

A TEZAUROSZ MINT „KISVILÁG”

Ontológiai helyzetkép a tezaurszokról az OntoClean szemlélet alapján, 1. rész

Ungváry Rudolf, ungvary@hungary.com, OSZK

MEO-PROJEKT

A Guarino terminológiája szerinti rigid generikus relációk alkotják a tezaurszokon belül az erős kapcsolatokat. Ha a rigid relációk nem teljesek, akkor a tezaursz szétesik több különböző tezaurszrészre, azaz a tezaursz nem lesz kompakt. A generikus relációk biztosítják a társadalmi hálózatokban korábban már leírt szigorú kapcsoltságot az információkereső nyelvekben.

A szemantikus vagy tematikus szerepeket (Parsons, Jackendoff) betöltő nem-rigid relációk ezzel szemben nem igényelnek teljességet, legfeljebb arra van szükség, hogy egy csomóponthoz legalább néhány partitív, oksági és egyéb rokonsági reláció kapcsolódjék. Ha ezeket a gyenge kapcsolatokat megszakítjuk, a tezaurszuk kompaktsága nem szűnik meg, nem esnek szét. A gyenge relációk biztosítják, hogy egy-egy tezaurszon belül viszonylag kevés — max. 5–7 ugrással — el lehessen jutni a tezaursz egyik csomópontjától (lexikai egységétől) a másikhoz. Ezáltal az információkereső-nyelv az ún. „kisvilágok” viselkedésével jellemezhető (Barabási). (Az ilyen világokban 5–6 lépésen belül mindenki mindenkit „ismer”). Ezekben érvényes a csomópontok közötti hat ugrásnyi átlagos távolság.

Az ilyen mérvű kapcsoltság lehetővé teszi egyrészt a rendszer stabilitását, másrészt a rendszeren belüli optimális navigációt a legtalálhatóbb kifejezés kiválasztásakor. Az információkereső nyelvek struktúrája tehát a Barabási és munkatársai által felfedezett csomópont alapú skálafüggetlen hálózatokéhoz hasonlít.

A fenti tulajdonságok vizsgálata az olyan méretű tezaurszokon, mint az OSZK-tezaursz/Köztaursz ideálisan elvégezhető, mivel lexikai egységeinek száma már lehetővé teszi a tetszés szerinti kapcsolóugrások s ezzel a belső csatlotság modellezését.

A számítástechnikai fejlődés három szakasza

A számítástechnikai fejlődés első szakaszában a gépi eszközök kialakításán volt a hangsúly. 1975 után, amikor már tömegméretekben is megfelelő számítógépek álltak rendelkezésre, értékelődött fel a szoftverfejlesztés jelentősége. A kereskedelmi forgalomban megjelenő szoftverek gyors terjedése érlelte meg a számítástechnikai fejlődés harmadik, a múlt század utolsó évtizedével kezdődő és máig tartó szakaszát, amelyben a fejlődés egyik súlypontját a másodlagos információik kezelése képviseli. E másodlagos információk az információ kereshető tárolását és keresését biztosító tárgyi kifejezések és osztályozási jelzetek, tulajdonnevek, idők, megjelenési, terjedelmi és egyéb azonosító adatok. Pusztán programozási szempontból ezek az információk ún. kulcsokat (hozzáférési pontokat) képviselnek az elsődleges információkhoz (a dokumentumrekordokhoz, tényadatokhoz).

A fejlődés eme harmadik szakaszában az informatikai szóhasználatban a másodlagos információ megnevezésére a metaadat kifejezése terjedt el.

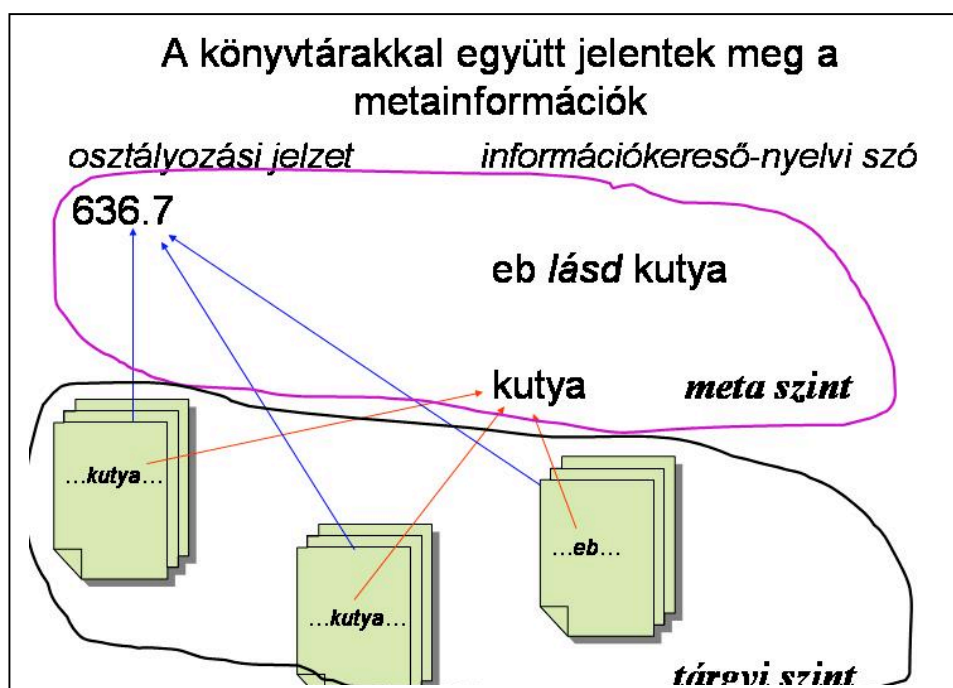
Hangsúlyok a számítástechnikai fejlődésben:

hardver (a „vas”) 1950 után

szoftver (a programok) 1975 után

másodlagos vagy metainformációk 1990 után

Noha a felhasználót mindig csak az elsődleges adatok (maghatározott mű, tanulmány, eljárás, gazdasági, műszaki, statisztikai tényadatok stb.) érdeklik, ezeket a metaadatok nélkül nem tudná megtalálni. A metaadatok a múlt század végéig csak a könyvtári világ kitüntetett figyelmét élvezték. A könyvtárakban ezeket az adatokat a dokumentumleírásokat tartalmazó katalóguscédulák rendezésére használták, és ezért besorolási adatoknak nevezték. Másfél évszázad szerves fejlődés során alakultak ki az ún. egységesített besorolási adatok (újabbban: rekordhozzáférési pontok) szabványai. Az 1990 utáni fordulatot jelzi, hogy immár elsősorban az informatikusok kezdeményezésére megjelentek a legfontosabbnak tekintett alapmetaadatok szabványai, mintegy jelezve, hogy a metaadatok a számítástechnikai innováció kitüntetett tárgyai lettek.



A metaadatokat valójában az értékelte föl, hogy az adatbázisokban tárolt, ill. a weben elérhető elsődleges vagy tárgyi információk mennyisége rendkívül megnőtt, és megfelelő metaadat-rendszerek hiányában ez a mennyiség egyre kevésbé kezelhető. Metaadatok, főleg pedig megfelelő metaadat-rendszerek (hagyományosan: tárgyszórendszerek, teauruszok, osztályozási rendszerek — összefoglalóan információkereső-nyelvi szavak szótárai) nélkül ugyanis az elsődleges vagy tárgyi információk elérhetetlenek. Valójában tehát minden adat- és ismeretbázis-kezelő tevékenység alapja a megfelelő metaadatrendszer. A metaadatok mintegy „megfogják”, „rögzítik” a tárgyi információkat, afféle rendszerstabilizáló szerepük van.

Metaadat és fogalom

A metaadatok valójában — ha az osztályozási jelzetektől eltekintünk — természetes nyelven kifejezett szavak és kifejezések. Végző soron a közbeszéd keretei között maguk a beszélt nyelv szavai is — önmagukban — metaadatoknak tekinthetők. A szavak és kifejezések mondatok formájában megvalósított összerendezésével fogalmazhatók meg az elsődleges információk, miközben maguk a mondatalkotó szavak és kifejezések egyben az információkereső nyelv lexikai egységeit is képviselik. Mindezek a szavak és kifejezések ugyanakkor fogalmak megnevezései, összességük fogalmi rendszert alkot. A forma (a nyelv) szempontjából információkereső nyelvekről, a tartalom (a fogalmak) szempontjából pedig rendező rendszerekről van szó. Ebben az értelemben a természetes nyelven megnevezett fogalmak rendszere (az értelmileg strukturált szókinccs) is egyetlen, hatalmas, és tulajdonságai tekintetében még meglehetősen kevésbé ismert rendező rendszer (s egyben információkereső nyelv).

A metaadatok fontossága azért értékelődött fel az informatikában, mert az internet általános elterjedése miatt az információkeresés a weben tömegmértévé vált. A laikus végfelhasználókat a világhálón végzett kereséshez az általuk beszélt (köz- vagy természetes) nyelven kell támogatni (ma szemantikus webnek nevezik az internetnek ezt a — egyelőre csak jövőbeli — változatát). Mindez olyan problémákat vet fel, amely miatt a nyelvészek és logikai szakértők figyelme is a metaadatok felé fordult, noha e két szakterületen nem metaadatokról vagy információkereső-nyelvi szavakról hanem természetes nyelvű kifejezésekről, elsősorban fogalmak megnevezéseiről beszélnek. (Például a „kutya” vagy a „Julius cézár” a tárolás szempontjából ismérvi vagy (rekord)hozzáférési pont (kulcs), a keresés szempontjából információnyelvi keresőszó, az informatikában metaadat, a nyelvészet nézőpontjából kifejezés, a fogalomelmélet nézőpontjából általános név, ill. egyedi név, a logika nézőpontjából pedig egy argumentumú predikátor vagy reláció, ill. individuumnév.)

Metaadat-hálózatok

Az információkereső-nyelvi vagy fogalmi rendező rendszerek hálózatként is viselkednek. A teljes nyelv szókinccse is afféle virtuális hálózatot alkot, mivel a szavak és kifejezések között a jelentésük (a fogalmak tartalma) alapján egy rendkívül komplex — és ma még lényegében feltáratlan — értelmi összefüggésrendszer húzódik meg. A 19. század közepétől az osztályozási rendszerek, a 20. század közepétől pedig a teauruszok formájában kialakultak ugyanakkor a metaadatok strukturált szótárai, melyekben az értelmi összefüggések hálózata rögzített, explicit formában jelenik meg.¹

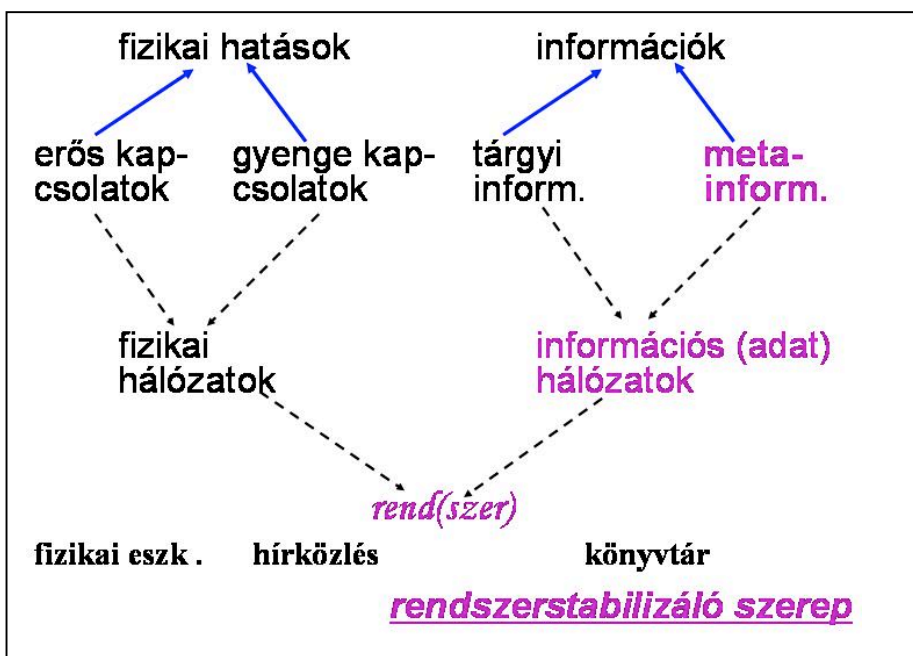
¹ A legkorábbi metaadat-rendszerek hierarchikus osztályozási rendszerek formájában már a Kr. e-i 3. évezred elején megjelentek, de csak a 19. sz. közepétől alakultak ki a nagy — 50.000–100.000 lexikai egységet tartalmazó — rendező rendszerek.

E nyelvi-fogalmi hálózatokban az értelmi összefüggések egyébként áttekinthetetlen hálózatát leegyszerűsítve, szabványosított relációtípusokba összevonva tüntetik fel. Ábrázolásuk egyik formája a címkézett, irányított gráf, mely egyben a hálózat térképe is. Ennek azért nagy a jelentősége, mert szemben a természetes nyelvvel (melynek térképével ma még nem rendelkezünk), ezek a metaadat-rendszerek a térképeik alapján vizsgálhatók. Noha a természetes nyelv komplexitásához képest rendkívül leegyszerűsített nyelvi összefüggésrendszert alkotnak, elemzésükből némi következtetések már levonhatók a beszélt nyelv szavai és kifejezései közötti szemantikai összefüggésekről.

Hálózati szempontból a helyzet a műszaki életben megfigyelhető kettősséghez hasonlít. A fizikai hálózatokba összefogott műszaki eszközök között ún. erős kapcsolatok állnak fenn. Ilyen hálózatot alkot például egy gépjármű, melyben a tartó és erőátviteli elemek erős fizikai kölcsönhatásokon keresztül kapcsolódnak egymáshoz.

Ahhoz azonban, hogy ez az erős kapcsolatokon alapuló fizikai rendszer kezelhető legyen, szükség van az irányítást biztosító, gyenge kölcsönhatásokon alapuló hírközlő hálózatra. Ilyen hálózatot alkot például a gépjármű járművillamos és elektronikus, újabban számítógéppel is támogatott irányító rendszere, melyben pl. gyengeáramú kölcsönhatások biztosítják az irányítási adatok átvitelét.

Az információs hálózatokban — mint amilyen egy könyvtáré, adatbázisé, ismeretbázisé, vagy mesterséges intelligencián alapuló rendszeré — az „erős” kapcsolatok a felhasználót érdeklő elsődleges vagy tárgyi információk között állnak fenn (egy adott dokumentumon belül vagy dokumentumok, illetve tárgyi információkat leíró rekordok között). Adott dokumentum például a hálózatokról szól, és e tény virtuálisan összekapcsolja mindazokkal a dokumentumokkal, melyeknek ugyancsak ez a témája. A dokumentumokat és általában az elsődleges információkat a kereshető célú információtárolás érdekében valóságosan a metainformációk kapcsolják össze, és éppen általuk jönnek létre az információs vagy adathálózatok. A hálózatokról szóló dokumentumokat például ellátják a „hálózat” ismérvvvel, vagy lehetővé teszik, hogy a dokumentumok szövegében szabadon lehessen keresni, többek között a „hálózat” keresőszóval (ebben az esetben a dokumentum szövegén belüli „hálózat” szó egyszerre az elsődleges és a másodlagos (meta-)információ szava).



Metainformációs hálózatot alkotnak az osztályozási rendszerek, a tárgyszójegyzékek és teauruszok, pontosabban a bennük összefogott információkereső-nyelvi szavak, de valójában maga a természetes, beszélt nyelv összes szava és kifejezése (rajtuk keresztül pedig az összes lehetséges fogalom az összes lehetséges világban) is.

Ontológiák

A metainformációs hálózatok terén az elmúlt másfél évtizedben egyik alapvető fejlemény az ontológiák megjelenése. Az ontológiák a generikus hierarchián alapuló fogalmi struktúrák, melyek az előbb említett osztályozási rendszereken (és elsősorban a teauruszokon) belül mindig is léteztek², két alapvető különbséggel.

Az ontológiáknak olyan, szabványokba is foglalt logikai leíró nyelvük van, mely a fogalmak legfontosabb tulajdonságain alapuló automatizált kezelésüket lehetővé teszi. Eme ontológiai leírónyelvek alapján elsősorban ismeretbázisok, de webes keresőrendszerek számára is megfelelő programrendszerek készíthetők.

Az így értelmezett ontológia olyan fogalmi sémát (konceptualizációt) képvisel, mely lefordítható a számítógépes rendszer tudásreprezentációs eszközeinek nyelvére. Ennek megfelelően az ontológiáról mint mérnöki produktumokról beszélünk, amelyet az egyes információrendszerekhez és tárgyterületekhez hozzá lehet majd rendelni. Az ontológia biztosítja, hogy az ismeretalapú rendszer által tárolt ismeretbázis mindig összhangban legyen a (valós világbeli) fogalmi sémával.

Fogalomtulajdonságok elemzése

A FOGALMAK TULAJDONSÁGAI (METATULAJDONSÁGOK)

A másik, az ontológiák megjelenését szükségszerűen követő fejlemény, hogy születő félben vannak olyan logikai fogalomtulajdonság-elemző eljárások, melyek lehetővé teszik az ontológiákba foglalt generikus fogalmi rendszerek következetes, formális elveken alapuló szerkesztését. Az egyik legismertebb ilyen fogalomtulajdonság-elemző és egyben ontológiaszerkezet-leíró módszer a Guarino nevéhez fűződő OntoClean.

Az OntoClean lényege, hogy a fogalmak terjedelmébe eső előfordulások (egyedek, terjedelmi elemek) tulajdonságai alapján fogalmak tulajdonságai állapíthatók meg (melyeket az OntoClean terminológiája szerint metatulajdonságnak neveznek). Más szóval a fogalmakat jellemzi, hogy egyedeire vonatkozóan bizonyos térbeli és időbeli tulajdonságok, ill. függési viszonyok fennállnak-e vagy sem. A metatulajdonságok alapján szabályok és fogalomtipológia alakítható ki, mely az ontológiák helyes szerkezetének kialakítását támogatja és logikai programok készíthetők a szerkezet formális helyességének ellenőrzésére.

Mivel az ontológiák nem egyebek, mint első rendű logikai leírással ellátott (és ezáltal logikailag gépi eszközökkel kezelhető) generikus fogalmi hierarchiák, ezért az OntoClean módszertan minden természetes nyelven alapuló hierarchikus osztályozási rendszer, továbbá a teauruszok generikus hálójának formálisan helyes szerkezetének a kialakítására és ellenőrzésére is használható.

² A tudományoknak már az ókorban megkísérelt osztályozásai, Arisztotelész (Kr. e. 4. sz.), Kant stb. fogalmi kategóriarendszerei, Porphüriosz (3. sz.) Lullus (1272), Wilkins (1668), stb. fogalomosztályozási rendszerei, Roget teaurusza (1852), a 19. század közepétől kezdve pedig a Tizedes és a később az Egyetemes Tizedes osztályozási rendszer (a maga mára kb. 130.000 osztályozó fogalmával) az ontológiák őseinek tekinthetők.

A metatulajdonságokkal kapcsolatban két probléma adódik. Az egyik alapvető és a fogalmak értelmezésével függ össze, a másik terminológiai jellegű.

a) Függés az ontológiai elkötelezettségtől.

A metatulajdonság megállapítása függ az ún. ontológiai elkötelezettségtől: attól, hogy hol, például a közgondolkodásban (egy általános, köznyelvi ontológiában), vagy melyik speciális információs rendszerben, ontológiában használják az adott fogalmat, melynek tulajdonságait meg kell határozni. Más szóval alkalmazásonként eltérhet a fogalmak értelmezéséből következő ún. ontológiai elkötelezettség. Ennek következtében ugyanaz a fogalomnak más-más metatulajdonságot állapítanak meg a különböző rendszerekben. Ez mindaddig, ameddig az ontológiákat elszigetelten használják, nem okoz problémát, mert a lényeg az, hogy adott rendszeren belül formálisan egyértelműen alkalmazzák a szerkesztési szabályokat.

b) Terjedelmi megalapozás, fogalmi használat.

A terminológiai nehézséget az okozza, hogy előfordulások (a fogalom terjedelmi elemeinek) tulajdonságai alapján kell a metatulajdonságot (tehát a fogalom tulajdonságát) megnevezni. Az előfordulás tulajdonsága (azonosítható, függ, csak ha teljes akkor létezik, részei vannak stb.) viszonylag könnyen megnevezhető, de ha ezt az elnevezést a fogalom tulajdonságára alkalmazzuk, az a benyomás kerekedik, mintha a fogalomnak lenne a megnevezett tulajdonsága, ami nem igaz (attól például, hogy egy fogalom előfordulásai térben és időben egymástól elkülönítve nem azonosíthatók, a fogalomra nem következik, hogy az se azonosítható). Mindkét problémát szemléltetni fogjuk a későbbiekben.

AZ EGYED ÖNAZONOSSÁGA ÉLETTARTAMÁNAK IDEJÉN (RIGIDITÁS)

A legfontosabb metatulajdonság azon alapul, hogy adott fogalom terjedelmébe eső egyedet minden időben reprezentálja-e a fogalom vagy sem. Ezt a metatulajdonságot rigiditásnak („rigidity”) nevezik.

A fogalom akkor rigid (+R), ha a terjedelmébe eső minden előfordulást minden lehetséges világban (minden lehetséges időben és helyen, ahol az egyed létezik) reprezentálja. Például:

- bárki, aki ember, létének első pillanatától az utolsó pillanatáig ember marad;
- az a konkrét embercsoport, amelyet magyar nemzetnek hívunk, fennállása egész ideje alatt magyar nemzet, nem lehetett korábban vagy később valami más (a hangsúly a konkrét előforduláson van: a magyarok egy korábbi összessége nem feltétlenül számított még magyar nemzetnek a szó mai értelmében);
- aki angol származású, az életében soha nem szűnik meg angol származásúnak lenni
- az étel akkor is étel, ha nem eszik meg és megromlik (ha viszont kidobják a szemétkébe, ételként már megsemmisül a szétesés miatt).

További példák: valami, folyamat, tulajdonság, lény, anyag, tárgy tulajdonság, élőlény, ember, csoport, foglalkozás, tanítás, gép, vér, élelem, téglá, homok

A fogalom akkor antirigid (~R), ha a terjedelmébe eső minden előfordulást nem minden lehetséges világban (nem minden lehetséges időben és helyen, ahol az egyed létezik) reprezentálja. Például:

- tanárnak nem születik valaki, és az is lehetséges, hogy valaki abbahagyja a tanári foglalkozást;
- magyar állampolgárrá válhat valaki, és el is veszítheti;
- a csecsemő egy idő múlva felnőtt ember, később aggstyán (miközben mindvégig ugyanaz a valaki);

- a hernyó a következő évben báb, azt követően lepke (holott mindvégig ugyanaz az egyed);
- a munkanap valamelyik naptári előfordulása nem feltétlenül munkanap egy másik évben;
- a piros alma fejlődésének korábbi szakaszában nem piros színű.

További példák: szerep, ágens, aktor, rész, egész, ok, okozat, eszköz, rendeltetés, tulajdonos, tulajdon, tanár, színész.

A fogalom akkor semirigid ($\neg R$), ha a terjedelmébe eső előfordulások egy részét rigiden, másik részét antirigiden reprezentálja. Például:

- az építőanyagok egy része rigiden (pl. téglá), másik része nem rigiden (homok) az;
- a piros színű dolgok egy része mindig piros (pl. vér), más részük nem mindig (pl. piros arc);
- angol állampolgár lehet valaki születésétől a haláláig, de meg is szerezheti meg el is veszítheti azt;
- mindazok a dolgok közül, melyek jellemzője, hogy angol, az egyik rigiden az (pl. angol származás), más részük antiirigiden az (pl. angol állampolgár);
- a nőtlenek egy része mindig nőtlen marad (tehát rigiden nőtlen), más részük megházasodik;
- az ennivalók egy része fennállása egész ideje alatt ennivaló marad (abból a célból kszül, pl. a torta), más része akkor lesz ennivaló, ha erre a célra felhasználják (pl. az alma).

A rigid fogalmakat nevezik megtévesztő módon „természetesnek” is (pl. Guarino [1992], Roles3) (ilyen alapon a nem rigid és a semirigid fogalmak „mesterségesnek” minősülnek). Később látni fogjuk, hogy az OntoClean ontológiaszerkezetében a rigid fogalmak a rendszer gerincét alkotják, ezért a „természetes” helyett kézenfekvőbb volna a „törzsfogalom” elnevezés. Az eltérő ontológiai elkötelezettség jól szemléltethető az építőanyag és az étel fogalmával. Aki szerint az építőanyag az építményekbe beépített anyag, annak ez a fogalom rigid. Az építőipari terminológia szerint azonban építőanyag az, ami építés céljából felhasználható; egy építőipari ontológiában tehát ez a fogalom semirigid. Az OntoClean megalkotója, Guarini szerint az élelem antirigid, mert lehet, hogy nem eszik meg a kenyeret (e felfogás szerint a „kidobott élelem” nem az élelem egyik fajtája). Más felfogás szerint azért antirigid, mert megszűnik élelem lenni, ha megromlik (e felfogás szerint az alma csak akkor élelem, ha élelem céljából szedték le a fáról, és a „romlott élelem” nem az élelem egyik fajtája). Az élelmiszertechnológiai terminológia szerint az élelem szemirigid, mert van, ami rigiden (töltött pulyka) és van, ami antirigiden (alma) élelem. Ez az egyetlen metatulajdonság, melynek neve valóban a fogalom tulajdonságát nevezi meg: egy fogalom, ha rigid, valóban megváltoztathatatlanul reprezentálja az egyedet, függetlenül annak időbeli és térbeli állapotától.

AZ EGYED ÖNazonossága TÉRBELI ÉS IDŐBELI Tulajdonságai Alapján (AZONossági FELTÉTEL)

Az azonossági metatulajdonság (az OntoClean terminológiája szerint: azonossági feltétel, „identity condition”) azon alapul, hogy adott fogalom terjedelmébe eső egyedeket jellemzi-e olyan közös térbeli és időbeli tulajdonság, amelynek alapján az egyedek egyértelműen azonosíthatók (+I), illetve nem jellemzi ilyen tulajdonság (-I). Ebből a szempontból nevezhetők a fogalmak identikus (+I), ill. antiidentikus (-I) fogalomnak.

Az első (+I) esetben az azonossági feltétel alapján kimondható vagy a fogalom által reprezentált egyedek azonossága (+ME), vagy a fogalom által reprezentált egyedek különbsége (-ME). Az OntoClean terminológiában csak az első esetnek adtak speciális nevet: mereológiai (összetettségi) azonosságnak nevezik. A második esetet (-ME) értelemszerűen nevezhetnénk individuális [egyedi] azonosságnak. Példa a kétfajta azonossági feltételre:

- a fizikai tárgyakat (darabokat, alkatrészeket, gépeket, kövületeket) egymástól a térben, illetve az időben elfoglalt helyük különbözteti meg (egyedi azonosság);
- az élőlényeket (vírusokat, gombákat, növényeket, állatokat, embereket) a genetikus kódjuk különbözteti meg (egyedi azonosság);
- a speciálisabb fizikai tárgyakat (kavics, tengely, eszterga, foraminifera-ház stb.) a speciális fizikai tulajdonságaik (egyedi azonosság);
- a személyeket a személyi adatai, az egyéneket az egyéni tulajdonságai, a személyiségeket a társadalmi helyzete, foglalkozások szerinti személyeket (férj, anya, agglegény, magányos kutya, tudós, mérnök, tanár, tanár, tanuló, jogi ügynök, színész, politikus, asztalos, szakmunkás, segéd stb.) a speciális cél, rendeltetés, eszköz stb. (mereológiai azonosság);
- az anyagokat az összetételük, szerkezetük azonos volta határozza meg (anyag, vas, acél, homok, kolloid, tej, levegő, pára, selejt stb.) (merológiai azonosság);
- a csoportok, összességek fogalmainak egyedeit (konkrét népet, nemzetet, élő szervek és szervezetek társulását/rendszerét, társadalmi összességet/rendszert és társulást) az egyes konkrét csoportok elemeinek (tagjainak) azonossága határozza meg (merológiai azonosság);
- a tulajdonságot, állapotot, formát időlegesen (antirigiden vagy semirigiden) hordozó egyedeket (folyamatos gyártás, anyagi haszon, piros alma, tuberkulotikus csont, munkanap, csecsemő, felnőtt ember, imágó, hernyó stb.) a közös tulajdonság, forma (mereológiai azonosság);
- a speciális folyamatokat, tevékenységeket, cselekvéseket az aktoiraik, okaik/okozataik, céljaik/rendeltetésük stb. (az áramlást az adott közeg és történés ideje, a lélegzést az adott légzőszerv és a történés ideje, az evést az emberi szervezet és a történés ideje, az ugatást a kutyafélék és a történés ideje, a gondolkodást az elme működése és a történés ideje, a betartást a személy és a történés ideje stb.) (egyedi azonosság);
- a konkrét és elvont tudományos dolgok fogalmait által reprezentált dolog (téridő, energia, hő, információ, gravitáció, atom, kvark, univerzum, tejút, csillag, hegység, talaj, fagy, éghajlat, eső, réteg, város, ösztön, indulat, én, lélek, tudat, fogalom, szabály, elv, demokrácia, jogegyenlőség, pénz, nyelv, reláció, függvényjel stb.) azonosítási feltétele az adott szakterület ismerete alapján állapítható meg.

A második (-I) esetben nincs a fogalom terjedelmébe eső egyedekre megállapítható azonossági feltétel. Ezek a legáltalánosabb fogalmak, továbbá a hordozott tulajdonságok és hordozott formák/alakzatok. Például:

- az elvont fogalmi kategóriák („category”), melyek három csoportra oszlanak (az OntoClean terminológiájában a kategória fogalmán belül nincs meg az alábbi megkülönböztetés):
 - tartalmi kategóriák (valami, folyamat, tulajdonság, állapot, lény stb.);
 - formai kategóriák (dolog, tárgy [elvont értelemben], jelenség, előfordulás, történés, eset, tett stb.);
 - szerepkategóriák, az OntoClean terminológiájában formális szerepek („formal role”)(szerep, ágens, résztvevő, ügynök, aktor, beneficiens, faj, nem, rész, egész, eszköz, rendeltetés, kiindulás, cél, következmény, eredmény, alap, feltétel, stb.);
- az ismeretterületek (tudományok, tanok, elméletek, tevékenységi és ügykörök (ügyek), ágak, ágazatok, iparok, technikák és technológiák, mesterségek, foglalkozások [foglalkozási körök]); az osztályozáselméletben ezeket szakkategóriáknak nevezik; a szakkategóriák terjedelmi elemei azok a fogalmak, melyek az adott ismeretterületen használatosak, és csak

rajtuk keresztül az általuk reprezentált egyedek (az OntoClean terminológiájában külön nem térnek ki erre a típusra)³;

- a jellemzők („attribution”), melyek két csoportra oszlanak (az OntoClean terminológiájában nincsenek megkülönböztetve):
 - hordozott állapotok és tulajdonságok (folyamatos, anyagi, cseppfolyós, forró, szín, piros, élő, tuberkulotikus, emberi, társadalmi, jogtalan, munka-, autós, nőtlen, csecsemős, felnőtt, hernyószerű stb., továbbá szabadság, rettenet, térszerű, időbeli stb.);
 - a hordozott alakzatok és formák (alak, forma, háromszög, kör, hiperbola, gömb, nyeregfelület, fejlődési forma stb.)

Ha egy fogalomnak nincs azonossági metatulajdonsága (-I), attól a terjedelmébe eső egyedei még felismerhetők, de ezeket az egyedeket (legalábbis első lépésben) már csak az adott fogalom ismertetőjegyei — és nem a fogalom terjedelmébe eső előfordulások tulajdonságai — alapján különíthetjük el, ill. azonosíthatjuk. Például a „lény” fogalmához nem kapcsolódik azonossági metatulajdonság. Nincs olyan konkrét térbeli vagy időbeli tulajdonság, melynek alapján az egyes konkrét lények (tehát a „lény” fogalmának a szintjén az angyalok, a kísértetek, az élőlények stb.) azonosíthatók volnának. Ennek ellenére a „lény” fogalmának ismertetőjegyei alapján meg tudjuk mondani, hogy adott esetben angyalról, szellemről, Istenről, istenekről vagy élőlényről van szó. A „folyamat” fogalmának egyedei (az áramlások, a lélegzések, a gondolkodás stb.) esetében sincs közös azonossági feltétel (külön az áramlások, a lélegzések és a gondolkodások esetei azonosíthatók), de a folyamat fogalmának birtokában (az ismertetőjegyei alapján) azt mindegyik esetben felismerjük, hogy folyamatokról van szó.

Az azonossági feltételek megállapítása többnyire nehéz, logikai felkészültséget igénylő művelet. Az ontológiák készítésekor adott fogalom metatulajdonságának megállapításához (s ezzel a fogalom tipizálásához) azonban elég annak mérlegelése, hogy a fogalom által reprezentált egyedekhez kapcsolódik-e azonossági feltétel vagy sem.

Az azonossági metatulajdonsággal rendelkező fogalmak (+I) — éppen azért, hogy egyedei egyértelműen megkülönböztethetők, illetve azonosíthatók — közvetlenül meghatározzák az egyedek (előfordulások) univerzumát. Más szóval minden egyednek tartoznia kell legalább egy azonossági metatulajdonsággal rendelkező fogalomhoz. Például egy adott kavics rigidén a „kavics”, antirigidén adott esetben a „lövedék” fogalmának terjedelmi eleme.

Az OntoClean terminológiája szerint ezeket a (+I) fogalmakat — megtévesztő módon — osztályozó fogalmaknak („sortals”) nevezik. Az elnevezés azért önkényes, mert minden más, azonossági metatulajdonsággal nem rendelkező fogalommal (-I) is (a „sortals” fogalmainál lényegesen nagyobb) osztályokba fogjuk össze a terjedelmükbe tartozó egyedeket (ezek összefoglaló neve ugyancsak megtévesztően „non sortal”), csak éppen nem az előbbieken tárgyalt, egyértelműen megfogalmazható azonossági feltételek, azaz nem az egyedek konkrét tulajdonságai, hanem a (-I) fogalmak ismertetőjegyei alapján.

Például egy konkrét kavics a tárgy „non sortal” fogalmának terjedelmébe tartozik, mivel a „kavics” fogalmának egyik ismertetőjegye kényszerűen tartalmazza a „tárgy” fogalmának ismertetőjegyét: „különálló létezőként érzékelhető anyagi jelenség”.

³ Az egyes fogalmak ugyan rendelkezhetnek a terjedelmükbe eső előfordulások alapján azonossági feltétellel (+I) de a szakkategóriák terjedelmi előfordulásait alkotó az egyes fogalmakhoz mint olyanokhoz nem tudunk olyan tulajdonságokat találni, melyek alapján formálisan elkülöníthetők lennének. A fogalmak mint olyanok csak a tartalmuk ismertetőjegyei által különböznek egymástól, de az ismertetőjegyek nem a fogalmak tulajdonságai. A „kutya” fogalmának ismertetőjegye, hogy ugat, tulajdonsága viszont ugyanaz, mint minden más fogalomnak is: elvont.

Fogalomelméleti nézőpontból helyesebb elsődleges (elsődlegesen reprezentáló, „sortal”) és másodlagos (másodlagosan [meta-]reprezentáló) fogalmakról beszélni (ezek lennének a „non sortal” fogalmak). Az elsődleges fogalmakat egy másik szokásos kettősség alapján nevezhetjük tárgyi fogalmaknak, a másodlagosakat pedig metafogalmaknak (LÁBJ: ezeket az elnevezéseket az OntoClean terminológiájában nem ismerik, feltehetően kialakítói és a vele manapság foglalkozók nem az információkeresés és a fogalomelmélet szakterületének művelői közül kerülnek ki). Fontos felismerni, hogy az utóbbiakkal alapvetően a tárgyi („sortal”) fogalmakat osztályozzuk, és csak rajtuk keresztül egyedeket. Ezért nem kapcsolódhat hozzájuk azonossági feltétel, hiszen az egyes fogalmaknak mint olyanoknak nincs a tárgyalt értelemben azonossági feltétele, csak egyedeknek.

Az egyes fogalmaknak ismertetőjegyeik vannak, fogalmi szinten ezek felelnek meg az általuk reprezentált előfordulások tulajdonságainak. Az azonossági metatulajdonsággal nem rendelkező fogalmak (-I) olyan általánosak, a gondolati elvontság olyan magas hierarchiaszintjén helyezkednek el, hogy valójában már nem az egyedek, hanem a generikusan alájuk rendelt speciális fogalmak osztályozására valók. Gondoljunk csak a folyamat, a tevékenység vagy a tulajdonság (mindegyik -I) fogalmaira, hogy belássuk átfogó rendező-osztályozó („sortal”) szerepük jellegét. Például a „tevékenység” osztályozó fogalma az összes tevékenységet jelentő azonossági metatulajdonsággal rendelkező fogalomnak (pl. az „ugatás”, a „futás” „gondolkodás” fogalmainak).

A metafogalmak soha sem tárgyai a konkrét gondolkodásnak, a legkritkább esetben irányul rájuk elemző figyelem, mivel elvontságuk következtében többnyire csak finomító, modifikáló szerepük van a gondolkodásban és a mondatalkotásban. A leginkább ez a helyzet a „valami” fogalmával, mely minden elképzelhető fogalom legáltalánosabb generikus fölérendeltje (kivéve a „semmi” fogalmát). Ezekre a fogalmakra csak a mesterségesintelligencia-kutatások és az ontológiák megjelenésének jóvoltából irányult a tudomány történetében a figyelem (a ritka, korábbi kivételektől eltekintve). Nehezen megérthető szerepük feltárásának újabban tapasztalható igyekezete valójában forradalmi változás a logika, a nyelvészet és a filozófia történetében.

AZ EGYED ELŐFORDULÁSAINAK FÜGGÉSE MÁS EGYEDELŐFORDULÁSOKTÓL (FÜGGÉSI FELTÉTEL)

Egy A fogalom **függ** a B fogalomtól, ha A egy előfordulása csak akkor létezhet, ha létezik B egy előfordulása is, a triviális kivételekkel. Ezeket a kivételeket Guarino sem sorolja fel, csak példákat hoz. Az így értelmezett függést Guarino „külső” függésnek nevezi.

A külső függés meghatározásához azt az elvet kell szem előtt tartani, hogy nem lehet minden függő. Ezért a következőket érdemes szem előtt tartani (ez sem teljes felsorolása a kivételeknek):

- a része reláció nem fejez ki külső függést, tehát sem az egész a részeitől, sem a rész az egésztől nem függ;
- valami nem függ a tulajdonságaitól, anyagától, alakjától stb.;
- csak a fogalmi függést veszünk figyelembe, olyat nem amely a világ törvényszerűségeiből ered (pl. a metatulajdonság szempontjából a panda nem függ a bambusztól).

Azt mondjuk, hogy egy fogalom **függő**, ha függ egy másik fogalomtól. Jelölés: **+D**, ill. **-D** ha nem. Példa: az ember nem függő fogalom, de a tanár függő (**+D**) (függ a tanítástól, a tanított tárgytól és a tanítványtól).

Az ember ugyan függ a neveltetéstől, a környezetétől, de ezek nem fogalmi — nem a fogalmak ismertetőjegyeiből következő — függések.

A „függőség” esetében valójában fogalmak közötti különféle oksági relációkról van szó. De itt is érvényes, hogy ezek a relációk nem fogalmak, hanem előfordulások között állnak fenn. Nem a „tanár” *fogalma*, hanem a konkrét tanár van kapcsolatban a konkrét tanítással stb. Valójában nem arról van szó, hogy a tanár *függene* a tanítástól, hanem arról, hogy a tanár rendeltetése a tanítás. A formális logikai megközelítés ezt a finomságot nem veszi figyelembe, és minden oksági összefüggést függésnek tekint.

A függés az egyik legértékesebb metatulajdonság, dacára annak, hogy meghatározása nem teljes; elsődleges szerepe a fogalmak osztályozásánál lesz, ahol is a szerepek elkülönítését szolgálja. A szerepek ismertetésénél részletesebben elemezzük.

ÖSSZEGEZÉS

A metatulajdonságok definícióiból, illetve egészen általános ontológiai elkötelezettségekből szabályok szűrhetők le, amelyek segítségével a generikus reláció alkalmazása vizsgálható, különböző hibák felfedhetők

Az OntoClean metodológia a metatulajdonságok függvényében osztályozza a fogalmakat. Az osztályozás:

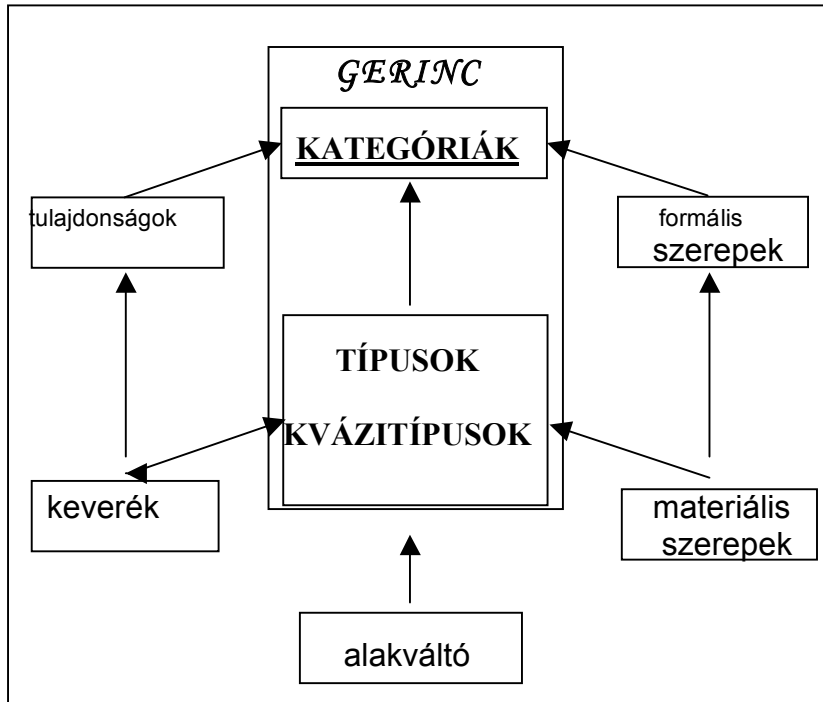
+O	+I	+R	±D	típus (type)	„sortal”
		+R	±D	kvázi-típus (quasi-type)	
		~R	+D	materiális szerep (material role)	
		~R	-D	alakváltó (phased sortal)	
		~R	±D	keverék (mixin)	
	-I	+R	±D	kategória	
		~R	+D	formális szerep (formal role)	
		~R	-D	tulajdonság (attribution)	
		~R	±D		

1. táblázat. Az OntoClean fogalomfajtái.

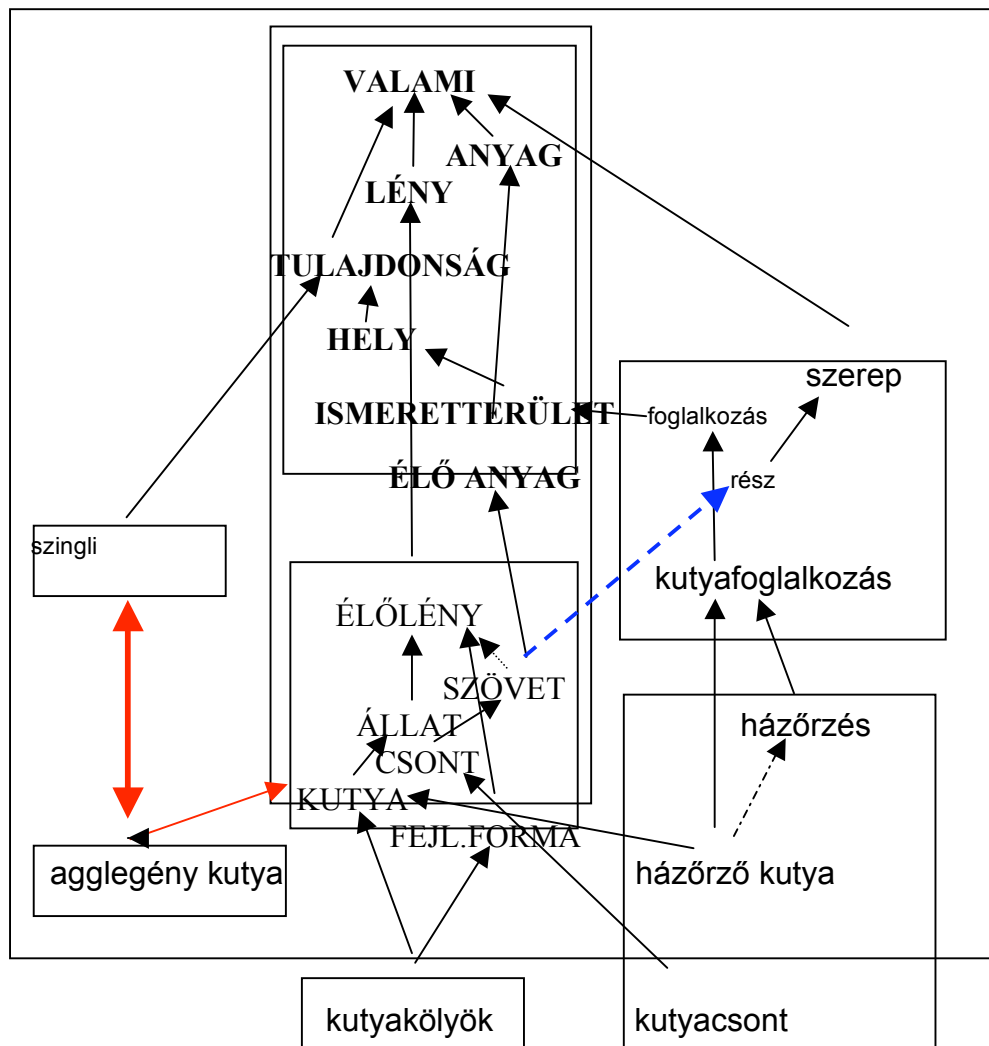
A +O és ~R vagy ~R kompozíciók azért nem szerepelnek, mert a definíciók szerint összeférhetetlenek.

A fenti osztályozás alapján egy ontológiaszerkezet állítható fel, az ábrában a nyilak a kötelező öröklődést mutatják:

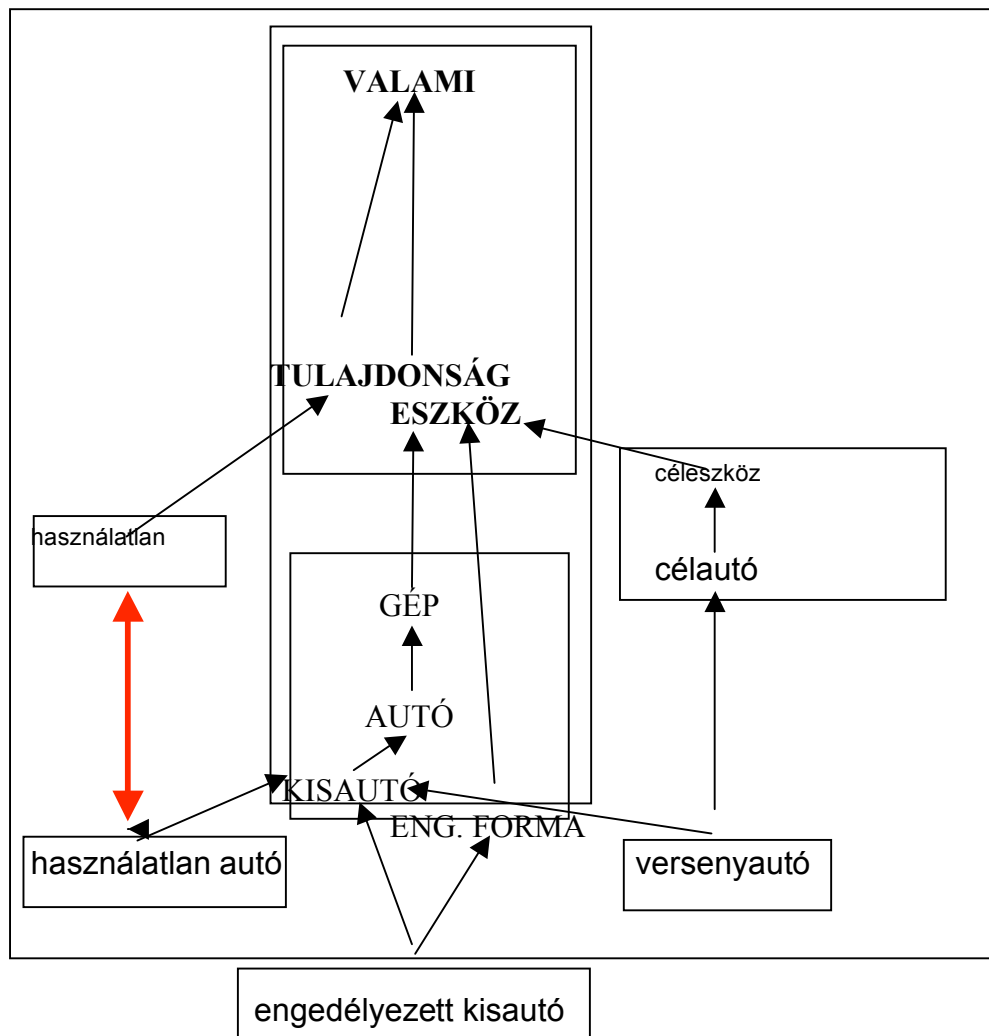
A GUARINOI ONTOCLEAN SZERKEZET AZ ALÁBBI:



1. behelyettesítési példa a szerkezetbe:



2. behelyettesítési példa a szerkezetbe:



Figyeljük meg e három fajta metatulajdonságait:

+I	~R	-D	alakváltó
+I	-R	+D	keverék
+I	-R	-D	keverék
-I	~R	-D	tulajdonság
-I	-R	+D	tulajdonság
-I	-R	-D	tulajdonság

Példákkal:

+I	~R	-D	alakváltó	báb, piros alma, kör?, kocka?, lányom, betű
+I	-R	+D	keverék	rúgnivaló, szabadság?
+I	-R	-D	keverék	aggregény

-I	~R	-D	tulajdonság	alak, jel, információ, energia
-I	~R	+D	tulajdonság	mód, stílus, összefüggés?, helyzet?
-I	~R	-D	tulajdonság	piros

Minden jel szerint nincs megoldva a formák, alakok minősítése. A természetes érzék szerint ugyanis a –I esetben a tulajdonságoknak és a formáknak kellene megjelenniük (piros, kemény, egyedi, ill. jel, forma, információ, energia, mód, stílus). A +I esetekben pedig ezek speciális megjelenései (szabadság, kör, kocka, lábnyom, betű).

A teljes rendszer kialakulása minden jel szerint további elemzéseket követel.

Hivatkozások

DOLCE. Descriptive Ontology for Linguistic and Cognitive Engineering. <http://www.loa-cnr.it/DOLCE.html>

Ferrario, Roberta; Oltmari, Alessandro: Towards a computational ontology of mind. In: <http://www.loa-cnr.it/Papers/focomf2.pdf>

Gärdenfors, P. 2000. *Conceptual Spaces: the Geometry of Thought*. MIT Press, Cambridge, Massachusetts.

Guarino, N. and C. Welty (2002). Evaluating Ontological Decisions with ONTOCLEAN. Communications of the ACM 45(2), 61–65. <http://www.loacnr.it/Papers/CACM2002.pdf>.

Masolo, Claudio; Borgo, Stefano; Gangemi, Aldo; Guarino, Nicola; Oltramari, Alessandro; Schneider, Luc: The WonderWeb Library of Foundational Ontologies. *Preliminary Report. ISTC–CNR*

Overview of OpenCycCycorp, Inc. . 2002–2006. <http://www.cyc.com/cyc/opencyc/overview>

Sowa, J. F. 1999. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. PWS Publishing Co., Boston.

SUMO ontology [Suggested Upper Merged Ontology]. Standard Upper Ontology Working Group (SUO WG). Last modified: 10/10/2005 18:23:48. <http://ontology.teknowledge.com/>

Szöts Miklós: Az ontoclean metodológia ismertetése, problémái és továbbfejlesztési lehetőségei. NKFP-2/042/04. Budapest, 2005. november 28.

<http://ontologia.hu/forum/Members/szots/OntoClean.pdf>

Ungváry Rudolf: A fogalom fogalma.

http://ontologia.hu/forum/MEO_forum_toplevel_ontology/fogalmi_definicio/

Ungváry Rudolf: Tezaurusz és ontológia, avagy a fogalmi ismertetőjegyek generikus öröklődésének formalizálása In: Tudományos és Műszaki Tájékoztatás, 51. évf. 5. sz. p. 1–10.

http://tmt.omikk.bme.hu/show_news.html?id=3615&issue_id=450