontológia valamilyen formális nyelven fejezi ki az ontológiai állításokat annak érdekében, hogy számítógéppel lehessen kezelni azokat.

A szemantikus web kezdeményezés az univerzális állításokat tripletekben (subject – predicate – object) írja le, de ettől természetesen el lehet térni. Egy triplet kétargumentumú predikátumnak feleltethető meg.

A **RADAR** sémában vannak három-, öt-, hatelemű állítások (rekordok). Az általános fogalmak közti kapcsolatokat tripletekben írjuk le, az individuumokra vonatkozó egyedi állításokban több argumentumot használunk (utóbbiakat nem igazán szokás ontológia fogalma alá tartozó állításoknak minősíteni).

Értelmezéstámogatás

A **RADAR** séma predikátumai, szerkezete

A körülöttünk levő földrajzi térben zajló konkrét eseményekre, egyedi individuumokra vonatkozó világtudásunkat a fent bemutatott entitások alapján rögzíthetjük adatbázisba. Egyelőre ötféle állítástípust vettünk fel, amit lehet bővíteni, bár a modell annyira absztrakt, hogy sok minden képes befogadni.

Mindegyik állítástípusban/táblában közös az első négy elem.

- Mindig meg kell adnunk az időbeli kötést az eseményeken keresztül. Az egyszerűség kedvéért két mezőben lehet kifejezni az individuumok időbeliségét úgy, hogy megadjuk – ha ismerjük – az folyamatok (összetett események) kezdő- és záróeseményét. Sok jelenség esetében csak egy eseményt érdemes kezelni (például születés), akkor a záróeseményt nem vesszük fel. Gyakran előfordul, hogy nem ismerjük az eseményt, ilyenkor maradhat üresen az adott mező.
- Az időbeli lokalizálást lehetővé tevő két eseménymező után mindig meg kell adni az individuum azonosítóját.
- Minden állítástípusban/táblában meg kell adni azt az általános fogalmat, ami kifejezi a rekord szemantikai tartalmát (például a statisztikai adatok felvitelekor magát az adatváltozó reprezentáló fogalmat).

Az – egyelőre – ötféle állítástípus abban tér el lényegileg egymástól, hogy mire hivatkozik az ötödik mezőben.

A **radar.individual_type** táblában van módunk arra, hogy az individuumra vonatkozóan tetszőleges klasszifikációs állítást tegyünk. Ehhez a nyelvi modul általános fogalmait használhatjuk. Az ötödik oszlopban egy általános – tipizálást, klasszifikációt jelentő - fogalomra hivatkozunk (**lang.concept**).

A **radar.individual_name** táblában az individuumok megnevezéseit kötjük a nyelvi modul tulajdonnév gyűjteményéhez. Az ötödik oszlopban a – merev jelölést, referálást elvégző – tulajdonnevekre hivatkozunk (**lang.name**).

A **radar.individual_geometry** táblában a földrajzi entitások földrajzi térhez való kötését fejezhetjük ki. Az ötödik oszlopban a földrajzi kötést jelentő geometriára hivatkozunk (**radar.geometry**).

A **radar.individual_value** táblában az individuum olyan tulajdonságát adhatjuk meg, amelyet nem „névterjesztettünk”. Ez az egyetlen pont a radar sémán belül (**value**), ahová szabad karaktereket lehet beírni, és nincs külső – idegenkulcs – hivatkozási kényszer előírva.

A **radar.individual_individual** táblában két individuum kapcsolatát lehet kezelni, az ötödik oszlopban egy másik individuumra kell rámutatni (**radar.individual**).

A **radar.individual_value** és a **radar.individual_name** táblák esetében felvettünk még egy hatodik oszlopot – eltérő megfontolások miatt.

Előfordulhat, hogy az individuumoknak úgy kell valamilyen értéket adni, hogy az az érték nem önmagában köthető hozzá, hanem egy másik individuummal való kapcsolatán keresztül. Példát hozhatunk erre a választási adatok köréből. Amikor egy pártra adott szavazatokat akarjuk kezelni, akkor, ha felvesszünk időváltozókat, felvesszünk fogalmi kapcsolatot (pártokra leadott szavazatok), felvesszük a települést, amely területre a szavazatokat aggregálták, és megadjuk az összesített szavazatszámokat, ebből még nem tudjuk, melyik pártról van szó, ezért fel kell venni egy újabb mezőt, ahová már beírhatjuk a párt individuális azonosítóját. Ez a megoldás nem kezeli azt a helyzetet, amikor három individuumot kell ilyen módon összekapcsolni, de ilyen esettel még nem találkozunk a gyakorlatban.

Elvileg meg lehetne oldani tetszőleges számú egyedi entitás összekapcsolódó jellemzését (közös értékkel való ellátását), ha felvennénk a rendszerbe egyedi individuumok egyedi gyűjteményeit (például egy gyűjteményben kapcsolnánk össze egy területet, egy pártot és egy személyt), és erre a – gyűjtemény típusú – individuumra hivatkoznánk a **radar.individual_value** táblában, de ezt megnehezítené a lekérdezést.

Az individuumok megnevezését kezelő táblába azért kellett felvenni egy hatodik mezőt, hogy az individuumokhoz tudjuk kötni azokat a megnevezési aktusokat, amikor elnevezték őket valaminek (**namingid**). A jelenlegi rendszeren az elnevezéseket is individuumokként kezeljük.

Történeti nézetek

Az adatmodell megengedi, hogy minden individuális állításhoz időbeli kötést lehessen tenni. Amilyen mértékben ki tudjuk tölteni az időbeli változásokat jelző adatmezőket, olyan mértékben válik a rendszer képessé arra, hogy történeti metszetben mutasson korábbi állapotokat.

A települések neveinek, határainak változásai 1990-ig visszamenőleg bent vannak a rendszerben. Ez teszi lehetővé, hogy bármilyen napra meg lehessen mondani, hogy aznap milyen települések, milyen néven, milyen határvonalakkal léteztek éppen.

Az utcanévek változásai is rendelkezésre állnak a napjainktól a 2000-es évek közepéig visszamenőleg.

Hasznos függvények

A normalizált adatbázisokban nézettáblák segítségével lehet emberek számára is áttekinthetővé tenni az adatbázis tartalmát. Hasonló célokat szolgálnak azok a függvények, amelyek paraméterek megadásával fontos összefüggéseket, adatkészleteket adnak vissza.

- **radar.polygon_90(date,typeid)**
1990 utáni dátumot és egy poligonpartíciót megadva visszaadja az adott napon érvényes helyek listáját, geometriáját
települések 1998.02.14-én: `select * from radar.polygon_90('1998-02-14', 80409)`
- **radar.all_universal_part_of(conceptid)**
egy fogalmat megadva visszaadja az összes, részeként alá tartozó fogalmat
emberi testrészek: `select * from radar.all_universal_part_of(134865)`
- **radar.subordinated_concept_of(conceptid)**
egy fogalmat megadva visszaadja az összes, generatív
KSH TSTAR adatváltozók: `radar.subordinated_concept_of(121287)`

Fontos relációk

A **RADAR** adatrendsze számára fontos a különböző entitás típusok és a köztük felvehető relációk egyértelmű és pontos meghatározása (Bittner et al.)

generatív alárendeltje	faja, altípusa, alosztálya, is-a, is-a-type-of	100118
generatív fölérendeltje	neme, osztálya	inverse-of(100118)
individuális partitív alárendeltje	része, is-part-of	69702
individuális partitív fölérendeltje	egésze	inverse(69702)
univerzális partitív alárendeltje	része, is-part-of	134860
univerzális partitív fölérendeltje	egésze	inverse(134860)

-ből áll	is-consists-of	134695
instanciája	példánya, instance-of	100117
inverze	inverse-of	118233
eleme	is-element-of	

rész (124034,109470)

Külső és belső névterek

Amikor a *RADAR* sémában adott fogalom alá tartozó individuumokat egyértelműen megkülönböztetjük egymástól, illetve más fogalmak alá sorolható individuumoktól, és az egész sémán belül egységesen hivatkozunk ezekre az individuumokra, akkor voltaképp résznévttereket kezelünk a rendszer egészén belül. Ezek belső névtterekre mutató kapcsolatok. De természetesen az individuumokat kapcsolhatjuk külső névtterekhez is.

Belső tulajdonnévtterek

A *RADAR* sémában több névtér is van, és bármikor tovább bővíthető a névtterek listája. Jelenleg a legfontosabb névtterek a következők:

- személynévtér (~350 ezer)
- területi névtér (~80 ezer)
- földrajzi névtér (~100 ezer)
- szimbolikus névtterek
 - devizanévtér (~200)
 - természetes nyelv (~8 ezer)

A belső névtér funkciót szolgálja ki a séma nagyobbik része: az **individual**, **individual_type**, **individual_name**, **individual_geometry**, **individual_individual**, **event**, **geometry** táblák szinte csak az individuumok azonosíthatóságát szolgálják. Az **individual_value** táblában vannak azok az adatok, amelyek tartalmas állításokat közvetítenek a különböző individuumokról, a köztük levő kapcsolatokról.

A Radar séma tábláiban – egy kivételével – 1-3 millió rekord van, míg az **individual_value** táblában 80 millió.

t

A közeljövőben várható, hogy a földrajzi névtteret további elemekkel bővítjük.

residential area, park, commercial,

iskolák, polgármesteri hivatal, művelődési ház, gyógyszertár, posta

Külső tulajdonnévtterek

A földrajzi adatok egy része, különösen a geometriai adatok az OpenStreetMap (OSM) rendszerből kerülnek be a *RADAR* sémába. Az OSM egyfajta névtérként is értelmezhető. A *RADARBA* bekerülő OSM-elemek geometriai jellemzőin túl tároljuk az elemek OSM-azonosítóit is. A személynévtér esetében sok külső névtteres kapcsolat is szerepel (például Wikipedia, FIAF, ISNI), bár ezek léte egyetemi környezetben nem annyira jelentős.

Részontológiák

Kezdetleges állapotban vannak ugyan, de léteznek szakontológia-kezdemenyezések.

- földrajzi fogalmak
- foglalkozások
- színek
- emberi testrészek
- rokonsági fogalmak
- családtámogatási törvény fogalmai

RADAR további adatok

A *RADAR* sémában tárolt alapadatok további adatrétegeket lehet elemzésekhez igénybe venni. Vannak olyan származtatott adatrétegek, amelyeket az alapadatokból lehet kiszámolni. Ezeket bárki, bármikor kiszámolhatja természetesen, de akkor értelme lehet önálló adatrétegekként értelmezni (és eltárolni, felkínálni) őket, ha a számolások időigényesek, hiszen az elemzések számára könnyebben igénybe vehetők. Vannak olyan adatrétegek is, amelyek szótári, lexikai adatokat tartalmaznak. Ezek növelhetik a későbbi elemzések szemantikai képességeit.

Származtatott adatrétegek

A különböző forrásokból származó partikuláris adatok biztosítják a **RADAR** legnagyobb adatrészét. Ezekből többféle értelmekben lehet származtatott adatrétegeket képezni.

Földrajzi adatok aggregálása

A földrajzi entitásokhoz kapcsolt adatok (társadalom-, gazdaságstatisztikai, természetföldrajzi adatok stb.) adott területi felbontásra vonatkoznak, amelyeket poligonszerű geometriával reprezentálunk. A földrajzi adatok felbontása, granularitása, elemi aggregáltsági szintje szerint a geoentitások sok esetben hierarchikus rendszerbe állíthatók, és ilyenkor van értelme a magasabb aggregáltsági szintekre összegezni az adatokat, önálló adatrétegeket képezni.

A területi adatok értékelésekor fontos, hogy azok milyen időpontokban, illetve milyen aggregáltsági szinten érhetők el.

Poligonpartíciók

A poligonos földrajzi helytípusok között vannak, amelyek példányai azonos intenzióval rendelkeznek, és földrajzi kiterjedésüket tekintve teljesítik a jepd-elvárast, vagyis a kölcsönösen kizáró és együttesen kimerítő feltételt (jointly exhaustive and pairwise disjoint). Ezen geotípusba tartozó geoentitások számossága időben változott/változik, és ezek egy részére már létezik időbeli változáskövetés (a települési szinten, illetve afelett), így megfelelő függvény segítségével a történelmi adatokhoz is hozzá lehet férni.

választókerület	~100-200	jepd-teljes	86015
település	~3100-3200	jepd-teljes	80409
járás	~150-200	nem jepd-teljes	37259
kistérség	~150-200	nem jepd-teljes	43605
megye	19	nem jepd-teljes	55855
kistáj	230	jepd-teljes	107804
középtáj	35	jepd-teljes	46984
nagyta	6	jepd-teljes	107876
vegetációtáj	109	jepd-teljes	134913
régió	5-10	jepd-teljes	69145
főváros kerület	22-23	jepd-teljes	103972
PIR-körzet	~3000	jepd-teljes	108469
távhió körzet	54	jepd-teljes	134487
egyházmegye (katolikus)	12	~ jepd-teljes	15821
espereskerület	~100	~ jepd-teljes	20391

Gyűjtemények

Vannak olyan földrajzi helyek, amelyek – praktikus megfontolások miatt – úgy fogunk össze egy-egy gyűjteménybe, hogy azok elemei nem feltétlenül rendelkeznek azonos intenzióval és általában nem is jepd-teljesek.

város	~250	nem jepd-teljes	86504
megyei jogú város	25	nem jepd-teljes	103973
agglomerációs térség	23	nem jepd-teljes	134881
településrész		nem jepd-teljes	86538
község	~2700-2800	nem jepd-teljes	47222
nagyközség	~100-120	nem jepd-teljes	58626
város + fővárosi kerület	~270-280	nem jepd-teljes	134879
település + fővárosi kerület	3100-3160	jepd-teljes	135320
járás + főváros	~150-200	jepd-teljes	135321
kistérség + főváros	~150-180	jepd-teljes	135322
megye + főváros	20	jepd-teljes	135323
zsákfalú (számolt adat)	~450-500	nem jepd-teljes	134215

Az elemi adatszinten rendelkezésre álló adatköröket aggregálni lehet magasabb területi szintre, amivel természetesen veszítünk az adatok granularitásában, viszont összevethetővé tehetünk adatköröket egymással.

A KSH TSTAR adatai település szinten állnak rendelkezésre, de ezek aggregálva vannak járási/kistérségi, megyei, országos szintre, így azonnal térképre tehetők vagy elemzésbe emelhetők át.

Számolt centralitásmutatók

A közlekedési hálózatokat gráfként értelmezve ezek struktúrájára többféle mutató kiszámolható, és bár az ilyen műveletek bármikor (és bárki által) elvégezhetők, az olyan esetekben, amikor nem várható gyakori és jelentős változás a struktúrákban, érdemes mindent kiszámolni és azonnali használatra elérhetővé tenni.

Az OSM utcahálózati adatok alapján többféle centralitásmutató érhető el a Radarban, amelyek kiszámolása a nagyobb városok esetében sok időt és komolyabb gépi kapacitást kíván.

Trendmutatók

A fogalmak sok esetben valamilyen idősorba rendezhető adatváltozókra utalnak, és sok esetben hosszabb időszakra vonatkozó individuális idősorok állnak rendelkezésre. Ezekre az változókra regressziós egyenest állítva, vagyis előre kiszámolva az adatsor meredekség és tengelymetszet (slope és intercept) értékeit, új, származtatott adatként állíthatunk elő. Ezeket nézettáblákba írva az új adatréteggként érhetők el.

A KSH TSTAR adatrendszer azon változói esetében, amelyek legalább egy évtizednyi idősorral rendelkeznek, lineáris regresszióval kiszámolva az adatsorok trendmutatóit, ezek ugyanolyan módon automatikusan elérhetők, mint maguk az alapadatok.

Lexikonok

- Értelmező szótár, I-VII. Akadémiai Kiadó
- Czuczor-Fogarasi szótár
- Ki kicsoda

Hivatkozások

- Kripke 2010
- Zvolenszky 2015
- Bittner et al.