

Horváth Zoltánné

Taxonómia – az egyezményes nyelvek szerepe és rokonságai – útközben a szemantikus webhez

Az információkeresés eredményessége és az információforrások elvileg határoktól mentes elérése között feszülő ellentmondás hosszú évek óta arra készíti a kutatókat és az informatikai szakembereket, hogy a tudás- és ismeretszerzés folyamatában a számítógép és az ember közötti együttműködést a jelenleginél magasabb szintre emeljék. Az információvisszanyerés iránti figyelem megnőtt a könyvtáron kívül is, amely együtt jár a hagyományos információkereső nyelvek szakmai háttérének erősödő tiszteletével, elméleti alapjainak felhasználásával, pl. a gazdasági szervezetek körében is. Az ún. szemantikus vagy „intelligens web” fejlesztési irány elméleti szakemberei a portálok és egyéb weblapok tartalmának átfogó és gyors áttekintését célként megfogalmazva olyan webes nyílt és osztott keretrendszer megteremtésén fáradoznak, amely szabványosított leíró nyelveket kíván alkalmazni az információforrások leírására, és akkor éri el a teljes lehetőségét, ha az adatok megosztása automatikusan is lehetővé válik. A szoftvermérnökök bevonásával ezért új programokat és módszereket fejlesztenek, amelyek jelentős állomásairól a W3C munkacsoportjainak beszámolóí adnak hírt. A különböző hagyományos és számítógéppel kezelt tartalomfeltáró nyelvek összehangolt, egymásra épülő számítógépes fejlesztése jelentős eredményeket mutat, de az automatikus módszerek még csak néhány szoftverre jellemzők. A számítógéppel kezelt osztályozási nyelvek csoportjában a taxonómiákat és a rokon eszközöket könnyen átvehető és aránylag gyorsan létrehozható osztályozási rendszernek (kontrollált, alacsony hierarchiájú szótárnak) tekinti az informatikai rendszerek elméleti kérdéseivel foglalkozó szakemberek egy része. Írásomban ezért főként a taxonómia kérdéseire összpontosítottam.

Bevezetés

A cikket szabadkozással és köszönettel kezdem. A taxonómiákkal és a kapcsolódó szakterületekkel foglalkozó írások átnézése során napról napra bővülő, megújuló, egymásnak ellentmondó forrásokkal, felfogásokkal és nézetekkel kerültem szembe, amelyek között az eligazodás rendkívül nehéznek tűnt. Valószínűnek tartom ezért, hogy nem minden kérdésben tudom képviselni azt az álláspontot, amely az elméleti tisztázottság vagy a tudományos megalapozottság kritériumainak megfelelő. Ennek ellenére nagyon érdekesnek tartottam azokat az írásokat, amelyek alapján ez a cikk megszületett. Köszönetem *Ungváry Rudolf*nak szeretném kifejezni, aki lektori tanácsaival még ebben az utolsó fázisban is sokat segített a kérdések, meghatározások helyes értelmezésében és tisztázásában.

A szöveges és tényadatok közvetlen elérése számítógépről ma az élet minden területén természetes igény, de a megfelelő és fontos információkhoz való gyors és teljes hozzáférés akadályokba ütközhet. Még az integrált adatbázisokban is gondot jelent, ha több adatbázisban szeretnénk egyidejűleg keresni, mert nem mindig alkalmaznak egységes információkereső nyelvet. A webes források számának növekedésével és az internet nyitottságával nincs arányban az információkhoz való hozzájutás lehetősége. A különféle források rendezetlen halmaza vagy a keresőeszközök hiánya miatti sikertelenség okai a *Delphi Group* jelentése szerint a következők [1]:

- a források elérhetőségének változása (41%);
- a szegényes funkciójú keresőeszközök (26%);
- a kereső tanácstalansága, vagyis nem tudja, hogy mit is kellene keresnie (11%).

Az ismeretgazdálkodással és -menedzsmenttel foglalkozó tanulmányok közvetlen kapcsolatot mu-

tattak ki az ismeretszervezés sikeressége és a gazdasági haszon között. A tudományos és szakmai intézmények mellett a vállalatok környezetében is jelentős figyelmet és forrásokat áldoznak a „jéghegy csúcsa” alatti, kb. 80%-ot képviselő strukturálatlan információk szoftverrel támogatott feltárására. A vállalatoknál inkább a feladatorientált, gyakorlatias, számítógépes programokkal jól algoritmizálható folyamatokra és eszközökre helyeznek hangsúlyt, mint a tudományos megalapozottságra, elsősorban a gyors eredmény érdekében.

Amióta kialakult az internet, olyan (automatizáltan használható) keresnyelveket és egyéb eszközöket igyekeznek létrehozni, amelyek segítségével a teljes világháló áttekinthető. A jelenlegi törekvések arra irányulnak, hogy a portálok és egyéb weblapok tartalmának átfogó és gyors áttekintése a szemantikus web keretei között valósuljon meg. A szemantikus web közös keretrendszert kínál arra, hogy az információk megosztása és újrafelhasználása lehetővé váljon a különböző információs forrásokra épülő más-más alkalmazásokban. A szemantikus web univerzális médium az adatok cseréjére, amelynek során automatikusan, vagy emberi közreműködéssel osztják meg az információkat [2]. A szemantikus web létrehozása érdekében olyan számítástechnikai megoldásokra van szükség, amelyek segítségével szabványos formában, ugyanakkor automatizáltan lehet létrehozni a kereséshez szükséges metaadatokat. A keresést támogató fogalmi rendszerek alapszerkezetét és tartalmát RDF nyelven szabványosítja a *W3C SKOS Core Guide* dokumentum [3] a fogalmi sémák modelljének és alapstruktúrájának kidolgozásával. A fogalmi séma a fogalmak olyan gyűjteménye, amelyek opcionálisan kijelentéseket tartalmazhatnak a fogalmak közötti szemantikai rokonságokról a különböző osztályozási rendszerekre vonatkoztatva (tezaurusz, tárgyszólista, taxonómia, terminológiák és egyéb kontrollált szótárak). Ma a taxonómiák képviselik az egyik olyan osztályozási rendszert, amelyet várhatóan bevonnak majd a szemantikus web keretei között automatikusan is kezelhető szabványosított szótári rendszerek közé.*

A taxonómiák

A taxonómia és a vele párhuzamosan megjelenő egyéb, az intelligens, automatikus gépi keresést támogató webes eszközök (tudástérkép, topic map, szemantikus háló, ontológiák stb.) tárgyában végzett kutatások, kifejtett elméletek, megjelenő programok, alkalmazott technológiák és szabványok mind a tömeges rendezetlenség megszünte-

tésére hivatottak – *H. G. Wells* 1930-ban megfogalmazott „World Brain” eszméjéhez híven [4].

A taxonómia olyan ellenőrzött szótár (egyben osztályozási rendszer), amely a dokumentumokat, illetve egyéb információforrásokat az általuk képviselt fogalmak hasonlósága alapján rendezi és csoportosítja – automatikus osztályozás esetén klaszteres, nyelvi analízis, illetve numerikus, statisztikai, vagy kombinált módszerek alapján, viszonylag alacsony hierarchiaszinten –, ezért számítógépes kezelése áttekinthetőbbnek tűnik, mint a bonyolultabb struktúrákat képviselő osztályozási rendszereké.

A taxonómia definícióiban vannak eltérések, amelyek értékelésére nem térek ki, mert az meghaladná e cikk terjedelmét és a szerző felkészültségét. Ezért örömmel használtam fel Ungváry Rudolf lektori segítségét a taxonómia meghatározásához, és idézek a *TMT*-ben megjelent cikkéből is [5], amelyben az osztályozási rendszerek, főként a tezauruszok szerepét, jövőjét és a szemantikus webhez való viszonyát rendszerezte.

„Taxonómiának egyrészt azokat a hierarchikus struktúrákon alapuló osztályozási rendszereket nevezik, melyeket tudományos rendszerező céllal készítenek. A biológiában az élő szervezetek (és szervek) rendszerei taxonómiák. A tudománytanban az ismeretterületek rendszerét nevezik taxonómiának. A tanuláselméletben és a mesterséges intelligenciakutatásban is kialakultak taxonómiák és a nyelvészetben is. A mérhető különbségeken és hasonlóságokon alapuló hierarchiákat numerikus taxonómiának nevezik. Gazdasági szervezetekben véletlenszerű, hogy a kialakított hierarchikus rendező rendszert osztályozási rendszernek vagy taxonómiának nevezik. A szoftverkínálók egy része pl. automatikus és szemiautomatikus taxonómiákat kínál, más részük meg automatikus osztályozási rendszereket.

Az információkereső és osztályozási célból készült hierarchikus rendszereket általában osztályozási rendszereknek nevezik, de az átmenet cseppfolyós az adminisztráció célból készült osztályozási rendszerek/taxonómiák felé. Az áruk esetében áruosztályozásról (TEÁOR = Termékek Átfogó Osztályozási Rendszere), a foglalkozások esetében ugyancsak osztályozásról (pl. FEÁOR) szoktak beszélni, de egyes publikációkban használják (szinonim kifejezésként) olykor a taxonómia szót is. Ha értékeken alapuló adminisztratív hierarchiákat készítenek, akkor inkább a taxonómia szót használják (lásd a numerikus taxonómia esetét is).

*Ungváry Rudolf megállapítása.

Vagyis:

A tudományos, numerikus és értékalapú, hierarchián alapuló rendező rendszereket taxonómiának nevezik. Vállalati körülmények között is előfordul, hogy az alkalmazott osztályozási rendszert taxonómiának nevezik.

Az információkereső, osztályozási, besorolási célú, hierarchián alapuló (gyakorlati) rendező rendszereket általában és főleg könyvtári-dokumentációs körülmények között osztályozási rendszernek nevezik.

Elvileg, legáltalánosabb értelemben a taxonómia és a hierarchikus osztályozási rendszer kifejezések szinonimák, mindkettő hierarchián alapuló rendező rendszert jelent.

Rendező rendszer: fogalmak meghatározott célból kialakított rendszere (lehet hierarchikus taxonómia és osztályozási rendszer) és enumeratív (ilyen a felsoroló rendező rendszer, mint például egy egyszerű online mutató), továbbá összetett (ilyen a deskriptoros.)

A rendező rendszereket foglalják ellenőrzött szótárakba (tárgyszójegyzékekbe, teauruszokba, osztályozási táblázatokba, taxonómiai struktúrákba stb.).”

A Wordmap meghatározása szerint [6] a taxonómaszoftver lehetőséget ad egy szervezet számára, hogy a szervezeti információk rendezésére szabványos sémát hozzon létre. A taxonómia-eszközzel ellenőrizhetők azok a kategóriák, terminológiák, és metaadatok, amelyeket a szervezet az információs rendszerében használ(hat), elősegítve a navigáció sikerességét. Egy másik forrás leszögezi, hogy a taxonómia olyan navigációs eszköz, amelyben a szabványos metaadatok és az ellenőrzött szótár lexikai egységeinek fejlesztése kombináltan adja azt a navigációs struktúrát, amely segít a felhasználónak a szükséges tartalmat elérni [7].

Összességében a következő megállapítások jellemzők [8]:

- a taxonómia webes környezetben biztosítja az információforrások, illetve dokumentumok csoportba rendezését hasonlóságaik alapján a tartalomfeltárás céljára;
- szisztematikus és szintetikus osztályozási eszköz a tudás-, illetve ismeretrepresentáció céljára, hierarchikus struktúrában tartalmazza a fogalmakat, és terminusai a tudásmegosztás nézőpontjából fejlesztik az adott közös nyelvet.

A szemantikus web napjainkban még nem valósult meg, mert sem a technológiák, sem a szoftverek,

sem az ösztönző társadalmi feltételek nem állnak általánosan rendelkezésre. Ha jelenlegi tudásunk szerint egy minden ismeretet és információt rendezni képes egységes fogalmi rendszer és annak formalizált leírása a struktúrák és a szintaktikai kapcsolatok feltárásával még nem jött, illetve jöhet létre a webes tartalmak rendezésére, akkor meg kell valósítani a különböző területeken létrejött és szabványosított nyelven leírt tartalomfeltáró eszközök elemzését, átvételének és egyesítésének feltételrendszerét – több szakmai terület összefogásával. Ezt a célt szolgálják a W3C azon törekvései, hogy elméleti szakembereket, programozókat vonjon be az RDF és OWL nyelv alkalmazásába és fejlesztésébe, valamint azokba a kutatásokba, munkacsoportokba és az ún. *jó gyakorlatok* (best practices) elterjesztésébe, amelyek a keresőnyelvek és ontológiák, a szabványok és leírónyelvek kialakítását, valamint a keresés gyakorlati szempontjainak figyelembevételét célozzák.

A taxonómiai megközelítési mód

A taxonómia szintetikus szemlélet alapján kialakított információkereső nyelv, amelynek készítésekor egy meghatározott felhasználói környezet által használt és elfogadott fogalomrendszert építenek ki kapcsolatainak feltárásával együtt, és ezekhez a fogalmakhoz rendezik az információforrásokat, dokumentumokat, karaktersorozatokat (adatokat) vagy ugrópontokat, amelyek természetesen dokumentumhoz is, és közvetlenül információkhoz is vezethetnek, például egy szolgáltatás fogalmi meghatározásához. Gazdasági szervezeteknél nem mindig jellemző a tudományos vagy ismeretterületi megalapozottság. Gyakran egy adott közösség által kezelt tudás összetételét, és az ott alkalmazott fogalmi megnevezéseket veszik figyelembe, amelyek eltérő értelmezést hordozhatnak más területekhez, vagy a szaknyelvhez viszonyítva. A taxonómia akkor valósítható meg jobban, ha a fogalmak alapvetően generikus hierarchiát alkotó osztályok rendszerében helyezkednek el, és az egyedi fogalmak és osztályok között kiterjedt, de logikailag egyszerű kapcsolatrendszer van.

A taxonómia környezete és háttere

- Taxonómiát manuális, automatikus és szemi-automatikus módon lehet létrehozni – vállalati környezetben elsősorban elektronikus információforrások (portálok, dokumentumkezelő rendszerek segítségével tárolt információk stb.) keresésének céljából. Vannak már példák a taxonómiák szoftverrel támogatott kialakítására is;

- egy közösség, amely használja és elfogadja a konszenzuson alapuló megnevezéseket, vagy amelyeknek üzleti nyelvét a taxonómia megfelelően formalizálja;
- egyéb osztályozási eszközök (tezauruszok, szótárak, indexek, tárgyszókészletek stb.), amelyek részeit, fogalmi struktúráját a taxonómia kialakításakor célszerű felhasználni, átvenni;
- folyamatok, szolgáltatások stb. jól meghatározott leírásának gyűjteménye. Gazdasági szervezetnél, például a folyamatleírások általános dokumentumai auditált szakkifejezéseket tartalmaznak, és tartalomjegyzékeik, fejezetcímeik adhatják azt az általános rendezési háttérrel, amelynek alapján a kiválasztott fogalmak struktúrába rendezhetők:
 - Értékesítés
 - Beszerzés
 - Marketing
 - Reklám
 - Fejlesztés stb.

A taxonómia nyelve – gyakorlati szempontok – egyezményes megnevezések

A taxonómiákat vállalati környezetben gyakran nevezik a tudásmenedzsment (TM) közösség által fejlesztett tudásfeltárás (tudásreprezentálás) explicit eszközeinek. A taxonómia olyan terminusokat tartalmaz, amelyek a tudásmegosztás nézőpontjából megfelelnek adott környezet közös nyelvének. Nem szükséges egy tudományág teljes fogalmi reprezentációját képviselnie, de kialakításakor nélkülözhetetlen segítséget jelent a fogalmi elemzés. Gazdasági szervezeteknél az üzleti vagy szakmai tevékenység (misszió) a vezérelv, amelynél a legfelső kategória a misszióra vonatkozó általános megnevezés, és ezt követik olyan általános és gyűjtőfogalmak, mint: szolgáltatás, marketing, termékek stb. A megnevezések kodifikálását a TM-közösség vagy szerkesztőbizottság végzi egy tesztcsoport közreműködésével. A tesztelés társult eredménye a megnevezések „megtanulása” is. A konszenzuson alapuló megnevezések eltérések is lehetnek az általánosan használtaktól – a fogalmi definícióban szerepelnek az értelmezés helyi sajátosságai [9].

Az egységesség és az egyesítés kérdései

A taxonómiák készítésénél ésszerű feltérképezni az előzményeket annak érdekében, hogy vannak-e részben vagy teljesen átvehető eszközök. Ezért

jelentős eredmény a már létező osztályozási rendszerek feltérképezése a legfontosabb jellemzők feltárásával, amelyre már vannak jó kezdeményezések. Ezek közül egy példa (1. ábra) a *Dow Jones Reuters Business Interactive LLC (Factiva)* weboldaláról [10].

Az XML-ben tárolt taxonómia letölthető, amennyiben erre nézve a felek megállapodnak az átvételről. Arról, hogy az osztályozási rendszerek megnevezése (tezaurusz, taxonómia stb.) mennyiben felel meg a valóságnak ebben a rendszerezésben, nem mernék állást foglalni. A létező eszközök különféle technológiák alapján jönnek létre, frissítésük gyakorisága lehet rendszeres, rendszertelen vagy ismeretlen.

Az egységesség elve és a számítástechnikai eszközökkel történő átvételre való törekvés a következő kérdések átgondolását teszi szükségessé [11] a szemantikus web célja szempontjából. Hogyan lehet:

- heterogén információforrásokat egységes elvek és különböző technológiák alapján szemantikai szempontból elemezni;
- egyértelmű, gyakorlatias, helyhez, szakmához, környezethez illeszkedő egyeztetett megnevezéseket létrehozni és alkalmazni;
- dokumentumalapú megközelítés helyett/mellett ismeretalapú módon felderíteni a weben, a portálokon lévő, és a vállalati rendszerekben tárolt strukturálatlan elektronikus információforrásokat, tudáshelyeket, számítógéppel kezelt osztályozási eszközöket egységes elvek alapján fejleszteni;
- a különböző tartalomfeltáró eszközök által képviselt általános fogalmak hálózatával a teljes web-tartalmat lefedni (a teljes web tartalma egy kb. 130 000 lexikai egységből álló, egyetemes és átfogó ellenőrzött szótárral lenne lefedhető*)?
- Milyen típusú informatikai rendszerekben kell gondolkodni, és milyen eszközök állnak jelenleg és a közeli jövőben rendelkezésre az intelligens számítógépes keresésre, és hogyan kapcsolódhatnak egymáshoz?
- A szemantikai elemzés szoftvereszközeinek fejlesztéséhez milyen előkészületek szükségesek a szoftvermérnökök, a szabványalkotó testületek és más szakemberek részéről?
- Hogyan fog kialakulni az összefogás a távlati célokban megfogalmazott szemantikus web lét-rejöttéhez?

*Ungváry Rudolf véleménye.

Name:	Factiva Intelligent Indexing Taxonomy Feed
Publisher:	Factiva, a Dow Jones & Reuters Company (Subject Taxonomies and Company Codes)
Type:	Taxonomy
Categories:	General Business, Geographic Names, NEWS & CURRENT EVENTS
Description:	The Factiva Intelligent Indexing business taxonomy. Consists of an XML feed of Industry, Region and Subject codes. Hierarchical tree-structure; labeling of terms; definition of terms; multiple alias names; alternative language versions. Associated example document collections to assist with automatic categorization.
Total Terms:	1500
Top Terms:	
Preferred Terms:	1500
Non-Preferred Terms:	
Relationships:	
Levels:	7
Notation Scheme:	Yes
Notation Description:	Proprietary Factiva alphanumeric scheme
Relationship Types:	Equivalency, Hierarchical
Notes Fields:	Definition, Other
Multilingual:	Yes
Languages:	Chinese, English, French, German, Italian, Japanese, Russian, Spanish (European)
Additional Information:	Comprehensive business taxonomy as used in Factiva's information services. Taxonomy customization assistance available on request. Associated example document collections to assist with automatic categorization are available in English, French, German and Spanish.
Revision Cycle:	Quarterly
Last Updated:	200305
Formats:	Download - XML
Informational URL:	http://www.factiva.com/content/intindexing/codes.asp?node=menuElem1108
Online/Download URL:	Sample data available on request

1. ábra Példa a Dow Jones Reuters Business „Factiva” weboldaláról

Rendkívül izgalmas, és több mint 30 éve (a mesterséges intelligencia kutatásának kezdetétől) fontos és megválaszolatlan kérdés az, hogy vajon ténylegesen leírható-e a világ gazdagsága a formalizált, algoritmus alapján működő rendszerben? A jövő ugyanis nem azokat a megoldásokat hozza, amelyeket jelen tudásunk alapján lehetségesnek tartunk – mégis számos vízió megvalósult a múltból. A szemantikus web küldetését *Tim Berners-Lee* 2001-ben fogalmazta meg [12]. A közeljövőben talán várhatjuk, hogy a kitűzött célok szerint kialakult széles körű összefogás – az elméleti és gyakorlati szakemberek részvételével – meghozza a várt eredményt.

A taxonómia szerkezete

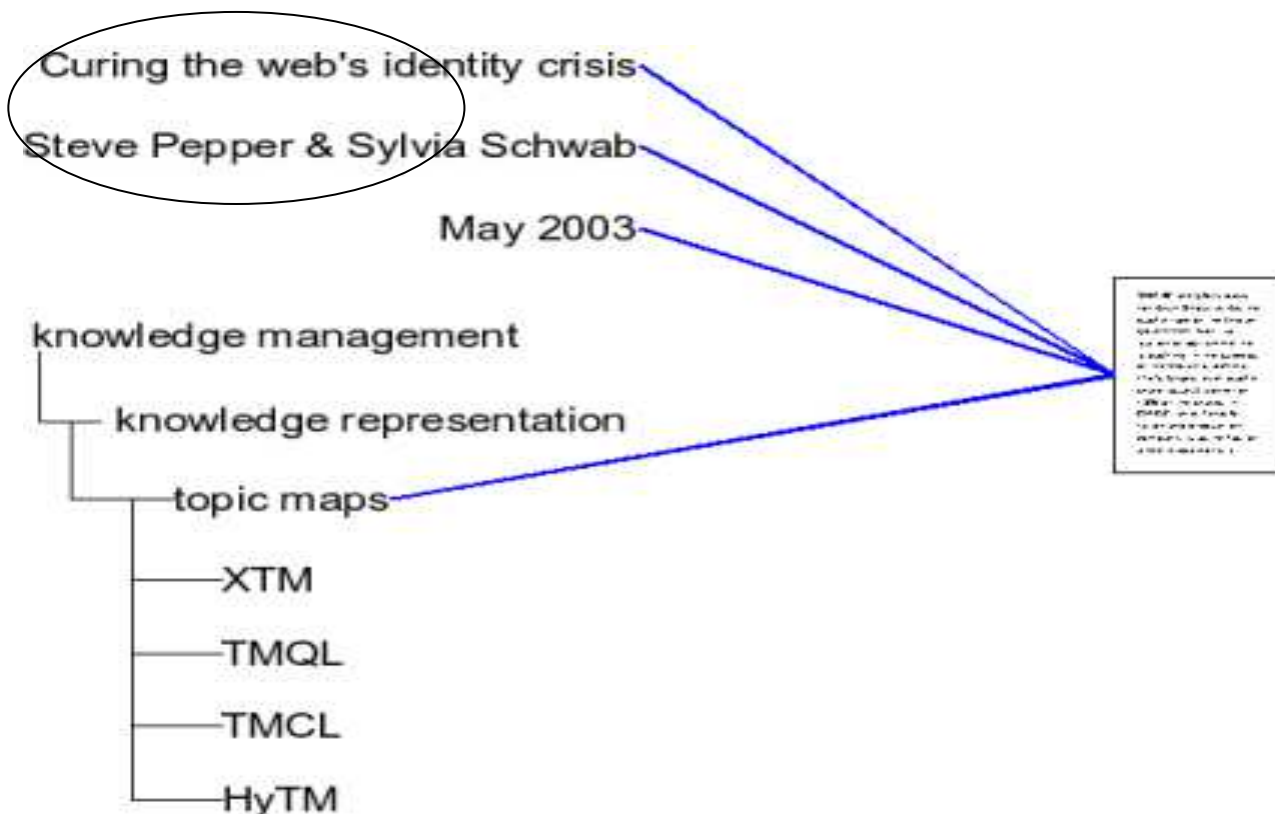
A taxonómia hierarchikus szerkezetében [13] a létrehozott struktúra nem szótárszerű összefüggésekből áll. A megnevezések sorrendje nem index-szerű, hanem szülő/gyerek viszonyt tükröz, és nem nevezi meg a rokonsági kapcsolatok típusait:

- tágabb,
- szűkebb,
- rokon.

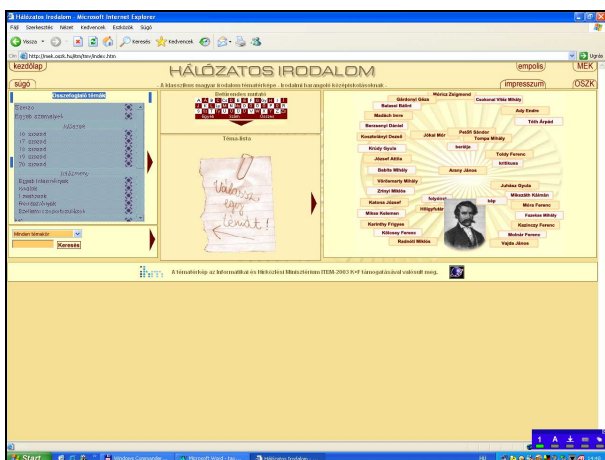
Egyéb leíró adatok

- Jelentéstani definíció, amely jelentős lehet az általánosan elfogadott vagy használt értelmezésektől eltérő esetekben;
- ún. referenciadokumentumok egy csoportja, amelynek tartalmában a taxonómia fogalmai hasonlóképpen és magas szinten vannak képviselve (2. ábra). A fogalom értelmezésének változása esetén a változást képviselő dokumentum(ok) hozzáadódnak a referenciadokumentumokhoz (ez a „forrás” stb. adat, melyet a MARC formátumokban szabványosítottak, és az OWL ugyancsak kezelni tudja).

A definícióhoz szerző és dátum is tartozik, amely frissítésnél kiegészül (ugyancsak adatmezők a MARC formátumban). Az egyezményes megnevezés és értelmezés két fontos eszköze a fogalmi definíció és a fogalmat leginkább képviselő dokumentumok együttese. A referenciadokumentumoknak az értelmezésben is, de főként az automatikus keresést támogató dinamikus taxonómiák esetében van fontos szerepe, amelyeknél a hasonlóságokat gépi eszközökkel, klaszteranalízissel, vagy statisztikai módszerekkel, illetve kombináltan is támogatják.



2. ábra Egyszerű taxonómaszerkezet a referenciadokumentum metaadataival és szövegével. A „topic map” = tématerkép rokonságai a rokonság megjelölése nélkül szerepelnek (XTM, TMQL stb.) [14]



3. ábra A Neumann-ház tématerképe (az OLIB rendszer teaurusz moduljából XML-be töltött adatok felhasználásával készült) [34]

A szükséges eszközök

- Taxonómia management system a taxonómia-építéshez. Integrált könyvtári rendszerek is alkalmasak az építésre, ha van teauruszkezelő moduljuk (ha csak a generikus relációt használják az F/A [MSZ 3418] relációkkal – de a ta-

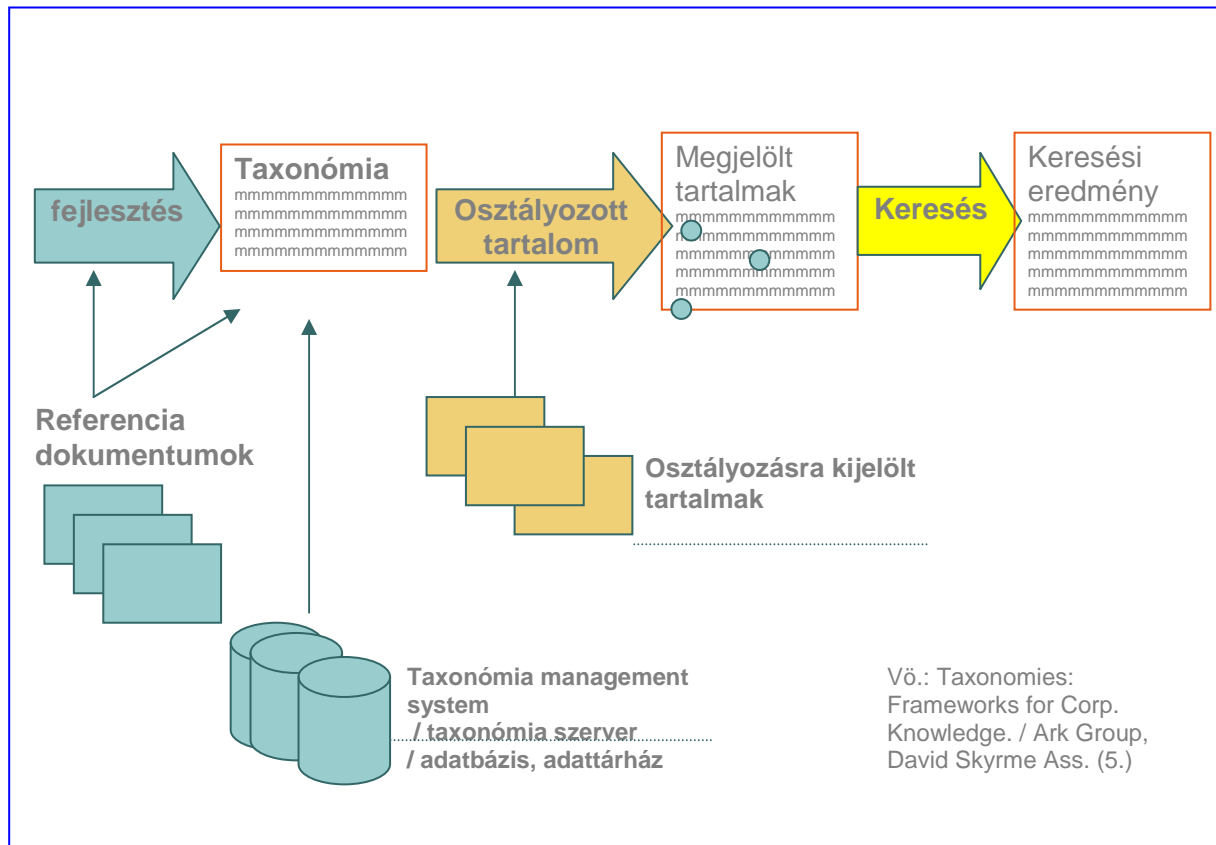
xonómia ettől még nem teaurusz és fordítva), amelyből az adatok XML fájlba kimenthetők. Lásd a *Neumann-ház* tématerkép megoldását (3. ábra), amely az OLIB teauruszkezelőből kinyert témacsoportok felhasználásával készült.

- Taxonómaszerver, amelyről minden használó számára megoldották az elérést.
- Tartalmakat kezelő valamely tartalommenedzsment (CMS) szoftver, adatbázis-kezelő, adattár, XML-ben (stb.) tárolt szöveges állomány, adattárház stb.

A taxonómia működési sémáját a 4. ábra szemlélteti [15].

A taxonómia dimenziói és a fazetták

A taxonómiának mint ellenőrzött szótárnak, illetve hierarchikusan strukturált fogalmi rendszernek több dimenziója van, mint amennyit a három fizikai kiterjedés lehetővé tesz, ezért alkalmaz állandó metszeteket (fazettákat, fogalmi mátrixokat) az azonosan és gyakran előforduló strukturális jellemzőkre. A fazettával olyan megnevezést használunk, amely egy dolgot az összes többitől meg-



4. ábra A taxonómia működési sémája

különböztet: mindez *Ranganathan* osztályozási elméletén alapul [16]. A közös elemek adott felhasználói környezetre jellemzők, ezért a szervezeteknek önállóan kell létrehozni őket. A fazetták jól algoritmizálható szerkezetet kínálnak az ismétlődő elemeknek, és jól szolgálják a fogalmi struktúra több dimenziójának feltárását. Üzleti intelligenciaszolgáltatás megvalósítása esetén különösen fontos metszetek adhatók például egy szolgáltatás környezetében: foglalkozás, életkor, vásárlási, szórakozási, internethasználati szokások, adó- és hiteltartozás stb.

Az automatikus osztályozás és fogalomalkotás kérdése

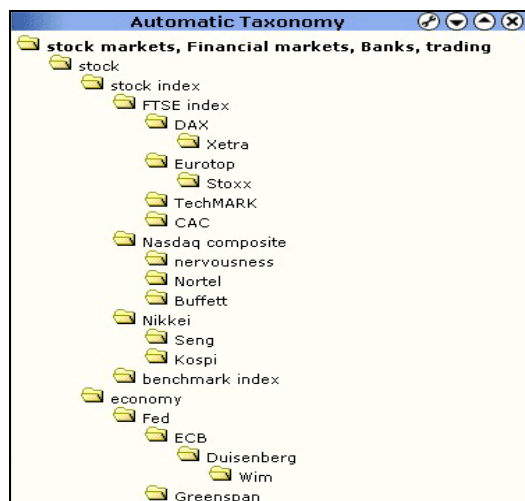
Az automatikus osztályozásnál algoritmusokat használnak a szoftver segítségével végzett analízisre, az adatbányászathoz és egyéb, az üzleti intelligencia keretébe tartozó megoldásokhoz hasonlóan. Az automatizálás nemcsak a keresés, hanem a kategóriák létrehozása szempontjából is lehetséges. A hasonló tartalmak kiválasztásának alapját az egyes taxonómiamegnevezésekhez kapcsolt ún. referencia- vagy mintadokumentumok adják,

az általuk képviselt fogalmakhoz viszonyítják egy másik információforrás tartalmát a klaszter-, a statisztikai vagy a kombinált elemzés alapján. A vizsgálat nemcsak szöveges formában történhet, hanem bináris kódok alapján is, amelynek során vizsgálják a megnevezések szöveggörnyezetét is. Példaként az *Autonomy rendszer* kombinált módszereit említhetjük. A kérdéseket természetes nyelven lehet megadni, az analízis az ún. *Bayes* matematikai, valamint statisztikai módszerek, és bináris kódok alapján történik (5. ábra) [17].

Automatikus kategorizálást támogató taxonómatechnológiák

- Példa alapján tanuló rendszer – a kategóriákat szerkesztő határozza meg –, később a *training set* támogatással valósul meg a rendszer felkészülése a pontos keresésre.
- Szabályalapú – a dokumentumok osztályozása speciális üzleti szabályok alapján történik.
- Statisztikai módszerek – a kifejezéseket az előfordulási gyakoriság, és egyéb kifejezésekhez, szavakhoz való kapcsolatuk alapján határozzák meg.

- Természetes nyelvi folyamat (konceptuális nyelvészet) az osztályozás kerete egy nagy kiterjedésű extenzív szótár és tezaurusz a fogalmak azonosításához.
- A fentiek kombinációi.



5. ábra **Autonomy rendszerben előállított automatikus taxonómia**
(a fogalmak az üzleti funkció, és nem a fogalmak elméleti összefüggései szerint rendezettek)

A vizualizáció kérdése

A vizualizáció nem új keletű megoldás, hiszen a tezauruszok is megjelentek grafikus ábrázolásban. A vizualizáció akkor nyer új jelentőséget, ha a vizuális elemekre kattintva közvetlenül eljuthatunk egy információforráshoz. Ilyen jellegű vizualizációt nyújtanak a klasztermegoldások, például az Autonomy esetében az alábbi módon:

- Színnel van jelölve a legkeresettebb dokumentumokat képviselő vektorhely, és adott színre kattintva megjelennek ezeknek a dokumentumoknak a metaadatai.
- Színnel vannak jelölve az adott fogalmat leginkább képviselő dokumentumok helyei, amelyek a színre kattintva szintén megjelennek.

Az egyéb vizualizációs lehetőségekre (hierarchikus faszervezetek és fogalmi térképek) a tudástérkép és tématerkép ismertetésekor térek ki, mindkét eszköz alkalmas megoldás a taxonómia ábrázolására oly módon, hogy annak legfelső kategóriái alapján történik a megjelenítés [18]. A különböző „térkép” megoldások a gyors áttekintést és az általánosítást támogatják, amelynek szintjét az áttekintés horizontjának függvényében célszerű meghatározni.

Ontológia és taxonómia

Az ontológia vállalati vagy intézményi környezetre alkalmazott értelmezése ma a megegyezésen alapuló fogalmi rendszer formális, egyértelmű leírása számítógépes használatra. A kérdés teljes áttekintéséhez Ungváry Rudolf már idézett cikkét ajánlom [19]. A „megegyezésen alapuló” kitétel fontos, azt a felfogást tükrözi, hogy az ontológiák szemantikai szabályrendszerek, a dolgok értelmezésére, rendezésére és keresésére szolgálnak, figyelembe veszik a felhasználói szempont érvényesülését. *Tárgyunk szempontjából az eredetileg tudományt (lételméletet) jelentő kifejezés nem a hierarchikus rendszert jelenti önmagában, hanem annak szoftverrel kezelhető, elsőfokú logikai kijelentésekkel bővített változatát. Távol áll a technikai zsargonról, vagy a mesterséges intelligencia szabályalapú rendszereitől, az ontológia fogalmi rendszer arra szolgál, hogy a webet saját, algoritmizálható intelligenciával ruházzuk fel* [20].

A webnek ez az intelligens képessége arra szolgál, hogy a szövegek keresése, értelmezése mellett bizonyos következtetéseket is létre lehessen hozni automatikus eszközökkel. Az „ismeret-technológia” (tudástechnológia, *knowledge engineering*) nézőpontjából tehát az ontológia meghatározott ismeretterület számítástechnikailag kezelhető, formális modellje. Az így értelmezett rendszer esetében helyes az ontológia megnevezés [21].

Az ontológia és a taxonómia mint a tudásreprezentáció és tudáskezelés eszköze

A tudáskezelés, tudásgazdálkodás fontos kérdése, hogy milyen hatékonysággal, milyen mélységben és terjedelemben tudjuk megoldani a létező „tacit és explicit” információk megszerzését, újrafelhasználását és adaptálását. Ha azonos dolgokat különbözőképpen nevezünk meg, és ezek között nincs meghatározva szabványosított és automatikusan létrejövő kapcsolatrendszer, vagy ha van, de nem ismerjük, vagy nem tudjuk használni, nem értjük, vagy nehézkesnek tartjuk, akkor hiányos, hosszadalmas vagy eredménytelen lesz az információ-elérés, amely akadályozza az ismeretszerzést.

A tudásmenedzsment szempontjából az ontológia „explicit” magyarázatot ad a fogalomalkotáshoz, ami azt jelenti, hogy egy alkalmazási környezetben megfogalmazott, kodifikált jelentésről van szó, amelynek háttérében egy formális logikai elemzés és egy elfogadási procedúra áll. Létrehozása so-

rán meghatározzák azokat a dolgokat, amelyek egyértelmű jelentést hordoznak, és szükségesek egy adott kontextusban szereplő tudás reprezentációjára.

Ontológia a számítástechnikában

A számítástechnikában az ontológia arra szolgál, hogy szigorú fogalmi rendszer alapján megkísérelje az adott doménhez rendelt fogalmakat egy hierarchikus adatstruktúrában definiálni úgy, hogy az tartalmazzon minden releváns entitást, kapcsolataikkal és szerepeikkel együtt. Kialakulása a mesterséges intelligencia szabályalapú szakértői rendszereitől, és a számítógépes nyelv programozott használatából származtatható [22]. Kulcskérdés, hogy milyen automatikus eszközök állnak rendelkezésre a tudásfeltáráshoz (pl. tartalomfeltárási eszközök, adattárház, adatbányászat, üzleti intelligencia, automatikus kategorizálás, természetes nyelvi keresés). Az ontológiák létrejötte alapvető előfeltétele az internet következő fejlődési irányának, vagyis a szemantikus web kialakításának, mert a fogalmi állandóság segítséget jelent a különböző számítástechnikai rendszerek összekapcsolásában. A szakterületi tudás/ismeret fogalomrendszerének megjelenítése, publikálása és kodifikálása az a lépés, amellyel az ontológiák az informatika talaján előrelépést hoztak, „ezért jelenleg talán az ontológiák jutottak a legközelebb ahhoz, hogy mi is lehet az emberi tudás (fogalmi szintű) részének a tartalma és szerkezete” [23].

A szemantikus web fejlődési irány

A pénzigényes, és jó üzletnek számító tartalommenedzsment kérdéseiben a könyvtári szakértelem felhasználását nem tekinthetjük megfelelőnek; emellett a könyvtári szolgáltatásokban is megjelentek a teljes szövegű, strukturálatlan információforrások keresésének támogatására épülő analízis és szintetizáló szolgáltatások és szoftverigények. Azzal a problémával kell szembenézni, hogy nagy tömegű elektronikus információt sem indexelni, sem visszakeresni nem lehet hatékonyan manuális eszközökkel. A számítástechnikai kutatások az 1950-es évek közepétől, de főként a 70-es évek elejétől foglalkoznak a szövegek gépi olvasásának, fordításának, indexelésének kérdésével, és a mesterséges intelligencia kialakításának igényével. A kutatások egyelőre nem hozták meg a kívánt eredményt, viszont rávilágítottak arra, hogy a web tartalmát egyezményes és automatizálható eszközökkel kell feltárni és rendezni.

A szemantikus web rendeltetése: olyan nyílt, osztott környezet létrehozása, amely nyelveken alapul (metaadatok, tezaurusz, taxonómia, ontológia, tématerkép, tudástérkép stb.), és amely akkor éri el a teljes lehetőségét, ha létrejön az a környezet, amelynél az adatok olyan mértékben lesznek géppel és automatikusan megoszthatók, mintha azt emberek végeznék.

A W3C megfogalmazásában [24]: A szemantikus web közös keret arra, hogy az adatok elosztottan, és az egyes alkalmazások által újra felhasználhatóan, vállalati szinten vagy közösség számára elérhető legyenek. Ebben a fejlesztésben a W3C sok kutatóval és ipari partnerrel működik együtt. A kialakítás az RDF nyelven (*Resource Description Framework*) alapul, amely XML használatával integrálja a nagyszámú, különböző alkalmazásokat a szintaxis összefüggéseinek tárolásával és az URI (*Uniform Resource Identifier*) forrásazonosító rögzítésével.

A szemantikus web – szabályozók és ontológiai nyelvek

A szemantikus webhez tartozó ontológia legfontosabb szabályozója az OWL szabványcsalád (*Web Ontológia Nyelv*), amely arra szolgál, hogy leírjuk vele a web dokumentumaiban és alkalmazásaiban előforduló lényeges kapcsolatokat, vagyis az OWL formalizálja a webtartalmakat a következő séma alapján [25]:

- formalizálunk egy tématerületet osztályok és tulajdonságai definiálása alapján;
- egyedeket határozunk meg, és tulajdonságaikról ún. kijelentéseket teszünk;
- többletinformációkat következtetünk ki ezekről az osztályokról és egyedekről, amilyen mértékben ezt az ontológia nyelv formális szemantikája lehetővé teszi.

Az OWL nyelv néhány fontos eleme

- A Web Ontológia Nyelv az RDF nyelven alapul, amely meghatározza a metaadatokat és a webforrások leírását, lehetőséget ad az adatállományok közötti kapcsolatok jelölésére, amely a szemantikus web kialakításában fontos feltétel. Az RDF lehetővé teszi, hogy egy adat kapcsolódhasson egy másik RDF-ben tárolt adathoz, vagyis az adat megosztása a weben decentralizált módon jön létre.
- Az RDF sémanyelv (kapcsolatok tipizálására szolgál).

- A SPARQL (*sparkle*) standardizált keresőnyelv a core-komponensekhez, amely létrehozza a kapcsolatot az RDF adatok gyűjteményében.

Összefoglalva: az „OWL igyekszik ötvözni az informatika jelenlegi szintjén az ember két fontos mentális képességét: a tudásreprezentációt és a következtetés képességét” [26]. A szemantikus web legnagyobb kihívása, hogy olyan nyelvek, szabványok, programok létrehozására van szükség, amelyek az adatokat és a szabályokat is létre tudják hozni a létező tudásreprezentációs rendszerben, és ezeket exportálni tudják a webre.

A szemantikus web információfeltáró eszközei

A hagyományos (automatizálható) osztályozási rendszerek, a taxonómiák és ontológiák mellett léteznek olyan eszközök, amelyek az intelligens webes keresést egyéb módon és egymással némi átfedésben támogatják, például akár vizuális eszközökkel vagy ugrópontkapcsolatok beépítésével.

Szemantikus háló

A szemantikus háló az ontológia megjelenítése adott felhasználói környezetben. A gráf csúcsa a fogalmak reprezentációja, a rácsok pedig a kapcsolatokat és az asszociációkat mutatják meg a fogalmak között. Természetes ábrázolási mód, ha a következtetés az elemek valamilyen rendszerén alapul (taxonómia), mert kifejezi a hierarchiát és az objektumok közötti távolságot. A szemantikus háló grafikusán ábrázolja az objektumokat, jellemzőiket és a köztük fennálló függéseket [27]. A szemantikus háló általában olyan problémák reprezentálására alkalmas, amelyekben

- az univerzum szerkezetére valamilyen taxonómikus hierarchia jellemző, vagyis az egyedek alafőlé rendelt osztályok rendszerében helyezkednek el;
- az univerzum egyedei és osztályai között kiterjedt, de logikailag egyszerű kapcsolatrendszer áll fenn.

A szemantikus hálókat eredetileg *Quillian* fejlesztette ki az 1970-es évek végén angol mondatok jelentésének leírására. Célja olyan reprezentáció kidolgozása volt, amellyel két szó összehasonlítását, szembeállítását, valamint egy szöveg jelentésének meghatározását az emberi gondolkodáshoz hasonló módon lehetne megtenni. A természetes

nyelv megértésének során a lexikai egységek ábrázolására alkalmazták [28]. A szemantikus háló ugrópont a jelentés típusai között, és feltárja az ismeretek fontos kapcsolatainak hálózatát. A háló lehetővé teszi a generikus kapcsolatokat, vagyis az osztálytulajdonságok öröklését.

A grafikus ábrázolás elemei:

- dolgok, objektumok, dokumentumok,
- jellemzőik,
- közöttük fennálló összefüggések.

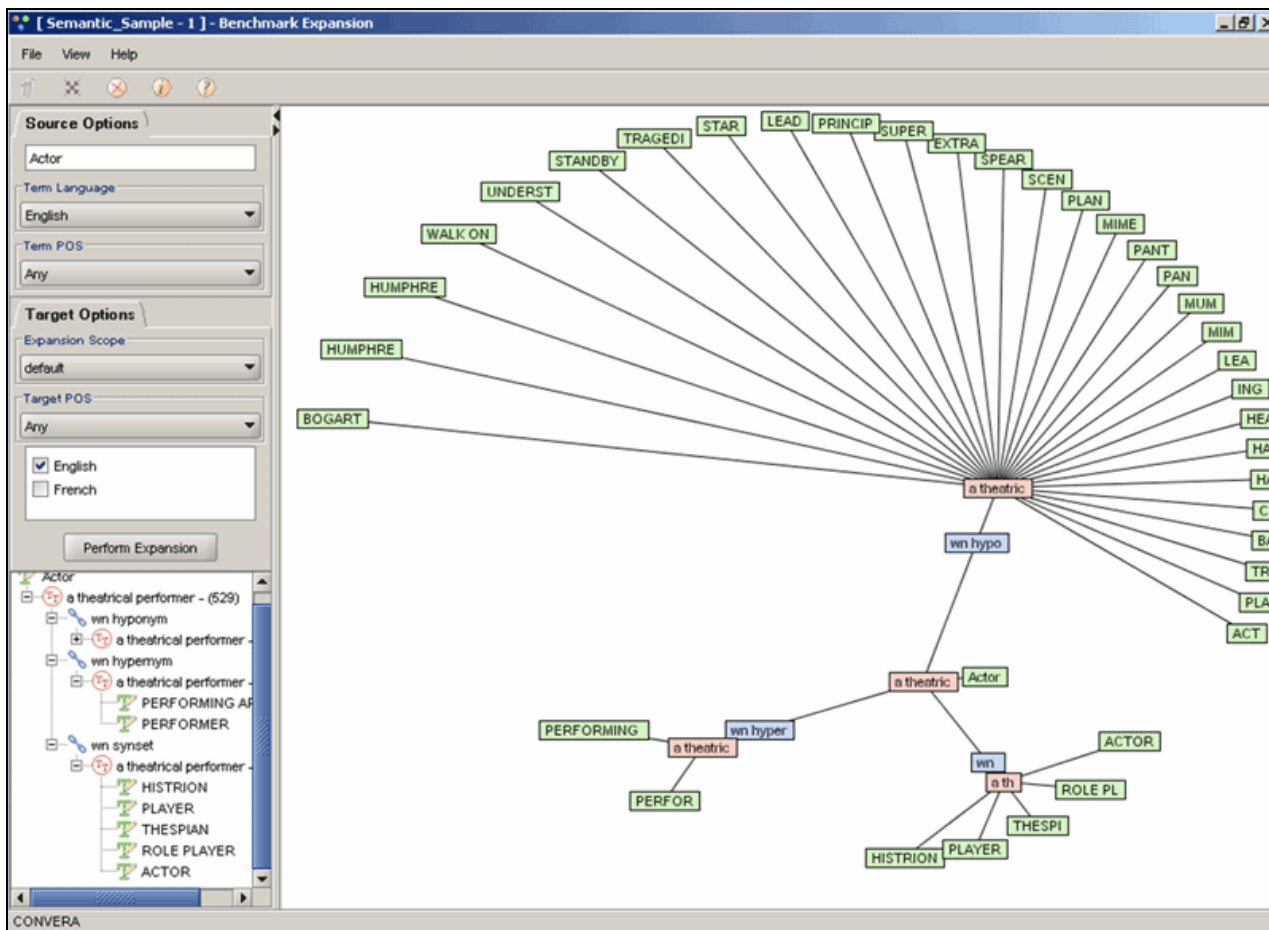
A szemantikus háló tágran használt eszköze a strukturált tudás megjelenítésének (6. ábra). Az ontológia gyakran jelenik meg a doménismeret megjelenítéseként szemantikus háló formájában. A szemantikus háló alapvető sajátossága az osztálytulajdonságok generikus öröklődése, és emiatt nem biztos, hogy egy objektum valamennyi bizonyítandó tulajdonságát közvetlenül hozzákapcsolva találjuk a tényhálóban. Lehetséges, hogy a szóban forgó tulajdonság az objektum osztályának, vagy még bővebb befoglaló halmazának általános tulajdonsága, és ezért a hierarchia magasabb szintjén találjuk meg [30].

Tudástérkép

A tudástérkép intellektuális infrastruktúrát szolgáltat a tudásmenedzsment lépéseihez; keretet ad az információ-visszakeresés céljára alkalmazott ontológiáknak [31], vizuális és konceptuális kapcsolatot teremt a tartalmak eléréséhez. Nemcsak szövegek, hanem más térképelemekhez hasonlóan szimbólumok, útvonalak, irányok stb. tartozhatnak hozzá. Célja a navigáció támogatása az egyes tudáshelyek, tudásgazdák, tudásbázisok között, másrészt biztosítani azt a mechanizmust, amelyet a térkép új fogalmak beépítése során használ a szoftverbe épített algoritmusok alapján. Kevésbé formalizált, és flexibilisebb, mint a taxonómia és az ontológia. Veszélyt jelenthet a tudástérkép és az általa lefedett működési terület összemosása.

Topic map – tématerkép

A tématerképnél minden téma valójában egy szinonimacsoportot képez, amelyet egyetlen megnevezés képvisel. A tárgyi osztályozás technikáján alapul a témahely tárgyának formális deklarálására és azonosítására. A tématerkép megfordítja az általunk ismert információfeltáró folyamatot; itt nem a dokumentumból, hanem a témától jutunk el az objektumhoz, az információforráshoz.

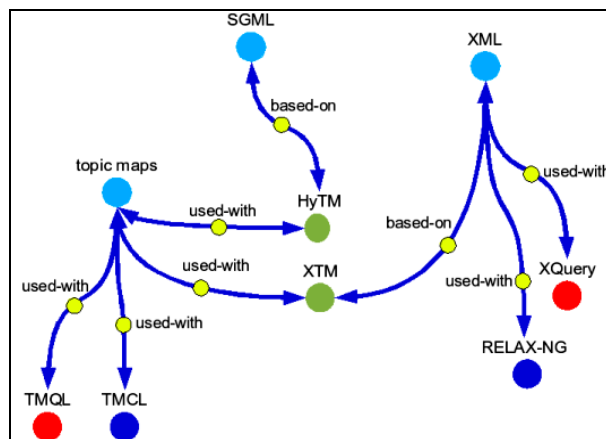


6. ábra Példa a szemantikus hálóra a Converától [29]

Ezért szerkezetében az alábbi jellemzők szerepelnek:

- téma (topic) – a neve azonosítja, amely több részből állhat;
- kapcsolatai (al- és fölérendelt);
- előfordulásai.

Szabályalapú tudásmenedzsment eszköz, amely leírja a téma tartalmát, és kifejezi a köztük lévő rokonságokat. Hasonlóan a szemantikus webhez, képes témahelyek közötti meghatározott kapcsolatokra épülő automatikus akciók generálására. Szerkesztése a taxonómiák felső szintje, illetve azok kapcsolatai és előfordulásai szerint történik. Kapcsolódik az osztályozási rendszerekhez, emellett felhasználja a tezaszusz szemantikai ugrópontjait és keresési módszereit is. A topic map kifejezés kapcsolatai egy tématerkép megoldásnál a 7. ábrán láthatók (hozzátartozik még egy dokumentum is, amely vagy képviseli, vagy meghatározza az ismeretet) [32].



7. ábra A topic map kifejezés kapcsolatai egy tématerkép megoldásnál

A megnevezés ad egy helyet (scope), ahol megjelenik a téma egy halmaza, amely egy tartalmat képvisel. A megnevezések lehetnek azonos alakúak is, azonban a típus, az előfordulás és a kapcsolat pontos értelmezést ad a megnevezésnek, pl.

Paris (mitológiai alak); Paris (város). A tématerkép a többnyelvű információk szolgáltatását is támogatja, a felhasználó saját nyelvén választhatja ki megnevezést, és a rendszer nem arra figyelmezteti, hogy egy másik, preferált kifejezést alkalmazzon, hanem belső kapcsolatai alapján „érti” a kérdést. A tématerkép háttérében a tudásintegrációra fejlesztett ISO szabvány áll [33].

Kitekintés és konklúzió

Ha az információs társadalom új elnevezéseinek megfelelően (a tudás társadalma, tudásközpontok, e-kormányzat, e-learning stb.) a fejlődés egyik központi kérdése marad az információs esélyegyenlőség, akkor a szemantikus web megvalósítása, a világ ontológiai szemléletű leírása kiemelt szerepet fog kapni a rendszerfejlesztők és a szoftvermérnökök programozási feladatai között, mert a szemantikus webnek a következő internetgeneráció részévé kell válnia, ahol a hálózat „érti” a szavak jelentését és a parancsokat, automatizálva állítja elő a tudásbázisokat és azok változásait az ontológiák beépített kapcsolatai alapján. A könyvtárosok által használt osztályozási rendszerekre felfigyelt a számítástechnikai szakértők közössége, és az osztályozási tevékenység is olyan szakmai minőségű, amely sem a számítástechnikusok, sem a vállalati tudásszervezők napi gyakorlatában nem honos, ezért e téren létrejöhetnek új típusú együttműködések. Kérdés viszont, hogy a szemantikus web automatizált folyamatai keretében kezelhető-e a komplex, félreérthető és irracionális mindennapi valóság? „Az okos ember jóval több információt lát meg az érzékszervi információkban, mint amennyi közvetlenül látszik, mert össze tudja kapcsolni korábbi ismereteivel” [35]. Ezért nagyon nehéz olyan automatikus módszereket alkalmazni, amelyek nem konkrét, hanem elvont információk formális leírása mellett teszik lehetővé az intelligens keresést.

Jegyzetek és irodalom

- [1] WYLLIE, Jan–SKYRME, David: Taxonomies: frameworks for corporate knowledge. - London, Ark Group, 2005. p. 53.; Taxonomy & content classification market milestone. Report. White paper. http://www.delphigroup.com/research/whitepapers/WP_2002_TAXONOMY.PDF
- [2] <http://www.w3.org/2001/sw/>; <http://www.consortiuminfo.org/bulletins/semanticweb.php>
- [3] <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20050510/>
- [4] L. [1], p. 130.
- [5] UNGVÁRY Rudolf: Tezaurusz és ontológia, avagy a fogalmi ismertetőjegyek generikus öröklődésének formalizálása. http://tmt.omikk.bme.hu/show_news.html?id=3615&issue_id=450
- [6] BOCK, Geoffrey E.: Mapping words for an enterprise taxonomy. How wordmap manages the terminology of an organization. <http://www.wordmap.com/downloads/whitepapers/wordmap-seybold-paper.pdf>
- [7] Preparing a controlled vocabulary for content management and access: an indexer's perspective. http://www.earley.com/Earley_Report/ER_Indexing_vs_Taxonomy.htm
- [8] Vö. [1], p. 53–54.
- [9] Vö. [1], p. 22.
- [10] Free Web directory of taxonomies... a comprehensive directory of taxonomies, thesauri, classification schemes and other authority files from around the world, plus information about taxonomy references, resources. www.taxonomywarehouse.com
- [11] The semantic Web: An interview with Tim-Berners Lee. <http://www.consortiuminfo.org/bulletins/semanticweb.php>
- [12] BERNERS-LEE, Tim–HENDLER, James– LASSILA, Ora: The semantic Web = Scientific American, 2001. május. <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>
- [13] L. [1], p. 53.
- [14] GARSHOL, Lars Marius: Metadata? Thesauri? Taxonomies? Topic maps. <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>
- [15] L. [1], p. 53–54.
- [16] <http://facetmap.com/>; [1], p. 17., 61.
- [17] www.autonomy.com/content/product/IDOL/classification.html
- [18] L. [1], p. 62.
- [19] L. [5]
- [20] L. [1], p. 17.
- [21] UNGVÁRY Rudolf–VAJDA Erik: Szemantikai (relációs) információkereső nyelvek = Könyvtári információkeresés. <http://www.hik.hu/tankonyvtar/site/books/b153/fig-4>
- [22] L. [6]; HENDLER, James–BERNERS-LEE, Tim–MILLER, Eric: Integrating applications on the semantic Web. <http://www.w3.org/2002/07/swint>; NOY, Natalya F.–McGUINNESS, Deborah L.: Ontology development 101. A Guide creating your first ontology. <http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html>

- [23] VÉRY Zoltán: Tudás és hálózat – beszélgetés Krauth Péterrel. <http://miau.gau.hu/miau/84/vz3.doc>
- [24] The semantic Web. <http://www.w3.org/2001/sw/>; Sematic Web activity statement. <http://www.w3.org/2001/sw/Activity>; HENDLER, James–BERNERS-LEE, Tim–MILLER, Eric: Integrating applications on the Semantic Web. <http://www.w3.org/2002/07/swint>
- [25] Az OWL Web Ontológia nyelv. Útmutató. A W3C Magyar Iroda fordítása. <http://www.w3c.hu/fordítások/OWL/REC-owl-semantics-200402106main.html>
- [26] L. [23]
- [27] Szemantikus (asszociatív) hálók (KLTE-tananyag). <http://www.math.klte.hu/~bogнар/mestint4/szemanti.htm>
- [28] What are the semantic network? <http://www.cogs.susx.ac.uk/books/computers-and-thought/chap6/node5.html>; [1], p. 9.; [27]
- [29] <http://www.convera.com/products/retrievalware/concept-search.asp###>
- [30] L. [1], p. 9., [27]
- [31] L. [1], p. 6.
- [32] L. [14]
- [33] SO/IEC 13250 and XTM (Topic map)
- [34] <http://mek.oszk.hu:8181/itm>
- [35] L. [23]

* * *

- BERNERS-LEE, Tim–SWICK, Ralph: The SEMANTIC Web (prezentáció, 2000. május 16.). <http://www.w3.org/2000/Talks/0516-sweb-tbl>
- The DELPHI report: insight for business and technology leaders. <http://www.delphigroup.com/research/whitepapers/20041012-journal.pdf>
- Developing enterprise taxonomies. http://www.earley.com/Earley_Report/ER_Taxonomy.htm
- Information intelligence: content classification and the enterprise taxonomy practice. <http://www.delphigroup.com/research/whitepapers/20040601-taxonomy.pdf>
- Knowledge representation, ontology / John Frederick Sowa [könyvismertetés]. <http://www.jfsowa.com/ontology/>
- Managing a vocabulary for the Semantic Web - Best practice. <http://esw.w3.org/topic/VocabManagementNote>

- McDONALD, Rebecca: Striking a balance: Global and community-specific taxonomies. <http://www.eimagazine.com>
- Az ontológiák. = Ungváry Rudolf–Vajda Erik: Könyvtári információkeresés. <http://www.hik.hu/tankonyvtar/site/books/b153/sec-2.9.2.html>
- Ontology driven architectures and potential uses of the Semantic Web in software engineering. <http://www.w3.org/2001/sw/BestPractices/SE/ODA/>
- Quick guide to publishing a thesaurus on the Semantic Web W3C Working Draft (10 May 2005). <http://www.w3.org/TR/2005/WD-swbp-thesaurus-pubguide-20050510>
- SOWA, John Frederick: Ontology: Words of wisdom, Definition and scope etc. <http://www.jfsowa.com/ontology>
- A Survey of RDF/Topic maps interoperability proposals W3C Working Draft 29 March 2005. <http://www.w3.org/TR/2005/WD-rdfm-survey-20050329>
- Taxonomy & content classification: Market milestone report. White paper. http://www.delphigroup.com/research/whitepapers/WP_2002_TAXONOMY.PDF
- Thesauri and controlled vocabulary. Definitions. <http://www.collectionscanada.ca/8/4/r4-282-e.html>
- WHITE, Martin: Managing enterprise content. London, Ark Group, 2004. 99 p.
- Wordmap taxonomy management system. <http://www.wordmap.com/downloads/whitepapers/wordmap-features.pdf>
- WYLLIE, Jan–SKYRME, David: Taxonomies: frameworks for corporate knowledge. 2. ed. London, Ark Group, 2005. 143 p.
- WYLLIE, Jan: Zen and the art of taxonomy maintenance. 1–4. = Eimagazine, 2004–2005. <http://eimagazine.com>
- Beérkezett: 2006. III. 14-én.



Horváth Zoltánné

IQSYS Rt.

E-mail: horvath.zoltanne@iqsys.hu