

Ungváry Rudolf  
Országos Széchényi Könyvtár

# Ontológiák és könyvtárak

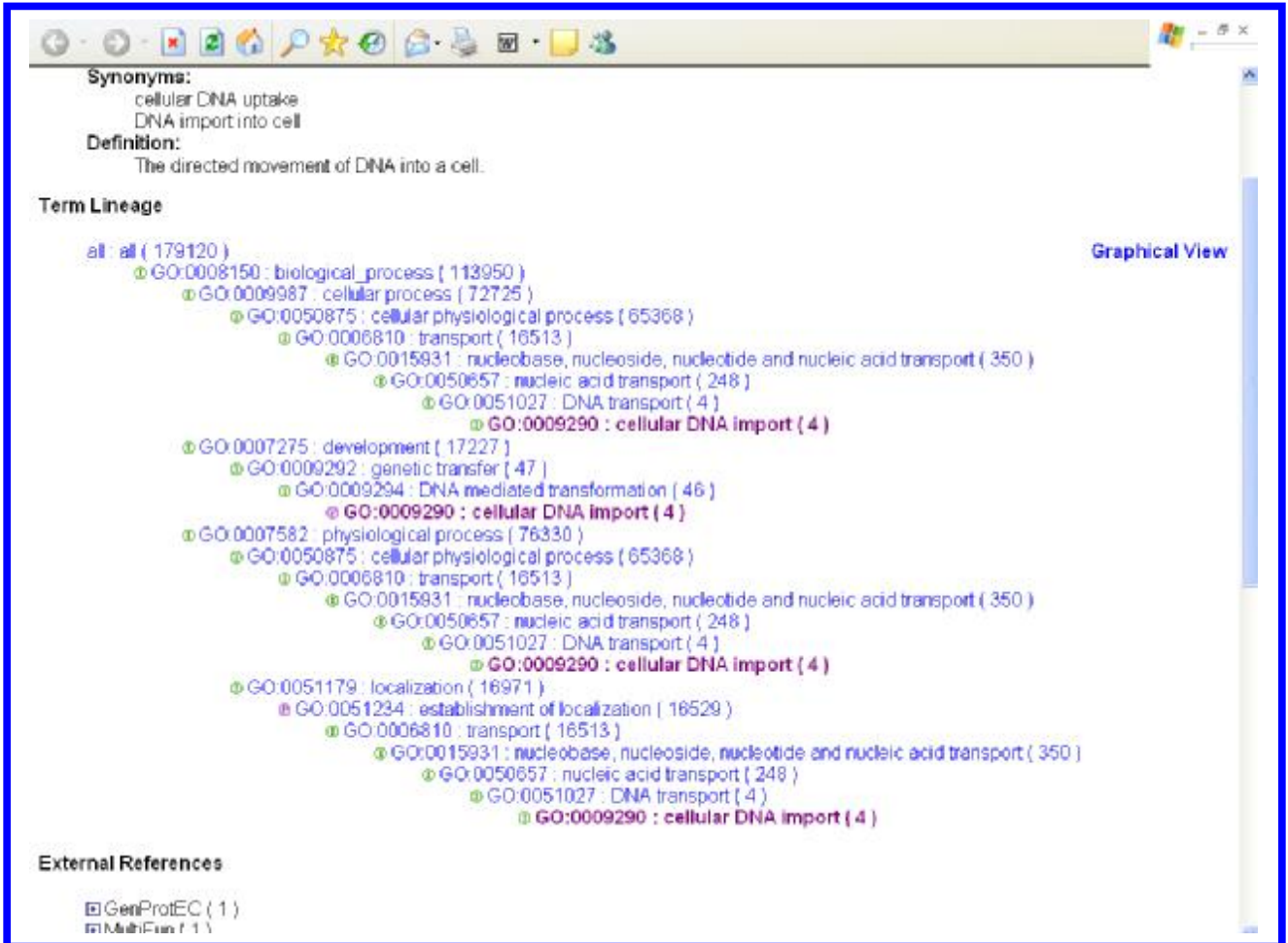
## Az ontológiák

### A fogalom magyarázata

Az ismeret- vagy tudásbázisokban<sup>1</sup> használt, *formális logikai leírásokkal ellátott* (az 1. és a 2. ábrán és az 1. és 2. táblázatban példaként bemutatott) generikus hierarchiaszerkezetekkel kapcsolatban kezdtek el a 90-es évek elejétől — jelentésátvitellel — ontológiákról beszélni. A fogalom azonban már a hetvenes években megjelent a mesterséges intelligencia kutatásban a szoftverekkel generált mesterséges világok megnevezésére.<sup>2</sup> Az eredetileg tudományt (lételméletet) jelentő kifejezés azonban nem az előbbi fejezetben értelmezett hierarchikus fogalmi rendszert jelenti önmagában, hanem annak elsőfokú logikai kijelentésekkel bővített változatát.<sup>3</sup> Az „ismerettechnológia” (tudástechnológia, knowledge engineering) nézőpontjából tehát az ontológia meghatározott ismeretterület afféle formális modellje. Az így értelmezett modell esetében korrekt az ontológia megnevezés.<sup>4</sup>

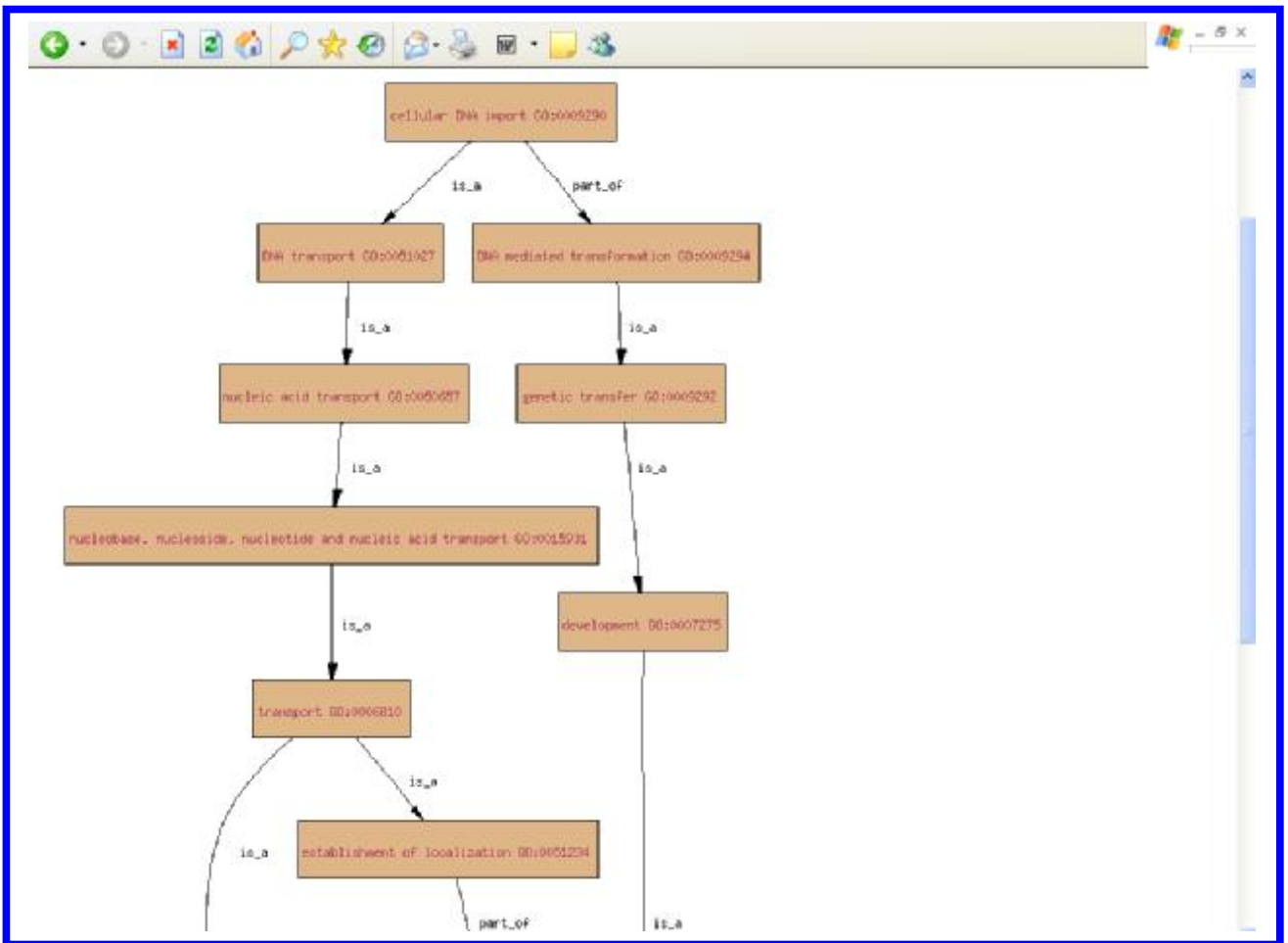
Az ontológia részét alkotó fogalmi hierarchia kifejezéseivel dokumentumok (források, elsősorban webforrások)<sup>5</sup> tartalma osztályozható, írható le. E hierarchia kifejezései tehát a dokumentumok (források) tartalmára vonatkozó ún. másodlagos vagy metaadatokat képviselnek, maga a hierarchia pedig tekinthető osztályozási rendszernek, illetve információkereső nyelvnek. A különbség a hagyományos osztályozási rendszerekhez (ETO) és információkereső nyelvekhez (tezauruszok) képest az, hogy az ontológiákkal a logikai szerkezet jóvoltából automatizáltan kihasználható a generikus hierarchia eddig csak intellektuálisan hasznosított tulajdonsága, a későbbi fejezetben részletesebben tárgyalt generikus öröklődés. Az öröklődés az ontológiákban például azt jelenti, hogy ha a „kutya” fogalmának ismertetőjegye az „ugatás”, akkor a kutyák minden fajtájára érvényes, hogy ugatnak. Ez a következtetés minden nyelvhasználó, és vele minden kereső számára magától értetődő, mégpedig anélkül, hogy a kutyák minden egyes fajtájára nézve külön-külön rögzítenie kellene magának ezt az ismeretet. Mind a mai napig azonban a logika egyetlen változata sem teszi igazán lehetővé ennek a következtetésnek a formális végrehajtását (azaz ezt az ismeretet egy formális rendszerben az egyes kutyafajták esetében külön-külön rögzíteni kell). Az ontológiákat azért találták ki, hogy ezt a problémát a mesterséges intelligencián alapuló rendszerekben és az ismeretbázisokban, rajtuk keresztül pedig valamiképpen az információkeresésben is valahogy áthidalják.

Az 1. ábrán egy weben elérhető, genetikai információkat tároló rendszer ontológiájának hierarchiája látható.



1. ábra. A Gene Ontology Consortium ontológiájának részlete <<http://www.geneontology.org/>>. (A kezdőoldalon a „Search GO” mezőben adjuk meg a „cellular DNA import” kifejezést, a megjelenő oldalon a „Go term” alatt pedig kattintsunk rá a „cellulár DNA import” kifejezésre.)

Ugyanennek a hierarchiának a grafikus változatán (2. ábra) látható, hogy a szóban forgó ontológiában már pontosan megkülönböztetik a generikus („is\_a”) és a partitív (part\_of) relációkat.



2. ábra. A Gene Ontology Consortium ontológiájának grafikus részlete <http://www.geneontology.org/>. (Az előző ábrán bemutatott képernyőn kattintsunk rá a „Graphical View” parancsra.)

## A fogalmi öröklődés szabálya

A reláció típusok között a generikus (faj–nem) relációnak kitüntetett szerepe van. A generikus reláció hierarchikus (az 1. táblázatban három ilyen hierarchia látható).

valami	valami	valami	
dolog		dolog	tulajdonság
élőlény	összesség	emberi tulajdonság	
állat	társulás	lelki tulajdonság	
gerinces		népesség	identitás
emlős		lakosság	nemzettudat
ragadozó		nemzet	
kutyafélék			
farkas			
kutya			
agár			
puli			
uszkár			

1. táblázat. Generikus hierarchialáncok (a relációlánc kisbetűs elemei a teauruszokban általában nincsenek is fölvéve, mivel a gyakorlati osztályozásban és keresésben nincs rájuk szükség).

A generikus rendezettség legfontosabb tulajdonsága, hogy az ismertetőjegyek a fölérendelt fogalomtól a fajfogalmak irányában öröklődnek.

A fajfogalom generikus alárendeltje a nemfogalomnak. Ez a hierarchia logikai nézőpontból típusok és altípusok láncából áll, alapvető tulajdonsága, hogy a fogalmak ismertetőjegyei („tulajdonságai”) az általánostól a speciális felé öröklődnek: az általánosabb típus ismertetőjegyei mind érvényesek az altípusra. Ha igaz az, hogy a kisebbségi kérdés a politikai kérdés(ek) „része”, és összefügg az „autonómia” kérdéssel, akkor ez érvényes alárendeltjére, a nemzetiségi kérdésre is. Fordítva viszont, alulról fölfelé nincs a generikus hierarchialánc menté generikus vagy fogalmi „öröklődés”.

Generikus vagy fogalmi öröklődésen (a továbbiakban: öröklődés) azt értjük, hogy az általánosabb fogalom ismertetőjegyeit a neki alárendelt speciálisabb fogalmak is tartalmazzák. Ha igaz például, hogy a „kutya” ismertetőjegye az „ugatás”, akkor minden „agár”, „puli”, „uszkár” stb. esetében is ismertetőjegy az „ugatás”. Azaz az „ugatás” faji tulajdonság (amin nem változtat, ha van olyan kutya, amely valamilyen véletlenszerű ok miatt nem képes ugatni).

A dolgok közötti relációk a dolgokat reprezentáló fogalmak ismertetőjegyein alapulnak. A „kutya” fogalmának ismertetőjegye például, hogy „jellegzetes hangadása az ugatás”, „felhasználják házörzésre”, „fertőző betegségük a veszettség”. Eme ismertetőjegyek alapján áll fenn reláció a „kutya” és az „ugatás”, a „kutya” és a „házörzés”, a „kutya” és a „veszettség”, szükségképpen pedig a kutya összes fajtája és az „ugatás”, a „házörzés” és a „veszettség” között. (Ha azonban a fenti példában említett „felhasználják házörzésre” csak bizonyos fajtájú kutyák fogalmainak ismertetőjegye, akkor nem a „kutya” és a „házörzés”, hanem csak eme kutyafajták és a „házörzés” között áll fenn reláció, tehát ez a reláció csak eme kutyafajták alfajtáira stb. öröklődik tovább).

Megfogalmazható az alábbi szabály:

A generikus relációban kapcsolódó fogalmak esetében az általánosabb fogalmak ismertetőjegyei (és vele relációi) érvényesek a speciálisabb fogalmakra is (de fordítva nem). Ez a tulajdonságok generikus öröklődése a fajfogalmak irányában.

## A fogalmi öröklődés tulajdonságai

Generikus reláció — s vele ismertetőjegyek öröklődése — csak fogalmak között, és ezeken belül is csak azonos fogalmi kategórián belül lehetséges. Elvont dolognak csak elvont dolog, összességnek csak összesség, tulajdonságnak csak tulajdonság stb. lehet a fajtája. (Szerszámnak, például a késnek nem lehet fajtája tevékenység, például a vágás, hanem csak egy másik, még speciálisabb szerszám, teszem azt a konyhakés meg a zsebkés.) Más reláció típusok is hierarchikusak, de láncaik esetében nincs feltétlen öröklődés. A partitív — rész–egész — reláció például nem fogalmak, hanem a fogalmak

terjedelmébe eső dolgok között áll fenn, s ezért nem lehetséges az ismertetőjegyek öröklődése az alárendeltek felé (a „hadsereg” ismertetőjegyei például nem feltétlenül érvényesek a „katona” ismertetőjegyeire). A relációk jelentős része pedig (mint például az oksági reláció, a „tulajdonsága” reláció stb.) különböző fogalmi kategóriák között is fennállhat, és ezért az öröklődés eleve lehetetlen<sup>6</sup>. Például az „ebtenyésztés” és a „kutya” között oksági reláció áll fenn: az ebtenyésztés „tárgya” a kutya; de a tevékenységet jelentő „ebtenyésztés” ismertetőjegyei nem lehetnek feltétlenül érvényesek az élőlényt jelentő „kutya” esetében, azaz nincs öröklődés.

Ugyancsak nincs öröklődés konkrét dolgok és azok fogalmai között. Abból például, hogy valaki magyar és elveti a magyar nyelv finnugor eredetét (vagy fordítva), nem következik, hogy a „magyar” fogalmának ismertetőjegye az „elveti a magyar nyelv finnugor eredetét” (vagy fordítva). Ha viszont igaz, hogy a „magyar nyelv” finnugor eredetű, akkor a palóc, a székely, a moldvai csángó nyelvjárás stb. (azaz minden magyar nyelvjárás) is finnugor eredetű, hiszen a magyar nyelvjárások a magyar nyelv fajtái. (Persze lehet olyan magyar nyelvjárás, melyben más nyelvi eredet is ötvöződik, de akkor ez az eredet a magyar nyelv egészére nem érvényes, csak erre a nyelvjárásra és e nyelvjárás fajtáira).

Az igazgatásban (pl. Szolgáltatások jegyzéke, Foglalkozások Egységes Osztályozási Rendszere) és a könyvtárakban (pl. Egyetemes Tizedes Osztályozás) használt osztályozási rendszerekben a fogalmakat alapvetően egyetlen, hierarchikus reláció szerint rendezik. Ez a hierarchia azonban sokszor elnagyolt, ú. n. „laza” hierarchia, azaz nem minden esetben generikus, hanem keverednek benne a különféle relációk.

Ilyen laza — nem típus–altípus — hierarchia az ETO esetében a 2. táblázatban látható

Osztály	fogalmi kategória
3 társadalomtudomány	<i>tudomány (főosztály)</i>
34 jog	<i>tudomány (osztály)</i>
343 büntetőjog	<i>tudomány (alosztály)</i>
343.1 eljárási jog	<i>tudomány (al-alosztály)</i>
343.10 büntetőeljárás	<i>folyamat</i>
343.12 vád állítás	
343.121 gyanúsított, vádlott	<i>személy</i>
343.121.4 társadalmi védő	<i>személy</i>
343.19 büntetőbíróság	<i>testület</i>

2. táblázat. ETO „laza” hierarchiája

Mindezek a hierarchikus relációszerkezetek egyben szemantikai hálók is.

A gyakorlatban a „laza” hierarchia is ismeretet (tudást) reprezentál, megszorításokkal az ETO is felhasználható az ismeret(tudás)alapú rendszerekben, noha a többféle relációtípusból más és más logikai implikációk következnek.

A hierarchiák tulajdonságaival, s ezen belül az ismertetőjegyek generikus öröklődésével azért szükséges részletesebben foglalkozni, mert az ismertetőjegyek öröklődésének az osztályozás és az információkeresés kezdeteitől fogva fontos gyakorlati szerepe volt és van az eredményes tartalmi feltárásban és keresésben. Ha például valaki túl sok információt talál a „kutya” keresőszó használatával, akkor — anélkül, hogy pontosan ismerné a kutyák egyes fajtáinak a tulajdonságait — valójában az ismertetőjegyek öröklődésére támaszkodva számíthat arra, hogy ha a „kutya” fajtáival keres, továbbra is öt érdeklő információra akad és fordítva.

A fogalmi öröklődés szabálya tudományos természetű és természetesen nem közismert. Ösztönösen, a nyelvérzék alapján azonban sokszor alkalmazzák, aminek jellegzetes példája az előbbi, kutyákra vonatkozó keresés esete. Világnézeti téren pedig fatális következménye lehet annak, ha valaki — ösztönösen — hibásan alkalmazza az öröklődés szabályát, és például azt hiszi, attól, hogy magyarként valamit igaznak tart, az minden magyarra igaz kell, hogy legyen.

A generikus relációnak ezt az ismertetőjegy-örökítő tulajdonságát általában csak emberi közreműködéssel (intellektuálisan) lehetett kihasználni.<sup>7</sup> Az ismeret(tudás)alapú rendszerekben az ún. ontológiák rendeltetése, hogy automatizáltan is kihasználhatóvá válják ez a tulajdonság. Az ontológia alapú ismeretbázisokban a teauruszokban is alkalmazott generikus hierarchiának éppen a fentiekben tárgyalt átörökítő tulajdonságát használják ki „ha... akkor” következtetések (implikációk) formájában.

Az ismeretet (tudást) ezzel a hierarchiával reprezentálják, és a logikai következtetéseket erre alapozva fogalmazzák meg. Ezért a generikus reláció nevezhető ontológiai (ontológiaképző) relációnak.

## Filozófiai kitekintés

Az ontológiák számítástechnikai megjelenésének mélyreható és a filozófiai alapkérdéseit érintő következményei is vannak, melyek részletes tárgyalására itt nem térhetünk ki.<sup>8</sup> Csak utalunk arra, hogy az ontológiák a filozófia meghaladottnak tűnt problémáit elevenítik fel és töltik meg új tartalommal.

Egyrészt belőlük kiindulva egy arisztotelészi világ tételezhető fel a szubsztanciák által képviselt referenciákkal.<sup>9</sup> (Például az „állat–emlős–ragadozó–kutya” hierarchialánc alapján feltételezhető a tényleges állat, emlős, ragadozó és kutya léte.)

Másrészt viszont értelmezhető az ontológia (s vele a példaként említett hierarchialánc) úgy is, mint terminológiai jellegű meghatározáshalmaz, más szóval egy üres formális rendszer szimbólumai (ahol az „állat”, „emlős”, „ragadozó” és „kutya” változónak nem kell megfelelnie egy állatnak stb., legfeljebb a programozó fejében). Ez utóbbi értelmezés alapja Quine ún. elkötelezettségi tézise, amely szerint minden elmélet maga után von egy ontológiát, azaz létezőnek tételezi fel azokat a dolgokat, amelyet az elmélet változói jelölnek. Ezek a dolgok azonban nem objektív referenciák.

E két — arisztotelészi, illetve a Quine nevével fémjelvezhető analitikus filozófiai — értelmezés egymást kizárja. Hogy melyiknek van igaza, az minden bizonnyal a filozófia soha el nem dönthető kérdései közé tartozik. Az azonban figyelemre méltó, hogy maguk az itt tárgyalt — ha tetszik „számítógépes” — ontológiák hasznot hajtóan működnek.

Az információgazdaság legkorábbi, klasszikus eszközei, az osztályozási rendszerek (pl. az ETO) és tezaszok mögött valójában szintisztán ugyanez a fenti filozófiai problémakör húzódik meg, de a legújabb korban alig váltott ki filozófiai érdeklődést. Az ontológiákon keresztül az informatikai-számítástechnikai világ (ha tetszik: közösség) tudta felhívni magára a filozófiai figyelmet. Nemcsak az ontológiák hatékonyságának növeléséhez, hanem a könyvtári világ felértékelődéséhez is vezet, ha sikerül megteremteni a kapcsolatot eme hagyományos osztályozási rendszerek és más információkereső nyelvek, ill. az ontológiák között.

Figyelemre méltó, hogy az informatikában beindított „ontológia-felhajtásból” eddig nem csak a könyvtári, hanem a nyelvészeti „közösségek” se nagyon tudtak „részt kapni”, holott történetileg az első, a „megjelenítés” dolgában pedig a másodikként említett szakterület művelői legalább annyira illetékesek lennének a dologban. Van az úgy néha, hogy osztályharc helyett szakmák harca is folyhat az „újrarendezésért”.

## Konkrét (praktikus) fejlemények

Az ontológiák kialakulása az elmúlt néhány év fejleménye. Azoknak a törekvéseknek az egyik praktikus eredménye, melyek célja, hogy a web forrásai a tartalmuk alapján jobban elérhetőek legyenek. A webnek ezt a mai böngészőknél jobban kereshető, a következő évtizedekben valószínűleg kialakuló változatát nevezik szemantikus webnek. E téren viharos fejlődés tanúi lehetünk.

Az egyik legkorábbi fejlemény, hogy 2000-ben közreadtak egy „tématérképnek” (topic map) nevezett hierarchikus fogalmi struktúrát kezelő szabványt [26]. A weben jelenleg található vizualizált fogalmi struktúrák jelentős része ezen vagy ehhez hasonló fejlesztéseken alapszik.<sup>10</sup>

A közhasznú webes fejlesztéseket támogató World Wide Web konzorcium (W3C)<sup>11</sup> égisze alatt egy másik irányban indult el a fejlesztés. Egyik első eredménye volt, hogy 2000-ben a web metaadatainak leírására szabvány született, az XML-en alapuló webforrás-(webdokumentum-)leíró nyelv (Resource Description Framework, RDF).<sup>12</sup>

A weben alkalmazható hierarchikus fogalmi struktúrák — a vele az előbb említett generikus öröklődés — formális leírására is ezt a nyelvet használták fel, amikor a W3C megbízásából 2002-ben hozzákezdtek az ontológiák szabványának tekinthető webontológia-nyelv (Ontology Web Language, OWL) kialakításához. Az ideiglenes szabványajánlást 2003 tavaszán adta közre a W3C [14].

Ma általában — amint ez Sowa és Gruber előzőkben idézett ontológia-meghatározásaiból is kitűnik — meglehetősen korlátoltan fogják fel a fogalmi struktúrákat, mintha ezek kizárólag generikus összefüggésekből állnának. A tezaszok e tanulmányban is szemléltetett példája bizonyítja, hogy a fogalmi

struktúrák ennél sokkal több relációból állnak. Az OWL is azon alapszik, hogy csak a tisztán generikus hálók lehetnek ontológiák, s vele az ismeretrepresentáció alapjai, afféle „háttértudás” (típus–altípus hierarchián elvileg csak generikus reláció értendő). Ezzel szemben a tapasztalatok fényében az az igaz, hogy az ismeretet (tudást) nem egyedül a dolgok közötti tipológiai (generikus), hanem számtalan egyéb partitív, oksági, tulajdonsági, ellentét stb. összefüggés is reprezentálja még. Az OWL szerint osztályok között csak típus–altípus (tehát generikus) összefüggések lehetségesek, minden más reláció csak „tulajdonságként” definiálható. A teauruszok ismert [8] [12] relációszerkezetének fényében az ismeret reprezentációja ennél bizonyára komplexebb formalizációt igényelne (mint ahogy arra az utolsó fejezetben még rámutatunk, teauruszokat ma elég körülményesen lehet ontológiaformátumban leírni).

## Az ontológia-fogalom inflálódása

Az ontológia fenti, korrekt értelmén kívül van egy sor más értelemben is elterjedt, nem valami szerencsés névhasználat is.

Sommásan ugyanis „ontológiának” nevezik sokszor magát a minden logikai leírás nélküli generikus fogalmi rendszereket és a szemantikai hálókat is. Ilyen alapon ontológia lehetne a teauruszon belül, annak részét alkotó generikus kapcsolatok hierarchialánca (1. táblázat), meg az ETO is (2. táblázat).<sup>13</sup> Az efféle használat többek között a Yahoo! webkatalógus szerkesztői révén terjedt el a 90-es évek közepétől, akik a webkatalógusukban használt osztályozási rendszert nevezték „ontológiának” — amely egyébként „laza” hierarchiája következtében az ETO-nak megfelelő rendszer, csak éppen természetes nyelven adják meg az osztályokat. Ezen az alapon „ontológia” a Yahoo!-nak megfelelő magyar (Alta)Vizsla internetkatalógus (osztályozási rendszere) is.

Az „ontológia” megnevezés alkalmazása és kiterjesztése a pusztán generikus fogalmi struktúrákra bizonyára azért következett be, mert az ismeretrepresentáció informatikai (értve ezen számítástechnikai) művelői és a webkatalógusok készítői nem az igazgatás, a statisztika, a különféle rendszertanok, dokumentációs, irattári és archiválási rendszer, és főleg nem a nyelvészet és a könyvtárak világából érkeztek. Ez utóbbiakban a sommásan „ontológiának” nevezett, de leíró logikai szerkezettel nem rendelkező rendszereket osztályozási rendszereknek, nomenklatúráknak, taxonómiáknak stb. nevezik és hagyományosan az információk tartalmi feltárására és keresésére használják. Innen nézve az ontológia tekinthető generikus öröklődést biztosító logikai nyelvvel kiegészített hierarchikus osztályozási rendszernek. A másik oka e kiterjesztett szóhasználatnak talán az, hogy az osztályozási rendszerek valóban a létről alkotott ilyen-olyan elképzelések tükröi, a lét tudománya pedig az ontológia.

Az „ontológia” kifejezés használatának inflálódására utal, hogy újabban a közös nyelv, szókinccs használatának szabályrendszerét, terminológiai szabályait (például egy testületen, intézményen belül), sőt tantárgytematikák leírását is „ontológiának” nevezik.

Hasonló jelenség játszódott le az elmúlt években az eleve téves „fogalomtár” kifejezéssel. Mesterségesen tárolni csak hordozón rögzített szavakat, kifejezéseket lehet, fogalmakat nem (melyek egyedül a tudatban léteznek), de ez az egyszerű tény talán annyira szubtilis, hogy bizonyára fel se fogható. „Fogalomtárnak” többnyire a strukturált terminológiai szótárakat nevezik, élvonalbeli újdonságot sejtetve egy fából vaskarikára kifejezéssel („fogalomtárról” beszélni olyan, mintha a személyi adatok nyilván tartását „embertárnak” neveznék).

## Ontológia és könyvtári rendszer

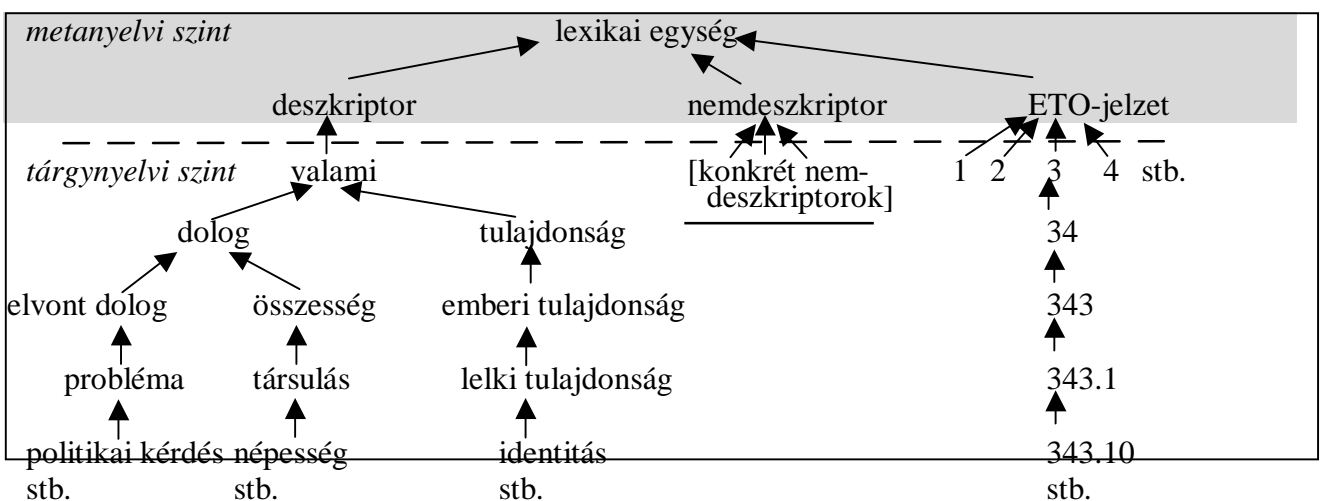
Egy könyvtári rendszerben a legfontosabb mindig a dokumentumok azonosítása marad. A találatok — keressék azokat teauruszok generikus hierarchiába szervezett kifejezéseivel, szerzők vagy testületek neveivel, évszámokkal, nyelv kódok stb. alapján — így vagy úgy, de a dokumentumok leírásait fogják tartalmazni. A tartalmi feltáráshoz használt osztályozási rendszer/információkereső nyelv állománya — a metanyelvi állomány — a szoftverkészítők számára sokáig afféle mellékes összetevőnek tűnhetett, talán ezzel is magyarázható, hogy a teauruszok és osztályozási rendszerek könyvtári rendszerbe integrált, rugalmas, sokoldalú és felhasználóbarát kezelésének a kérdése még ma sincs kielégítően megoldva.

Egy ismeretbázisból a távoli jövőben — a logikai szerkezet jóvoltából — szövegesen megfogalmazott kérdésekre talán szöveges válaszokat lehet majd kapni. Ezért a metanyelvi állomány jelentősége felértékelődik. Éppen ennek a felértékelődésnek a jele az ontológiák megjelenése. (E tényrel előbb-utóbb a könyvtári adatcsere-formátumok és teauruszszabványok gondozóinak és használóinak is szembe kell majd nézniük [3] [4] [24], és hatással lesz a teauruszokra, többek között az átfogó magyar teauruszra [5] [6] [8] [12]is.)

Noha az on-line katalógusok nem ismeretbázisok, ugyanakkor fejlődésükre az ontológiai alapú ismeretprezentáció hasznos befolyást gyakorolhat. Ez talán változtathat majd a teauruszok és az ETO jelenleg sanyarú kezelési komfortján is<sup>14</sup>, és jelentős mértékben javulhat az osztályozási rendszerek/információkereső nyelvek használatának eredményessége. Ezért fontos, hogy az informatikának ez az „ismerettechnológiai eszköze” ismertté váljék a könyvtári világban.

Mivel az ontológiai alapú ismeretbázisok potenciális felhasználói köre jelentős részben a pénzügyi és a gazdasági szféra, a fejlesztések mögött jelentős anyagi várakozások húzódnak meg. Ezzel magyarázható az újszerű, sokat ígérő vagy csupán más kifejezések használata („tudásintenzív”, „szemantikai erő”, „interoperabilitás” stb.), s vele az olykor promóciós célzatú, sulykoló megfogalmazások. Mindez a gyorsan fejlődő szakterületek velejárója. A könyvtártudomány, az osztályozás/indexelés és az információkeresés a maga lényegesen gyengébb önérvényesítő lehetőségei következtében egyelőre e fejlődés árnyékában marad. Fontos, hogy kapcsolatot teremtsünk a hagyományos és az új szóhasználat között, és megvilágítsuk, hogy a könyvtári világ az információgazdaság legkorábbi intézménye, melynek számos eredménye felhasználható az ismerettechnológiában, s egyúttal sikerrel integrálható a tartalomiparban (ahogy a webes információszolgáltatók saját működési környezetüket nevezik).

A teaurusz (és az ETO) ontológiai formája a hagyományos teauruszokhoz és ETO-hoz szokott szakember számára első pillantásra meglehetősen szokatlan, mivel ebben a formában nincs különbség a metanyelvi és a tárgynyelvi szint osztályai között. A teaurusz metanyelvét a „lexikai egységek”, a „deszkriptor” és a „nemdeszkriptor” alkotják. Az ETO esetében a metanyelvi osztályból csak egy van, az „ETO-jelzet”. A tárgynyelvi kifejezéseket pedig a konkrét lexikai egységek és a konkrét ETO-jelzetek alkotják (mint amilyenek az 1. ábrán, ill. a 2. táblázatban is láthatók). Azaz az ontológiákban a metanyelvi és a tárgyi kifejezések többnyire egyetlen — generikus — hierarchialáncban helyezkednek el, noha az igényesebb rendszerekben már a partitív relációszerinti hierarchiát is feltüntetik (2. ábra). A többi — oksági, egyéb — reláció elve nem alkot folytonos hierarchialáncot, és ezért ugyanúgy csak tulajdonságadat az ontológiában, akár csak a lexikai egységek vagy az ETO-jelzetek megnevezései, felvételi dátumai, magyarázatai, forrásadatai stb. (A konkrét megoldás egy példáját az utolsó fejezet tartalmazza.)



3. ábra. Teaurusz és ETO ontológiai csúcsfogalmai. A szaggatott vonal feletti szürke sávban a metanyelvi, alatta a tárgynyelvi kifejezések helyezkednek el. Az elágazásokat nagy fokú leegyszerűsítéssel ábrázoltuk. A nemdeszkriptorok alatti záróvonal azt jelzi, hogy ezeknek a lexikai egységeknek nincsen további alosztályai.



## Metaontológiák

Metaontológiák az ontológiák ontológiái: egyszerűbben fogalmazva az átfogó, elvileg minden speciális ontológiához felhasználható felső szintű ontológia<sup>15</sup>. Nemzetközileg ismert példája a Cyc-ontológia, melyet Douglas Lenat és Ramathan Guha többek közreműködésével a 80-as évek végén kezdtek kialakítani. Az „OpenCyc” kereskedelmi néven ismertté vált egyetemes célú, de elsősorban a teljes társadalomtudomány átfogó, „enciklopédikus” fogalmi rendszerük a generikus reláción alapul. Ennek legáltalánosabb csúcspfogalma a „thing” („valami”). A „cyc” elnevezés az „enciklopédia” angol nevére utal.

A rendszer mára közel 100.000 generikus hierarchiába szervezett fogalommegnevezést (osztályt) tartalmaz milliós nagyságrendű tulajdonság-adat és meghatározás mellett, Egyszerűsített változata kereskedelmi forgalomban kapható. Rendeltetése, hogy afféle újrafelhasználható „metaontológiaként” az egyes speciális ontológia-alkalmazások közötti leképezéseket közvetítse. Más szóval háttérinformációkat képvisel, melyek relevánsak a speciális alkalmazásokban, de nem specifikusak az adott feladatra nézve: a speciális ontológia átfogó részét alkotják és ezért tetszés szerinti speciális ontológia kialakításához kiindulásként használhatják fel.<sup>16</sup> A tezauruszok analógiájára nevezhetnénk „csúcsontológiának” vagy „makroontológiának”.

A Köztaurusz a jelenleg közel 70.000 lexikai egységével, és ezen belül a közel 35.000 nem-faj (generikus) relációban kapcsolódó, az ontológiában használt nyelven: osztály típus–altípus szerkezetébe rendezett hierarchiájával egy ilyen — ideális esetben szabadon hozzáférhető — kiinduló rendszert képvisel, ha OWL-formában is megvalósul.

Mindennek már csak azért is jelentősége van, mert ezáltal a könyvtári, osztályozási/indexelési valamint információkeresési igények is jobban megvalósíthatók az ismerettechnológia keretei között. Mindez visszahathat a könyvtári rendszerek további, felhasználóbarát korszerűsítésére.

## Az ontológiák egyik gyökere

A Cyc-ontológia megszületésében az indiai Guha részvétele több volt, mint szokott tudományos kooperáció.

A mai tartalmi osztályozás klasszikus megalapozóinak egyike, Ramamrita Ranganathan se véletlenül volt indiai. Az ind kultúrában hagyományosan többféle létfelfogás is megengedett volt. Természetesnek számított, hogy — vallási szemmel nézve — többféle úton is el lehet jutni az üdvözüléshez, vagy — tudományos szemmel nézve — a létnek többféle elmélete fér meg egymás mellett. Ranganathan személyében e hagyomány találkozott az európai műveltséggel és meghatározó szerepet játszott a 20. század első felében kialakított analitikus–szintetikus osztályozáselméletének és többdimenzionális osztályozási rendszerének a kialakulásában.

Ranganathan eredményei a 20. század közepétől megtermékenyítően hatottak a modern osztályozásra. Megvetették a szemléleti alapját a nem szisztematikus (nem a tudományok mechanikus felosztásán alapuló), hanem a generikus fogalmi kapcsolatokra épülő polihierarchikus osztályozási rendszereknek (taxonómiáknak, fogalmi struktúráknak), és rajtuk keresztül a tezauruszoknak és újabban az ontológiáknak.

A Cyc-ontológiában — s valójában az OWL szabványban rögzített ontológia-felfogásban is — ennek a szemléletnek a hatása is megjelenik.<sup>17</sup>

## Az ontológialeíró-nyelv (OWL)

Az ontológialeíró-nyelvvvel (OWL, Web Ontology Language)<sup>18</sup> a webes ismeretbázisok, szakértői rendszerek tartalmi feltárást és keresést biztosító ontológiájának géppel olvasható és interpretálható formája írható le. Olyan logikai leíró nyelv, melynek segítségével következtetésekre képes ismeretbázis alakítható ki. Az OWL segítségével (a) osztályok hierarchiája, továbbá (b) osztályok és egyedek jellemzői és az (c) osztályok és osztályok, osztályok és egyedek, egyedek és egyedek között fennálló kapcsolatok (asszociációk), nem utolsó sorban pedig az osztályok ismertetőjegyeinek az osztályok generikus hierarchiájából következő (d) öröklődése írható le. A logikai következtetések ezeken az ismertetőjegy-öröklődéseken alapulnak. Mindaz, ami az OWL segítségével leírt osztályokra jellemző,

érvényes az osztályok terjedelmét alkotó egyedekre. A jellemzők és a kapcsolatok összefoglalóan az osztályok tulajdonságai.

Az OWL értelmében vett ontológia tehát a következőkből áll (a felsorolás nem teljes):

- osztályok (Class);
- egyedek (ID);
- osztályok közötti generikus — taxonómiai, osztály–alosztály (típus–altípus) — és ekvivalencia-relációk (subClassOf; equivalentClassOf);
- öröklődő és nem öröklődő osztály- és egyedtulajdonságok (type):
  - osztályok közötti, osztályok és egyedek közötti és egyedek közötti nem generikus relációk (ObjectProperty);
  - relációtulajdonságok (tranzitív, szimmetrikus, inverz stb.)(TransitiveProperty, SymmetricProperty, inverseOf stb.);
  - osztályok és egyedek magyarázatai, hozzájuk kapcsolódó megjegyzések, felvételi dátumuk, forrásuk stb. (DatatypeProperty);
  - öröklődő osztály- és egyedtulajdonság-típus és -érték (FunctionalProperty);
  - nem öröklődő osztály- és egyedtulajdonság-típus és -érték (AnnotationProperty);
- feltételek, megkötések (Restriction), pl. előfordulás, (Cardinality), kizárás (disjointWith);
- rendszertájékoztatók (comment).

Fontos tudnivaló, hogy az OWL három fokozatban létezik (könnyű, közepes és teljes), melyek közül a jelenlegi ismeretalapú és szakértői rendszerek általában csak a könnyű, ritkább esetben a közepes változatot képesek kezelni. A teljes változat kezelése még a jövő zenéje, de a szemantikus web irányába tartó gyorsuló fejlődést feltételezve ez nem fog sokáig várni magára.<sup>1</sup>

## Irodalom

- [1] Copeland, B. J. CYC: a case study in ontological engineering. = Electronic Journal of Analytical Philosophy Vol. 5. 1997. p. 42–61. <<http://ejap.louisiana.edu/EJAP/1997.spring/contents.html>>
- [2] Gruber, Tom.: A translation approach to portable ontology specifications. = Knowledge Acquisition, Vol. 5. Nr. 2. 1993. p. 199–220.
- [3] Handbook on ontologies. Ed. by Staab, Steffen; Studer, Rudi. Hamburg, Springer, 2003. 243 p. (International handbooks on information systems). ISBN 354 040 834 7
- [4] HUNMARC, a bibliográfiai rekordok adatsere formátuma. Összeáll. Sipos Márta; kész. az Országos Széchényi Könyvtár Fejlesztési Osztályán. Budapest: OSZK, 1993. 129 p. ISBN 963 200 344 6.  
HUNMARC, a besorolási rekordok adatsere formátuma. Összeáll. Sipos Márta. 1998 március. 61 p. Tervezet
- [5] ISO/DP 5954–1985 Guidelines for the establishment and development of multilingual scientific and technical thesauri for information retrieval.
- [6] ISO/IS 2788–1986 Guidelines for the establishment and development of monolingual scientific and technical thesauri for information retrieval.
- [7] Kampis, György: A filozófia felfedezése a gépek világában. = Megismeréstudomány és mesterséges intelligencia (szerk.: Pléh Csaba), Budapest, Akadémiai Kiadó, 1998. p. 230–256. <<http://hps.elte.hu/~kampus/Mirror/makog98.html>>
- [8] Köztársaság. 2005. január 1. <<http://www.oszk.hu> — [Útmutató/Tezaurusz](http://mek.oszk.hu/adatbazis/thes.htm)> és <<http://mek.oszk.hu/adatbazis/thes.htm>>
- [9] Lenat, D. B.: Cyc: a large-scale investment in knowledge infrastructure. = Communication of the ACM. Vol. 38. Nr. 11. 1995. p. 33–38.

<sup>1</sup> Megjegyezzük, hogy tezauruszok is leírhatók az OWL segítségével.

- [10] Lenat, Douglas B.; Guha, Ramanathan V.: Building large knowledge-based systems. Representation and inference in the Cyc project. Reading, Addison–Wesley, 1990. 254 p.
- [11] Lenat, Douglas B, Miller, George A., Yokoi, Toshio: Cyc, WordNet and EDR. Critiques and responses. = Communications of the ACM, Vol. 38. Nr. 11. 1995. p. 45–48.
- [12] MSZ 3418–87 Magyar nyelvű információkereső teauruszok. Szerkezete, részei és formái.
- [13] OpenCyc. Formalized common knowledge <<http://www.opencyc.org>>
- [14] OWL Web Ontology Language Overview. W3C Candidate Recommendation 18. Aug. 2003. Ed. by Deborah L. McGuinness and Frank van Harmelen. <[OWL Web Ontology Language Overview](#)>
- [15] Sowa, John F.: Knowledge representations. Logical, philosophical, and computational. Pacific Grove, Brooks/Cole, 2000. 594 p.
- [16] Tezaurusz vizualizáció II. Budapest, Frutta Elektronika, 2003. Version 3.0 <<http://www.frutta.hu>>
- [18] Ungváry Rudolf: Tezaurusz-technológia. Az információkereső teauruszok készítésének folyamata. Közr. az Országos Széchényi Könyvtár Könyvtártudományi és Módszertani Központ. Budapest, NPI, 1979. 277 p.
- [19] Ungváry Rudolf: Az OSZK teaurusza és a KÖZTAURUSZ. = Könyvtári Figyelő, Új folyam 11. (47.) évf. 1. sz. 2001. p. 11–40. <<http://www.oszk.hu/kiadvany/kf/2001/1/ungvary.html>>
- [20] Ungváry Rudolf: Tezaurusz a felhasználói felületen. = Tudományos és Műszaki Tájékoztatás, 48. köt. 3. sz. 2001. p. 99–108. <<http://tmt.omikk.bme.hu/archiv.html>>
- [21] Ungváry Rudolf; Orbán Éva: Osztályozás és információkeresés. Kommentált szöveggyűjtemény. 1. kötet. Az osztályozás és elmélete. Budapest, Országos Széchényi Könyvtár, 2001. 543 p. ISBN 963 200 425 6.
- [22] Ungváry Rudolf; Vajda E.: Könyvtári információkeresés. 2. jav. kiad. Budapest, Typotex, 2002. 170 p.
- [23] Ungváry Rudolf; Vajda Erik: Az információkeresés szavai. = Tudományos és Műszaki Tájékoztatás, 50. köt. 12. sz. 2003. p. 1–27. <<http://tmt.omikk.bme.hu/archiv.html>>
- [24] USMARC format for classification data, including guidelines for content designation. 1991 Edition. Prep. by the Network Development and MARC Standards Office; Cataloging Distribution Service, Library of Congress, Washington, Update No. 1. July 1995. <<http://www.tlcdelivers.com/tlc/crs/clas0001.htm>>
- [25] Visual Thesaurus. A dictionary of the english language. New York, Plumb Design, 1998–2004. <<http://visualthesaurus.com/>>
- [26] XTM Topic-Map Standard, ISO/IEC 13250:2000. XTM TopicMaps Org. <<http://www.topicmaps.org/xtm>>

## Jegyzetek

- <sup>1</sup> „Tudásbázis”, „tudástechnológia”, „tudásalapú rendszer” helyett helyesebb az „ismeretbázis”, „ismerettechnológia”, „ismeretalapú rendszer” kifejezés, mivel az ismeret inkább jelenti a közvetíthető, adatok formájában reprezentált tudást, a tudás pedig az ismeret tudati „mélyszerkezetét”. A tudás felfogható értelmezett ismeretnek, az ismeret pedig kommunikálható tudásnak.
- <sup>2</sup> Történeti áttekintését lásd Staab és Steffer kézikönyvében [3].
- <sup>3</sup> John Frederick Sowa 2000-ben megjelent, az ismeretreprezentációról szóló könyvében az ontológiát így határozza meg: „Ontológia valamilyen tárgykörben létező vagy feltételezett dolgok kategóriáinak elmélete. Egy ilyen elmélet konkrét eredményét (is) ontológiának nevezik. Egy (konkrét) ontológia a tárgyak ama típusainak katalógusa, melyekről felteszik, hogy az adott T érdeklődési körben olyasvalakinek a nézőpontjából léteznek, aki L nyelvet használ a T tárgykörre vonatkozó gondolatainak megfogalmazására. Az ontológia típusait annak az L nyelvnek a predikátumai, szavainak jelentései vagy fogalmak és relációtípusok képviselik, melyet az adott T körben a kérdések megvitatására

használnak. ... A logika és az ontológia összekapcsolása révén nyelv keletkezik, mellyel adott érdeklődési terület dolgai, összefüggései megfogalmazhatók.

A formális ontológiát fogalmak és relációtípusok összessége alkotja, melyeket a típus–altípus kapcsolódások szerint részlegesen rendeznek. A formális ontológiák tovább finomíthatók azáltal, hogy milyen módon különböztetik meg az alárendelt típusokat a fölérendeltjeiktől: az axiomatikus ontológiákban a megkülönböztetést formalizált nyelven megfogalmazott axiómák és meghatározások segítségével végzik el, mint amilyen a logikai nyelv vagy valamilyen logikai nyelvre lefordítható számítógép-alapú jelzet; a prototípus-alapú ontológiákban a megkülönböztetés alapjai az altípust reprezentáló prototípus tipikus ismertetőjegyei. A nagyobb ontológiákat többnyire kevert módszerek jellemzik: formális meghatározásokat, axiómákat használnak a matematika, fizika és műszaki tudományok kifejezéseihez; prototípusokat használnak a növények, állatok és az elemi, közkeletű dolgok terén.” <[www.jfsowa.com/ontology/](http://www.jfsowa.com/ontology/)> és [15, p. 20. és 51]”.

1993-ban Tom Gruber még egyszerűbben fogalmazott: „Ontológia megegyezésen alapuló fogalmi rendszer formális, egyértelmű leírása.” („An ontology is a formal, explicit specification of a shared conceptualization”) [2] A „megegyezésen alapuló” kitétel fontos: azt a felfogást tükrözi, hogy az ontológiák — akárcsak az osztályozási rendszerek — nem feltétlenül a valóságban fizikai referenciákkal rendelkező struktúrák, hanem afféle szemantikai szabályrendszerek, melyek a dolgok rendezésére használhatók.

- <sup>4</sup> Az ontológiákról részletesen lásd [2], [3] és [15]. Az utóbbi egyben példákat is tartalmazó egyetemi szintű tankönyv.
- <sup>5</sup> A web körülményei között a dokumentumokat forrásnak nevezik. A weblap például könyvtári szempontból webdokumentum, a weblap része pedig részdokumentum; más nézőpontból viszont mindkettő webforrás (ahogy a részdokumentum és az azt tartalmazó [gazda]dokumentum is egyaránt dokumentum).
- <sup>6</sup> E relációk esetén is van azonban szabályszerűség. Például az eszköz–rendeltetés reláció csak anyag/tárgy és tevékenység között állhat fenn. A relációk kategoriális függőségére vonatkozóan lásd [18, p. 174, 178 és 185.].
- <sup>7</sup> E kihasználhatóság a ma kereskedelmi forgalomban található jelentősebb könyvtári rendszerekben (szoftverekben) meglehetősen gyatra (még ha találni is példát keresés céljából az ún. automatikus hierarchiaszint emelésre, ill. csökkentésre („upposting” stb.) mivel a kezelőrendszereket nem készítették még föl a teauruszok felhasználóbarát kezelésére [18] [19] [20]. Idővel ez a helyzet — talán éppen az ontológiák megjelenésének hatására — változni fog.
- <sup>8</sup> Részletesebben lásd Kampis György tanulmányát [7].
- <sup>9</sup> Szubsztancia a létező dolgok legáltalánosabb és legbensőbb lényege, esetünkben érthetjük rajta a dolgokról alkotott fogalmakat. Filozófiai és informatikai értelemben referencián jelnek vagy jelsorozatnak (szükségképpen szavaknak is) jelrendszeren (szükségképpen a nyelven) kívüli valóságra való vonatkoztatást értik: azt, hogy például adott szónak mi felel meg a fizikai valóságban.
- <sup>10</sup> Az AltaVizsla teaurusz nagyon szép vizualizációja látható a budapesti Frutta honlapjáról kiindulva [16]. Az egyszerűnek látszó, de meglehetősen talányos honlapon először a „Belépés az oldalra”, majd a „Referenciák” parancsra kell kattintani. A kellő ügyességgel rendelkező érdeklődő a bal oldali gördítő nyilat működtetve a szalagszerűen futó referenciák között el tudja kapni a „Teaurusz vizualizáció II” feliratot, és a letöltést megvárva célhoz ér. Más fajta küszködés a new yorki Plumb Design teauruszának vizuális megjelenítése [25]. Itt meg kell várni, míg a nyitó oldal letöltődik, a folyamatot azonban a böngésző nem jelzi, a kereső csak többszöri kilépési kudarc után jön rá arra, hogy türelmesen várnia kell, amíg a bal felső sarokban a „<look it up” felirattal ellátott keresőmező megjelenik, ahová az angol keresőszót beírhatja. A szenvedésnek azonban jutalma a láncreakciószerűen feltáruló fogalmi struktúrák csillagködöket és galaxisokat idéző szemantikai szépsége.
- <sup>11</sup> A World Wide Web Consortiumot (rövidített nevén W3C) 1994 októberében alapították azzal a céllal, hogy elősegítsék az olyan közös számítógépes protokollok kidolgozását, amelyek előremozdítják a Web fejlődését. Ma már a W3C-nek több, mint 450 tagja és közel 70 teljes munkaidős munkatársa van szerte a világon (és magyar irodája is).
- <sup>12</sup> Az RDF kidolgozása a W3C égisze alatt 1997-ben kezdődött el. A munka során többek között figyelembe vették a dublini metaadatszabványt és az OCLC weben elérhető WorldCat katalógusához való könyvtári csatlakozás „Connexion” nevű szolgáltatását is.
- <sup>13</sup> Ha a hierarchialáncok legfelső szintjeiről van szó, az angol nyelvű szakirodalomban a „top level hierarchy” kifejezést használják [15, p. 67.]. Magyarul „csúcshierarchia” a megfelelő kifejezés.
- <sup>14</sup> E sanyarúságról részletesen lásd [20].
- <sup>15</sup> Teauruszok esetén a „csúcsteaurusz”, „makroteaurusz” elnevezéseket használnak.
- <sup>16</sup> Lásd a rendszerről és annak vitájáról szóló publikációkat [9][10][11]. Az analitikus filozófia képviselőinek egy része az „Electronic Journal of Analytical Philosophy” 1997. évi 5. számában — melyet az ontológia filozófiai kérdéseinek szenteltek — igyekezett megsemmisítő bírálatban részesíteni a rendszert [1]. A Cyc-ontológia mögött ugyanis arisztotelészi alapokat orrontottak, azaz olyan rendszert, mely mögött készítőik szerint tényleges — szubsztanciális — dolgok vannak — holott az analitikus felfogás szerint — leegyszerűsítve — csak összefüggések létezhetnek. Ez nem akadályozta meg, hogy a Cyc-ontológiára eredményes kereskedelmi vállalkozás épüljön [13].
- <sup>17</sup> Ranganathan osztályozáselméletét és ontológiai szemléletét részletesen tartalmazza az osztályozás és információkeresés kommentált szöveggyűjteménye [21, p. 76–125.]
- <sup>18</sup> Teljes áttekintése megtalálható a W3C honlapjáról kiindulva a weben [14].