

MUNK Sándor

munk.sandor@uni-nke.hu

SZEMANTIKA AZ INFORMATIKÁBAN

Absztrakt

Az informatika szolgáltatásai hosszú ideig az információk adattá alakítására, a számszerűsíthető adatokon végzett, tartalomtól – az adatok 'jelentésétől' – elszakított műveletvégzésre, majd az eredményadatok felhasználó általi értelmezésére épültek. Az informatika fejlődése lehetségessé tette, hogy az adatokkal ne csak mechanikus, formai műveleteket, hanem az emberi gondolkodás egyes képességeit egyre jobban segítő, helyettesítő tartalmi, a jelentést is figyelembe vevő műveleteket is végezzenek. Ennek a képességbővülésnek a megnevezésére jelent meg az informatikában is a 'szemantikus' jelző, amelynek értelmezésére még nem alakult ki egységes álláspont. Jelen publikáció a szemantika, jelentés informatikán belüli értelmezésének vizsgálatát tűzte ki céljául. Ezen belül: bemutatja a szemantika fogalmi alapjait; elemzi a szemantikus információk és műveletek értelmezését; rendszerezi a szemantikus információk típusait, formáit.

For a long time the services of IT were based on the conversion of information to data, the operations on quantifiable data, separated from the content - the 'meaning' of data –, and finally the interpretations of the results by users. Developments in IT made it possible to carry out not only mechanical, formal operations on data, but also operations taking into account the meaning, to increasingly help, or substitute human mental skills. In IT to describe this capability increase has appeared the adjective 'semantic', but it has no common understanding yet. Recent publication aims to analyse the interpretation of semantics, and meaning in IT. In particular: presents the conceptual basics of semantics; analyses the interpretation of semantic information, and semantic operations; systematizes the types, and forms of semantic information.

Kulcsszavak: *szemantika, jelentés, szemantikus technológiák, szemantikus információk, szemantikus műveletek – semantics, meaning, semantic technologies, semantic information, semantic operations*

BEVEZETÉS

A számítástechnikából kinövő informatika szolgáltatásai kezdetben az információk adattá alakítására, a számszerűsíthető adatokon végzett, tartalomtól – az adatok 'jelentésétől' – elszakított műveletvégzésre, majd az eredményadatok felhasználó általi értelmezésére épültek. Ugyanez a megközelítés érvényesült az újonnan megjelenő (szöveges, rajzos, kép és hang) információ-reprezentációk esetében is: az új típusú adatokon a formai (szintaktikai) összetevőkre (karakterekre, rajzelemekre, képpontokra és digitális hangábrázolási elemekre) vonatkozó műveletek álltak rendelkezésre.

Az informatika technikai lehetőségeinek forradalmi ütemű fejlődése, az informatikai szolgáltatások és felhasználók körének erre épülő széleskörű kibővülése tette szükségessé és – egyre bővülő mértékben – lehetségessé, hogy az informatikai eszközök, rendszerek az általuk kezelhető formában rendelkezésre álló információkkal már ne csak mechanikus, formai műveleteket, hanem az emberi gondolkodás egyes képességeit egyre jobban segítő, helyettesítő tartalmi, a jelentést is figyelembe vevő műveleteket is végezzenek. A fenti képességbővülés megnevezésére jelent meg az informatikában is a 'szemantikus' jelző olyan szóösszetételekben, mint a szemantikus web, a szemantikus keresés, a szemantikus interoperabilitás, vagy a szemantikus technológiák.

A szemantikus jelző az informatikában divatos, a korszerűséget és képességet jelző kifejezéssé vált, azonban értelmezéséről, tartalmáról, lényegéről, sajátosságairól még nem alakult ki egységes álláspont. A különböző szemantikus megoldások szerepe a védelmi, katonai alkalmazásokban is folyamatosan bővül, így az ezekre irányuló kutatások megalapozásához hasznos lehet egy általános jellegű fogalmi, tartalmi vizsgálat.

Ennek megfelelően jelen publikáció célja, hogy az informatikában történő alkalmazás szempontjából összegezze, meghatározza a szemantika és a jelentés fogalmi alapjait, értelmezését, alapvető jellemzőit. Ezen belül:

- bemutatja a szemantika fogalmi alapjait, a jelentés jeltudományi, nyelvészeti, illetve informatikai értelmezését;
- bemutatja a szemantikus technológiák szakirodalmi értelmezését, elemzi a szemantikus információk és szemantikus műveletek értelmezését;
- végül rendszerezi az egyedi és az általános szemantikus információk megjelenési formáit, összegzi felhasználásuk lehetőségeit és feladatait.

A JELENTÉS ÉRTELMEZÉSE A JELTUDOMÁNYBAN, A NYELVÉSZETBEN ÉS AZ INFORMATIKÁBAN

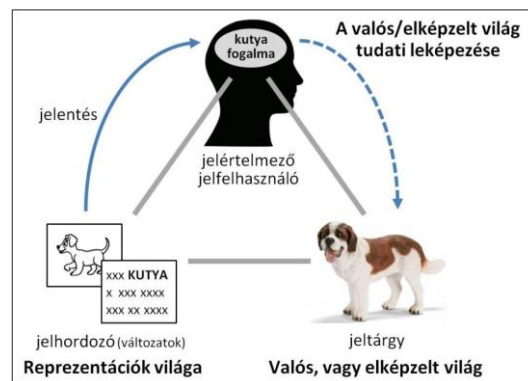
A *szemantika*, más néven jelentéstan, szűkebb értelmezésben a nyelv összetevőinek jelentésével foglalkozó, a nyelvészet részét képező tudományterület. Tárgya hosszú ideig a szavak jelentésére korlátozódott (lexikális szemantika), de a XX. század utolsó harmadától vizsgálati körébe került a kifejezések (szószerkezetek), mondatok, illetve szövegek jelentése is. A szemantika tágabb értelemben nem csak a nyelvi, hanem a társadalomban és a természetben előforduló jelek jelentésével is foglalkozik, a szemiotika (más néven szemiológia), az általános jeltudomány részét képezi. A múlt század végére aztán az informatikában is megjelentek a jelentést is figyelembe vevő szemantikus megoldások, eljárások.

A következőkben röviden összegezzük a jelentés fogalmi alapjait, értelmezési kérdéseit az általános jeltudományban, valamint bemutatjuk a nyelvi szemantika alapjait, a jelentés nyelvészeti értelmezését és az ehhez kapcsolódó alapvető fogalmakat. Ezt követően összegezzük az információk informatikai eszközökben történő reprezentációira (adatokra)

vonatkozó fogalmi alapokat; meghatározzuk az ezek által hordozott jelentés fogalmát; végül elemezzük ezen jelentés szerepét, megjelenését az információs tevékenységek informatikai eszközökkel támogatott megvalósításában.

A jelentés értelmezése a szemiotikában

A jeltudomány általánosan elfogadott álláspontja, hogy a jelek nem önmagukban, hanem meghatározott körülmények, viszonyok között töltik be szerepüket. A *jelviszony* összetevői közé – egy megközelítés szerint [1] – a jelhordozó, a jeltárgy és az értelmező tartoznak. A jelhordozó az a dolog, vagy folyamat, amely a jelentést hordozza az értelmező számára, aki ezt valamilyen más dolog, vagy folyamat jelének tekinti. A jel a jelhordozó és a hozzá kapcsolt jelentés együttese. A jeltárgy az a dolog, amit a jel jelöl, vagy helyettesít. Végül a (jel)értelmező, vagy jelfelhasználó az, aki a jelviszony keletkezési folyamatát elindítja, és aki számára a jelentés szól. [1, 107-109. o.] Más megfogalmazásban a fenti három összetevő három tartományt definiál: a jeltárgyak összességét képező (valós, vagy elképzelt) világot, az ezekre utaló (nyelvi, vagy tágabb értelemben jel-) reprezentációk összességét, valamint a világot visszatükröző, érzékelő, megjelenítő, megismerő emberi tudatot. [2, 13. o.]



1. ábra: A jelviszony és összetevői [1, 3. ábra, 104. o. alapján]

A fenti hármast alapján határozható meg a *jeltudomány három részterülete*:

- a jelek egymáshoz való viszonyával foglalkozó szintaktika;
- a jelzett dolog (jeltárgy) és a jel közötti viszonytal és az ebből fakadó jelentéssel foglalkozó szemantika;
- végül a jel és az jelértelmező közötti viszonyt vizsgáló pragmatika. [1, 33-35. o.]

A *jelalkalmazás három alapvető tevékenysége* a jelértelmezés, a jelhasználat és a jelalkotás. A jelalkalmazó:

- a jelértelmezés során valamely dolgot más dologgal jelviszonyba állít és ebből valamilyen megállapításra jut;
- a jelhasználat során a jelhordozót valamilyen cél érdekében kezeli, mozgatja, módosítja;
- végül a jelalkotás során a jelet meghatározott céllal, a hozzá való jelhordozó létrehozásával, vagy kialakításával megalkotja. [1, 38-40. o.]

A jelalkalmazás egyik – témánk szempontjából alapvető jelentőségű – területe a *kommunikáció*, amely egy közös jelrendszer segítségével megvalósított információcsere. Jelalkalmazásra a kommunikáció mellett sor kerülhet a megismerés során is, amikor a természeti környezetben felismert jelek értelmezésére kerül sor. A későbbiekben vizsgálatainkat az információcsere témakörére szűkítjük.

A *jelentés értelmezése* a jeltudományban két nagy csoportba sorolható. Az egyik megközelítés szerint a jelentés az a dolog (jeltárgy), amelyet a jel jelöl (helyettesít), a másik szerint ezzel szemben a jelentés a dolog tudati képmása (fogalma). A jelviszony keretében

azonban a jel mindig a jelértelmező tudatában megjelenő fogalom közvetítésével jelöli meg a jeltárgyat. A jelentés tehát első fokon a jel és a fogalom közötti kölcsönviszony. [3, 35. o.] Más megfogalmazásban a jelentés a jelértelmező vagy jelfelhasználó számára való jelentés, amely azt a jeltárggyal is viszonyba állítja. [1, 121. o.]

A jelentés értelmezése a nyelvészetben

A *nyelvi szemantika*, mint nyelvészeti részterület, a nyelv összetevőinek jelentésével, jelentésváltozásaival foglalkozik. A nyelv az emberi kommunikáció – információk átadásának, cseréjének – alapvető eszköze, egy speciális jelrendszer.¹ Minden nyelv az alkalmazott jeleket és a hordozott jelentést összekapcsoló folyamatokra épül, "azért tudja betölteni kommunikációs, ábrázoló, stb. szerepét mert bizonyos formákat összekapcsol bizonyos jelentésekkel". [4, 41. o.]

A *jelentés* a nyelvészetben egy széles, bár nem teljes körben elfogadott értelmezés szerint a nyelv összetevőinek a használatuk körülményeitől független, elvont jellemzője. Ettől különbözik az *értelmezés*, vagy *olvasat*, amely csak egyetlen konkrét használat esetén jelenik meg. Végül a *jelölet* (vagyis a jeltárgy) az a valós (vagy elképzelt) dolog, amire a nyelvi összetevő vonatkozik, utal. [4, 41. o.] Eszerint a megközelítés szerint a nyelvi jel a jelentés (fogalom) közvetítésével jelöli meg jeltárgyat. A jelentés és a jelölet párhuzamba állítható az intenzió és extenzió fogalmával.²

A nyelv jelentéssel bíró alapvető összetevői a szavak³, pontosabban a lexikai szavak (lexémák). A *lexikai szó* olyan jelentéssel bíró elemi nyelvi egység, amelynek jelentése nem vezethető le alkotóelemeinek jelentéséből. A lexikai szó általában önálló szó, de lehet képzett szó, összetett szó, szószerkezet, vagy akár teljes mondat is. [4, 76. o.] A *lemma* a lexikai szó egy megállapodás szerinti kanonikus (pld. lexikonokban, szótárakban használt) formája. Elméletileg a nyelv minden (toldalékokkal ellátott, vagy összetett) szavának jelentése levezethető a lexikai szavak jelentéséből.

A *szavak jelentése* több jelentéselemből tevődik össze. A különböző jelentésfajták között alapvető szerepet a fogalmi, más megközelítésben elsődleges jelentés (denotáció) játszik. Az elsődleges jelentést kiegészítheti, módosíthatja, árnyalhatja a járulékos, vagy másodlagos jelentés (konnotáció) és további jelentésfajták is megkülönböztethetők. [3, 85-92. o.]

A *szavak, kifejezések jelentésének meghatározása* (definíció) egy ismert jelentésű szavak (fogalmak) segítségével történő leírás.⁴ A köznyelv és a szaknyelvek esetében erre épülnek a különböző értelmező szótárak, lexikonok. A jelentés meghatározásának különböző fajtái lehetségesek: a szó által jelölt dolgok bemutatásával; szinonima megadásával; a szó tipikus használatának bemutatásával; vagy arisztotelészi definícióval (főlérendelt fogalom és megkülönböztető jegy megadásával). [2, 62-69. o.]

Az *összetett nyelvi formák jelentésének meghatározása* általában a kompozicionalitás elve alapján történik, amely szerint egy összetett kifejezés jelentését egyértelműen meghatározza összetevőinek jelentése és az összetétel módja. A mesterséges nyelvek erre az elvre épülnek,

¹ Szűkebb értelemben a nyelv az emberi beszéden alapuló nyelv, amelynek két alapvető formája a beszélt és az írott nyelv. Emellett léteznek más természetes (pld. jel-), illetve mesterséges nyelvek is.

² Az extenzió mindazon dolgok összessége, amelyekre egy fogalom, jel vonatkozik, utal. Az intenzió mindazon feltételek (tulajdonságok, minőségek) összessége, amelyekkel egy dolognak rendelkeznie kell, hogy egy fogalom, jel vonatkozzon, utaljon rá.

³ Bár a szó a nyelvtudomány alapvető fogalma, meghatározása rendkívül nehéz, széles körben elfogadott definíciója nincs.

⁴ A beszélt nyelv legtöbb szavának jelentését nem definíciók útján, hanem a köznapi tudás (common knowledge) elsajátítása során ismerjük meg.

azonban a természetes nyelvek esetében mindez tisztán nem érvényesül. A kompozicionalitás elve szoros kapcsolatot teremt a szintaktika és a szemantika között. [4, 52-54. o.]

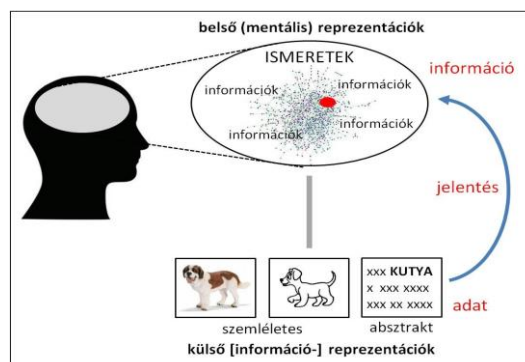
Információk, információ reprezentációk, adatok

Az informatika alapvetően az információk, ismeretek technikai eszközökkel kezelhető reprezentációinak a világa. Elsődleges szolgáltatása, hogy arra alkalmas reprezentációk segítségével támogassa információk tárolását, továbbítását, illetve feldolgozását/átalakítását. Ezek a reprezentációk a jel tudomány rendszerében a jelhordozók közé tartoznak, amelyek más dolgokra vonatkoznak; a jelalkalmazók alkotják meg, használják és értelmezik őket; és kapcsolnak hozzájuk jelentést.

Az informatika két alapvető fogalma az *adat* és az *információ*. A szakterület alapvető szabványa szerint az adat az információ formalizált, újraértelmezhető, továbbításra, értelmezésre és feldolgozásra alkalmas reprezentációja. Az információ pedig objektumokra, például tényekre, eseményekre, dolgokra, folyamatokra, illetve fogalmakra vonatkozó ismeret, amelynek egy meghatározott környezetben meghatározott jelentése van. [5, 6. o.] Ezen megközelítés szerint az adat információt reprezentál (hordoz), az információnak pedig – meghatározott körülmények között – jelentése van.

Az *információ és ismeret viszonya* megítélésünk szerint rész-egész jellegű. Az ismeretek a megismerő tevékenység eredményei, a valós vagy elképzelt világ visszatükröződései az emberi tudatban. Ide tartoznak a közvetlen érzékelés révén megismert empirikus tények és kialakult elképzelések, valamint a fogalmi megismerés útján létrejövő újabb tények, vagy elképzelések, fogalmak, összefüggések és tevékenységi elgondolások. Az információk az ismeretek részelemei, a világ egy megragadott aspektusának visszatükröződései, mentális reprezentációi, amelyek az emberi tudatban egymással összetett módon kapcsolatokba rendeződve léteznek. Az információk mind az ismeretszerzés, mind az ismeretátadás folyamatában közbenső pozíciót foglalnak el. [6, 32-33. o.]

Az *információk belső (mentális) reprezentációjának* két alapvető formája az érzéki megismeréshez kötődő analóg (képi, hang, illat, stb. jellegű) és a fogalmi megismeréshez, kommunikációhoz kapcsolódó propozicionális (ítélet-formájú, nyelvi jellegű) – a továbbiakban fogalmi alapú – reprezentáció. Az előbbiek a dolgokat implicit módon (nem önállóan, a valóság szerkezetét megtartva) reprezentálják, míg az utóbbiak a fogalmi tartalmat megragadva, explicit módon, absztrakt formában. Az információk az emberi tudaton kívül, külső reprezentációk formájában is megjeleníthetők. [6, 34-35. o.]



2. ábra: Ismeret, információ, adat, adat jelentése

A *külső [információ-]reprezentációk*, vagyis az adatok két fő típusa: a szemléletes és az absztrakt reprezentáció. A szemléletes reprezentációk az ember által érzékelhető környezeti hatásokat viszonylag valósághű formában reprodukálják. Érzékelésük, megszemlélésük, lejátszásuk lényegében mindenkiben az eredeti környezeti hatások érzetét kelti, biztosítja a valóság adott aspektusának megismerését. Az absztrakt reprezentációk nem hordoznak a

valóságához hasonló jellemzőket, kapcsolatuk az általuk reprezentált dolgokkal önkényes (megállapodáson alapul). Érzékelésük nem biztosítja automatikusan a reprezentált információ megismerését, ahhoz szükséges az ábrázolás szabályainak az ismerete is. [6, 35-36. o.]

Információ reprezentációk jelentése

Az informatika által kezelt *külső [információ-]reprezentációk, adatok, által hordozott jelentés* – mint az a jel tudomány eredményeiből is következik – nem az adatok belső, önmagában rejlő tulajdonsága, csak az őt kezelő (előállító, felhasználó, stb.) emberekhez kapcsolódóan értelmezhető. Egy adathoz különböző emberek különböző jelentést rendelhetnek, az előállítás során szándékolt jelentésről, a felhasználás során értelmezett jelentésről beszélhetünk. Informatikai szempontból a következőkben jelentés alatt általánosságban az adathoz (külső reprezentációhoz) az emberi tudatban kapcsolódó információt (belső reprezentációt) értjük. Az emberi együttműködés, az eredményes információcsere feltétele az adatok különböző értelmezéseinek, jelentéseinek – felhasználáshoz mérten – megfelelő mértékű azonossága.

A *szemléletes reprezentációk jelentése* szűkebb és tágabb tartalommal is értelmezhető. A jelentés szűkebb értelemben lehetne a tudatban megjelenő érzékszervi élmény, de tágabb értelemben az előbbi feldolgozására, értelmezésére épülő fogalmi információ-együttes, ugyanis az érzékszervi benyomások a megismerés során automatikusan, több szinten is feldolgozásra, hasznosításra kerülnek. A továbbiakban témánk szempontjából a második értelmezésre építünk és az álló- és mozgóképek, hangfelvételek jelentéséhez tartozónak tekintjük, hogy mit "tartalmaznak", mit "ábrázolnak", mit "fejeznek ki".

A fentiek alapján a továbbiakban egy *információ-reprezentáció, adat jelentése* alatt az emberi tudatban közvetlenül hozzá kapcsolódó fogalmi alapú információt, információ-együttest (mentális reprezentációt) értjük. Ez maga után vonja, hogy az adott információ-reprezentáció jelentése önmagában tulajdonképpen nem is értelmezhető. A külső reprezentáció által hordozott információ ugyanis csak mesterségesen, "információvesztéssel" szakítható ki az ismeretrendszerből, választható el a megfogalmazást biztosító fogalomrendszertől és az értelmezést támogató háttérismeretektől.

A jelentés szerepe, megjelenése az informatikában

Az *információ-reprezentációk, adatok jelentésének szerepe a támogatott információs tevékenységekben* nem azonos. Az *információtárolás és információtovábbítás* technikai támogatása esetében a hordozott jelentés szerepe másodlagos. Mindkét funkció során az alapvető követelmény az információt hordozó reprezentáció megbízható, változatlan formában történő 'eljuttatása' az idő, illetve a tér egy másik 'pontjára'. Az adattárolás és az adattovábbítás során legfeljebb a hordozott jelentést nem érintő, attól független, alacsonyabb szintű reprezentációs átalakításokra kerül sor. Ennek megfelelően nem soroljuk ide, mert nem ezen funkciók feladatának tekintjük az eltérő forrás- és célreprezentációk közötti átalakításokat. Hasonlóképpen nem játszik érdemi szerepet a jelentés az *információk technikai eszközök segítségével támogatott megszerzésében*, az adatgyűjtésben sem.

Eltérő a helyzet a belső és a külső reprezentációk közötti átalakítások – *információk adatokká alakítása, közreadása, illetve adatok felhasználása*, értelmezése – esetében, amikor megkerülhetetlenek az adott típusú adathoz tartozó jelentésre vonatkozó ismeretek. Ezek létezhetnek az informatika hatókörén kívül is, az adatokat kezelő személyek tudatában, vagy hagyományos formában (dokumentációkban, leírásokban) rögzítve, ebben az esetben felhasználásuk is az érintett személyek feladata, felelőssége.

Végül más a szerepe a jelentésnek az *információkat hordozó adatok feldolgozása* (szelektálása, átalakítása, stb.) esetében. Az adatokon számos – a felhasználók számára hasznos – művelet végezhető, amely független azok jelentésétől (számok sorba rendezése, szöveg

oldalra tördelése, kép színkorrekciója, hangfelvétel hangerejének módosítása, stb.). Az informatika szolgáltatásai kezdetben elsősorban ilyen műveletekre épültek. Vannak azonban olyan feladatok is, és ezek száma, jelentősége folyamatosan nő, amelyek csak a kezelt adatok jelentésének – legalább bizonyos mértékű – ismeretében valósíthatók meg.

Kezdetben az adatok jelentése az azokat feldolgozó algoritmusokban (programokban) – azokba beágyazva és ezzel egyben 'elrejtve' – került figyelembevételre, felhasználásra. A felhasználók számára az egyes adatok jelentését az adatbeviteli űrlapokon, illetve az eredménytáblákon szereplő megnevezések, magyarázatok orientálták. Ez a megoldás azonban nem támogatta a jelentésre vonatkozó, felhasznált információk hozzáférhetőségét, hatékony felhasználhatóságát.

Az adatok jelentésére vonatkozó információk elsőként a strukturált adatokhoz, az adatbázis-kezelő rendszerekhez kapcsolódóan, az adatmodellekben, illetve az adatleíró nyelvekkel leírt adatbázis sémákban jelentek meg. Ezekre épültek az általánosabb rendeltetésű metaadat nyilvántartások. Félig strukturált adatok esetében a formatizált üzenetekhez, illetve a de facto szabványt képező XML formátumhoz kapcsolódóan születtek elemi adatok és struktúrák leírását szolgáló megoldások. Végül a strukturálatlan (szöveges és multimédia) adatok esetében a dokumentumot leíró jellemzők jelentek meg.

SZEMANTIKUS TECHNOLÓGIÁK, INFORMÁCIÓK, MŰVELETEK

Az informatika és az informatika-alkalmazás fejlődése során az ismeretek, információk adatok formájában történő közreadásához, illetve az adatok feldolgozásához, majd ismételt információvá, ismeretté alakításához kapcsolódóan fokozatosan megjelent a 'hagyományos' megoldásokat meghaladó, a közös (szándékolt, egyeztetett) értelmezést is figyelembe vevő megvalósítás támogatásának igénye és lehetősége.

Az adatok közös értelmezés alapján történő kezelésének informatikai módszerekkel, eszközökkel történő támogatásához mindenekelőtt az alkalmazott külső információ-reprezentáció (adat) és az általa hordozott információ (jelentés) közötti kapcsolatra vonatkozó, informatikai eszközökkel kezelhető kiegészítő adatokra van szükség. A továbbiakban majd az ezen adatok által hordozott információkat nevezzük szemantikus információknak és az ilyen információkat (is) kezelő műveleteket tekintjük szemantikus műveleteknek. Egy adott csoportba tartozó eszközök, módszerek, eljárások összességét és az alkalmazásukra vonatkozó ismereteket technológiának nevezzük⁵, ebben az értelemben beszélhetünk szemantikus technológiákról is.

A következőkben elsőként áttekintjük, hogy a szakirodalom hogyan definiálja a szemantikus technológiák fogalmát, tartalmát, majd elemezzük a szemantikus információk és a szemantikus műveletek értelmezését.

⁵ Eredeti, szűkebb értelmezése szerint: mindazon módszerek és eljárások összefoglaló megnevezése, amelyekkel rendszeres átalakítás útján nyers-, illetve alapanyagokból ipari készterméket állítanak elő.

Szemantikus technológiák

A szemantikus technológiák a 2000-es évek elején kerültek a szakmai-tudományos vizsgálatok homlokterébe. A témakör népszerűségét a Szemantikus Web víziójának megjelenése alapozta meg, amely szerint "a Szemantikus Web nem egy különálló Web, hanem a jelenleginek egy kiterjesztése, amelyben az információknak jól definiált jelentése van, jobban biztosítva ezzel, hogy a számítógépek és az emberek együttműködve dolgozzanak." [7] Az eredeti vízió egy olyan rendszer volt, amely segítségével informatikai eszközök képesek 'megérteni' és jelentésüknek megfelelően reagálni összetett emberi kérésekre, kérdésekre.

A szakirodalomban a *szemantikus technológia* fogalmának meghatározására számos különböző változat született ([8], [9], [10], [11], [12] és [13]). A leggyakrabban hivatkozott értelmezések közé a következők tartoznak:

"A szemantikus technológia olyan szoftver technológia, amely futási időben lehetővé teszi az információk jelentésének és a köztük fennálló kapcsolatoknak a megismerését és feldolgozását. Ahhoz, hogy egy szemantikus technológia valóban működőképes legyen, egy rendszerben léteznie kell a világ egy része ismeretmodelljének, amelyet egy, vagy több alkalmazás futásidőben használ." [8, 4. o.]

"A szemantikus technológiákat a korábbi technológiáktól az különbözteti meg, hogy a jelentést az adatoktól, a tartalomtól és az alkalmazás kódoktól elkülönítve jelenítik meg, hogy azokat a számítógépek, illetve az emberek megoszthassák, értelmezhessek és feldolgozhatassák." [10, idézi 12, 9. o.]

A különböző meghatározások közös eleme a különböző (strukturált, félig strukturált, szöveges és multimédia) információ-reprezentációk jelentésének kezelése és azoktól elkülönített megjelenítése. Megjegyzendő azonban, hogy a hivatkozott publikációk nem részletezik az elkülönülten kezelt jelentés fogalmát, sajátosságait, nem foglalkoznak azzal, hogy az adatok maguk is a jelentést hordozzák.

A szemantikus technológiák mellett a szakirodalomban – mindenekelőtt a Szemantikus Web népszerűsége miatt – széles körben találkozhatunk a *szemantikus web technológiák* kifejezéssel is. Ennek értelmezése a Szemantikus Web hármasszögéhez kapcsolódóan adható meg. [részletesebben lásd 14] Eszerint a Szemantikus Web egyrészt egy jövőkép, amely kimunkált formában elsőként Berners-Lee és társai publikációjában körvonalazódott. [7] Másrészt a Szemantikus Web egy program is (W3C Semantic Web Activity), amelyet a World Wide Web Consortium az 1990-es évek végén hozott létre⁶ és amely szerint:

"A Szemantikus Web egy közös keretrendszert biztosít, ami lehetővé teszi az adatok alkalmazási, szervezeti és közösségi határokon átnyúló megosztását és újrahasznosítását. Ez a W3C által vezetett és nagyszámú kutató és ipari partner részvételével folyó közös munka." [15]

Végül a Szemantikus Web technológiák, eszközök és szabványok együttese is, amelyek a víziót megvalósító rendszer alapvető építőelemeit képezik. A szemantikus web technológiák tehát meghatározott W3C szabványok⁷ együttese, amelyek történetesen különösen alkalmasak szemantikus technológiák algoritmusainak, megoldásainak megvalósítására. [16]

⁶ A program 2013 decemberében lezárásra került és helyébe a tágabb tartalmú W3C Data Activity program lépett.

⁷ URI (Uniform Resource Identifier, Egységes Erőforrás Azonosító), XML (Extensible Markup Language, Kiterjeszhető Jelölő Nyelv), RDF (Resource Description Framework, Erőforrás Leíró Keretrendszer), RDFS (RDF Schema, RDF Séma), OWL (Web Ontology Language, Web Ontológia Nyelv), SPARQL (SPARQL Protocol and RDF Query Language, SPRQL Protokoll és RDF Lekérdező Nyelv), SWRL (Semantic Web Rule Language, Szemantikus Web Szabályleíró Nyelv), stb.

Metainformációk, metaadatok

A szemantikus információk fogalma megítélésünk szerint a jelentés informatikában történő kezelésének alapja, kiinduló pontja. Erre építhető fel a szemantikus műveletek, illetve a szemantikus technológiák fogalma, értelmezése is. Ennek érdekében a következőkben meghatározzuk a szemantikus információk javasolt értelmezését, majd ennek alapján megjelenési formáikat és felhasználási módjaikat az informatikában.

A szemantikus információk értelmezése a tárgyalási univerzum fogalmköréhez, ezen belül az *ismeretek elsődleges és meta-szintjei* fogalmához kapcsolódva lehetséges. A *tárgyalási univerzum*⁸ (a továbbiakban röviden a 'világ') egy adott alkalmazási terület, szakterület érdeklődésre számot tartó objektumainak, azok tulajdonságainak és viszonyainak összessége, amelyről ismeretekkel, információkkal rendelkezünk, információkat kezelünk (szerzünk, cserélünk, felhasználunk). Ezen ismeretek alapvető csoportját az objektumok tulajdonságaira, viszonyaira vonatkozó elsődleges, ún. objektum-szintű információk alkotják.

A *propozicionális (fogalmi alapú) elsődleges információk* az emberi tudatban a világ objektumait jelölő nevek (leírásokra), objektumok csoportjait meghatározó objektum-fogalmakra és kijelentéseket létrehozó tulajdonság-, illetve viszonyfogalmakra épülnek. Az elsődleges információk mindig egy adott értelmezési kerethez, fogalomrendszerhez kapcsolódóan léteznek, attól elszakítva értelmetlenek, értelmezhetetlenek. Ebből következően egy alkalmazási terület, szakterület ismeretrendszerének kiemelt részét képezi az alkalmazott terminológia, fogalomrendszer. Egy kijelentés formájában megfogalmazott információ értelmezéséhez tehát szükség van az abban szereplő fogalmak, beleértve a tulajdonság-értékekhez (számokhoz, logikai értékekhez, mennyiségekhez, stb.) kapcsolódó fogalmak ismeretére. A különböző fogalmak meghatározása csak más fogalmakat felhasználva, rekurzív módon lehetséges.

Az adott alkalmazási terület, szakterület információinak második szintjét tehát a *metainformációk* alkotják, amelyek nem a világ egyes objektumaira, hanem az ezeket leíró elsődleges információkra vonatkozó információk. A metainformációk egyik csoportját az elsődleges információkban szereplő fogalmakra, ezek körére, kapcsolatrendszerére, vagyis a szakterület fogalomrendszerére vonatkozó információk alkotják. Egy másik csoportba az elsődleges információkat minősítő, azok értelmezését, feldolgozását, felhasználását befolyásoló kiegészítő információk tartoznak. A metainformációk tehát más információk értelmezését segítő, módosító információk.

A szakirodalomban sokszor – nézőpontunkkal ellentétben – keveredik a metainformáció és metaadat fogalma, gyakran egymás szinonimájaként szerepelnek. A *metaadatok* fogalmát az informatika régóta használja, az *ISO 11179 metaadat-tárház szabvány* meghatározása szerint ezek más adatot meghatározó és leíró adatok [17, 4. o.] (amelyek csak meghatározott összefüggésben, másik adatra vonatkoztatva [lesznek] metaadatok). Az ISO 11179 szabvány adatelemek szemantikájának és reprezentációjának leírását és ezen leírások regisztrációs rendjét szabályozza. Célja, hogy ezzel elősegítse az adatok jelentésének pontos értelmezését és célszerű ábrázolásának megválasztását. [17, v. o.] A szabvány a metaadatok különböző típusait különbözteti meg. Egy részük az adatelem szemantikus, más részük reprezentációs jellemzőit írják le. A szemantikus metaadatok is két csoportra – az objektum- és tulajdonság típust, illetve az érték fogalmakat leíró adatokra⁹ – oszlanak. [17, vi. o.]

Tágabb értelemben a metaadat olyan strukturált adat, amely egy információs erőforrást ír le, magyaráz, tár fel, vagy más módon könnyíti meg visszakeresését, felhasználását és kezelését¹⁰.

⁸ Universe of Discourse (UoD), más megnevezéssel kommunikációs szöveggörnyezet (Domain of Discourse), vagy egyszerűen 'világ' (Universe), leegyszerűsített megfogalmazásban 'minden, amiről beszélünk, beszélhetünk'.

⁹ A szabvány szóhasználatában értelmezési környezeti (contextual) és szimbólum (symbolic) jellegű típusok.

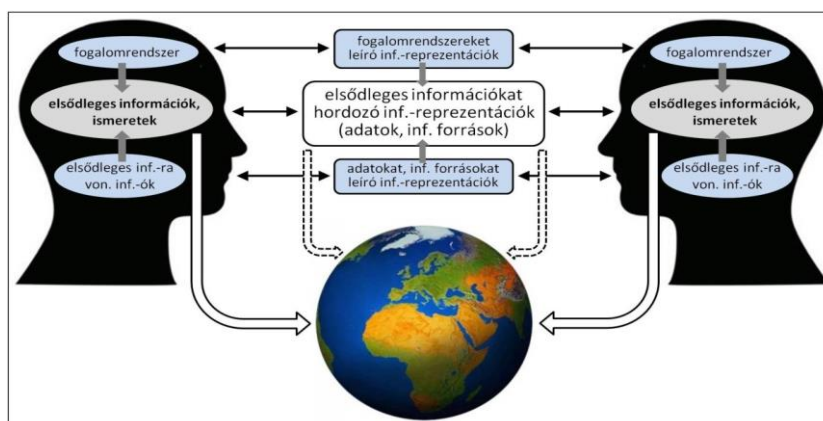
¹⁰ Ez az értelmezés magában foglalja a hagyományos könyvtári katalógusrendszereket is.

[18, 1. o.] A metaadatok típusai közé ebben a megközelítésben a feltárást és azonosítást támogató leíró metaadatok, a formai műveleteket támogató strukturális metaadatok és az erőforrások kezelését támogató adminisztratív metaadatok tartoznak. A metaadatok szűkebb és tágabb értelmezés mellett egyaránt (más adatokhoz hasonlóan) információ-reprezentációk, esetünkben metainformációk reprezentációi.

Szemantikus információk fogalma, típusai

A fentieknek megfelelően a továbbiakban *szemantikus információ* alatt – témánk szempontjából – információ-reprezentációk, információs erőforrások (a továbbiakban összefoglaló módon adatok) értelmezését, jelentésmegőrző feldolgozását segítő, jelentéshez kapcsolódó metainformációt értünk. A szemantikus információk informatikai eszközök, megoldások által kezelhető módon adatok – szemantikus adatok – formájában áll[hat]nak rendelkezésre, ugyanazon szemantikus információk természetesen különböző reprezentációkban is.

A szemantikus információk tárgya lehet a legegyszerűbb elemi adat (pld. egy számérték), egy összetett adatstruktúra (pld. egy dátum), vagy egy nagyméretű, bonyolult szerkezetű információhordozó (pld. egy könyv). Az érintett adat lehet strukturált, félig-strukturált, vagy strukturálatlan. Értelmezésünk szerint a szemantikus adatok nem az érintett adatok jelentését 'tartalmazzák' attól elkülönülten (mivel azokat maguk az alapinformációk hordozzák, esetleg önmagukban nem elegendő módon), hanem 'csak' ezen jelentés meghatározását segítik.



3. ábra. Elsődleges és meta-információk, adatok és metaadatok

Az egyes adatok értelmezése, jelentésmegőrző feldolgozása során három összetevő játszik szerepet: maga az értelmezendő, feldolgozandó adat; az adott adatra vonatkozó egyedi információk, ismeretek; valamint az értelmező, feldolgozó rendelkezésére álló általános (háttér-) információk, ismeretek, az értelmezési keret. Ennek megfelelően megkülönböztethetők egyedi és általános szemantikus információk.

Az *egyedi szemantikus információk* tehát mindig egy adott adatra (adategyüttesre) vonatkoznak, a szakirodalomban általában ezek adat formájában történő reprezentációira használják a metaadat megnevezést. Fontos hangsúlyozni, hogy nem minden metaadat hordoz szemantikai információt: mint az előzőekben is bemutattuk, vannak formai, illetve kezelési jellemzőket leíró metaadatok is. Az egyedi szemantikus információkat hordozó metaadatok megjelenhetnek az általuk leírt adatokkal együtt, vagy azoktól elszakítva.

Az *általános szemantikus információk* az elsődleges információkat hordozó adatok által felhasznált fogalmakat írják le. Mind az adatok értelmezésének, mind szándékolt jelentésüknek megfelelő felhasználásukhoz elengedhetetlen az érintett (objektum-, tulajdonság-, viszony- és érték-) fogalmak ismerete. Ez egyrészt magában foglalja a fogalmak emberek által értelmezhető definícióját és a fogalmak közötti viszonyok, a fogalmakra érvényes szabályok ismeretét,

másrészt az informatikai eszközök által (is) kezelhető reprezentációk meglétét. Ezek között kiemelt szerepet játszanak a következő fejezetben részletesebben ismertetésre kerülő kifejezéslisták, taxonómiák, thezauruszok és ontológiák.

Szemantikus műveletek, eljárások

Informatikai szempontból a szemantikus műveletek adatokkal, vagyis információ-reprezentációkkal végzett sajátos műveletek. Ennek megfelelően először összegezzük az adatokkal végzett műveletek értelmezését, majd meghatározzuk a szemantikus műveletek fogalmát és röviden bemutatjuk főbb típusait.

Az adatokkal végzett művelet – a matematikai művelet analógiájára – egy, vagy több (bemenő) adat alapján egy, vagy több (kimenő) adatot előállító eljárás. Ez a meghatározás értelmezhető szűkebben, 'egyszerű' adatokra és tágabban, tetszőleges (strukturált, félig-strukturált, strukturálatlan) elemi, vagy összetett információ-reprezentációkra. Elemi adatok, információ-reprezentációk (numerikus, logikai, szöveges, multimédia adatok) esetében a lehetséges műveletek köre és értelmezése az adott típushoz kötötten kerül megadásra. Összetett adatokkal végezhető műveletek az egyes összetevőkkel végezhető műveletekből építhetők fel.

Egy adatokkal végzett művelet rendeltetése, hogy az általa reprezentált információkkal végzett mentális műveletet helyettesítsen, vagyis a kimenő adatok jelentése egyezzen meg a bemenő adatok jelentésén végzett mentális művelet eredményével. Ez az értelmezés jelenik meg az ISO informatikai fogalomjegyzékében szereplő értelmező ábrán is [19, 25. o.]. Az információkon az emberi gondolkodás során végrehajtott mentális műveleteket helyettesítő adatuműveletek jellemzően egyszerűbb műveletekre bontva, illetve alacsonyabb reprezentációs szintek műveleteire épülően kerülnek megvalósításra.

Az adatokkal végzett műveletek által felhasznált, illetve előállított adatok között lehetnek olyanok is, amelyek a műveletben kezelt más adatokra vonatkozó jellemzőket hordoznak. Ezek a metaadatok – mint már többször hangsúlyoztuk – vonatkozhatnak az érintett adat szintaktikai, illetve szemantikai (tartalmi, jelentéshez kapcsolódó) jellemzőire. Összetett adatok esetében a leíró jellemzők lehetnek az adatok összetevői is (pld. XML elemet leíró attribútumok). Minél több leíró jellemzőt használ fel egy művelet, annál általánosabban, szélesebb körben használható fel.

A *szemantikus művelet* fogalma az informatikában a fentiekre építve a következőképpen határozható meg: adatokkal végzett olyan művelet, amely az érintett adatok tartalmára, jelentésére vonatkozó (szemantikus információkat hordozó) adatokat is felhasznál, vagy eredményez. A hagyományos műveletek csak a reprezentációkra épülnek (ami számos esetben elegendő is), a szándékolt jelentés esetükben – ha szükséges – az algoritmusba beépítve, egyben elrejtve kerül figyelembe vételre. A továbbiakban a szemantikus információkat és az ezeket hordozó (meta)adatokat – a reprezentációs szinttől (mentális, illetve külső) elvonatkoztatva – együtt szemantikus jellemzőnek nevezzük. A szemantikus művelet tehát szemantikus jellemzők felhasználásával végrehajtott művelet, amelynek eredménye lehet szemantikus, vagy 'hagyományos' jellemző.

Mint a fenti meghatározásból látható, a szemantikus műveletek terjedelmét a szemantikus jellemző fogalmának tartalma, értelmezése határozza meg. Szűkebb értelmezés esetén kevesebb műveletet nevezhetünk szemantikus műveletnek, tágabb értelmezés esetén többet. A szakirodalomban szemantikus művelet alatt egyes esetekben ontológiára épülő műveletet, más esetekben (tartalmi jellemzőket leíró) metaadatokat felhasználó műveleteket értenek. A fentiek alapján értelmezésünk szerint mindkét csoport a szemantikus műveletek közé tartozik.

A szemantikus műveletek típusai – eredményük és a felhasznált adatok jellege függvényében – tisztán logikai alapon is három csoportba sorolhatóak. Az első csoportba azon műveletek tartoznak, amelyek 'hagyományos' jellemzőkből szemantikus jellemzők felhasználásával

'hagyományos' jellemzőket állítanak elő. A második csoportot azon műveletek alkotják, amelyek 'hagyományos' jellemzőkből, esetleg szemantikus jellemzőket felhasználva, szemantikus jellemzőket hoznak létre. Végül a harmadik csoportba a szemantikus jellemzőkből szemantikus jellemzőket előállító műveleteket sorolhatjuk.

Az első csoport műveletei felhasználói szempontból talán a legfontosabbak, ezek nyújtják ugyanis a hagyományos megoldások lehetőségeit meghaladó eredményeket. A másik két csoport műveletei lényegében az első csoport műveleteinek feltételeit teremtik meg. A második csoport az egyedi szemantikus (leíró) információk emberi előállítását helyettesíti, míg a harmadik csoport jellemzően az általános szemantikai információk (pld. fogalomrendszerek) kezeléséhez kapcsolható.

SZEMANTIKUS CÍMKÉZÉS, FORMALIZÁLT FOGALOMRENDSZEREK

Az előzőekben bemutattuk, hogy az informatikában a jelentéshez kapcsolódó kiegészítő információk, illetve az ezeket hordozó adatok két jól elkülöníthető csoportba – az egyedi és az általános szemantikus információk közé – tartoznak. Ezek alkalmazása (előállítása, felhasználása) különbözteti meg a 'hagyományos' informatikai megoldásokat a szemantikus jelzővel ellátott informatikai megoldásoktól.

Az egyedi és az általános szemantikus információk a szakirodalomban más megnevezésekkel, illetve specifikusabb formáikban szerepelnek. A könyvtártudomány és a filozófia területén kialakult fogalmak (annotáció, teaurusz, ontológia, stb.) fokozatosan jelentek meg és kaptak sajátos értelmezést, felhasználást az informatikában. A következőkben külön-külön vizsgáljuk és összegezzük az egyedi és az általános szemantikus információk informatikai megjelenési formáit, rendeltetésüket, előállításuk és felhasználásuk lehetőségeit és feladatait.

Egyedi szemantikus információk, szemantikus címkézés

Az egyedi szemantikus információk lényegi sajátossága, hogy konkrét adatokhoz, információ-hordozókhoz, ezen belül azok tartalmához, jelentéséhez kapcsolódó kiegészítő információk. Rendeltetésük, hogy elősegítsék az érintett információ-reprezentációk tartalmának, jelentésének megismerését, tartalom-, jelentés-alapú kezelését, feldolgozását, ami a kiegészítő leírás hiányában nem, vagy nem kellő eredményességgel lenne lehetséges.

Az egyedi szemantikus információk két nagy csoportba sorolhatóak, amelyek eltérő célok megvalósítására alakultak ki. Az első csoportot az elemi információ-reprezentációk jelentéséhez kapcsolódó leíró jellemzők alkotják, a másodikba pedig az összetett információ-reprezentációk (szöveges, média, és vegyes összetételű dokumentumok) tartalmát leíró jellemzők tartoznak.

Az elemi adatok jelentését leíró jellemzők az informatikában elsőként a relációs adatbázisok, majd a félig-strukturált reprezentációk adatelemeihez, végül a szöveges reprezentációk szavaihoz, kifejezéseihez kapcsolódóan jelentek meg. Alapvető rendeltetésük, hogy elősegítsék az érintett összetevők szándékolt jelentésnek megfelelő értelmezését, illetve felhasználását (további feldolgozását).

A relációs adatbázisokban kezelt strukturált adatok esetében az egyes elemi adatok jelentését meghatározó objektum- és tulajdonság fogalmakat az adatdefiníciós/adatleíró nyelv (Data Definition/Description Language, DDL) az adatbázis táblák és az adatbázis oszlopok/mezők megnevezésével írta le, vagyis egy adatelemhez közvetve egy objektum-fogalom név és egy tulajdonság fogalom név került hozzárendelésre. Az így leírt adatbázis séma az elemi adatok által hordozott információk fogalmi szintű leírásának (modelljének) reprezentációja volt. Az értelmezést segítő további információk, meghatározások csak az adatbázis

hagyományos formátumú dokumentációjában, az informatikai felhasználás számára elérhetetlen formában álltak rendelkezésre.

A *félig strukturált adatok* esetében az egyes összetevők jelentését leíró jellemzők megjelenése és elterjedése az informatikában az XML formátum¹¹, illetve tágabb értelemben a jelölőnyelvek (markup languages) alkalmazásához kapcsolható. A jelölőnyelvek lehetővé teszik dokumentumok kiegészítését különböző (megjelenítési, strukturális, vagy tartalmi/szemantikai) rendeltetésű, az eredeti tartalomtól szintaktikailag elkülönített megjelölésekkel. A napjainkra egyeduralgódóvá vált XML-alapú reprezentációkban az egyes összetevőket fogalmakat, kategóriákat leíró (megnevezéssel azonosított) címkék jelölhetik.

A *nem strukturált szöveges adatok* esetében az egyes szavak, kifejezések értelmezését és ezzel a teljes szöveg feldolgozását segítő jellemzők elsősorban kategóriákba soroló kiegészítő leírások, címkék formájában jelennek meg. Ezek a jelentés egyértelműsítése, illetve az információ feldolgozása érdekében meghatározhatják például az adott szó, kifejezés szófaját, értelmezési tartományát¹², vagy a hordozott információ típusát¹³.

A *nagyobb információegységek tartalmát leíró jellemzők* elsőként a könyvtári gyakorlatban jelentek meg kulcsszavak, tárgyszavak, annotációk, referátumok formájában, napjaink informatikájában pedig elsősorban a címkézés formájában találkozhatunk vele. Alapvető rendeltetésük a tartalom alapján történő kiválasztás, előkereshetőség elősegítése az eredeti információegység (dokumentum) megismerése, egyedi feldolgozása nélkül.

Az elsődleges alkalmazási területek, a leírás tárgyát képező információk erőforrások közé tartoznak az informatikai rendszerek, eszközök segítségével elérhető önálló – akár hagyományos formában is létező – információhordozó objektumok: könyvek, tanulmányok, cikkek, multimédia anyagok (képek, filmek, hanganyagok), valamint az Internet lehetőségeire, szolgáltatásaira épülő új közösségi média 'termékei': blog, mikroblog, közösségi oldal bejegyzések, megosztott multimédia anyagok.

A tartalomleírás legegyszerűbb, legelterjedtebb eszköze az informatikában a *címkézés* (tagging). A *címke* (tag) formai értelemben egy információegységhez rendelt (metainformációt hordozó) metaadat. A címke egy címketípus¹⁴ és egy érték együttese, ahol az utóbbi jellemzően egy elemi strukturált adat (szöveges, numerikus, logikai). A hagyományos leíró jellemzőket megvalósító dokumentum tulajdonságok első változatai gyártófüggő formában már régebben megjelentek az informatikában¹⁵, azonban ezek kiterjedt alkalmazására nem került sor.

A címke, mint metaadat, az adott információegység számos jellemzőjét hordozhatja, amelyek közül a tartalomhoz kapcsolódó jellemzők csak egy csoportot alkotnak. A *szemantikus címke* alatt tehát egy információegységhez rendelt, annak tartalmát leíró szemantikus jellemzőt értünk. A tartalmi leírás, ezzel egyben a szemantikus címkék alapvető – legegyszerűbb és legegyszerűbben kezelhető – formája mindmáig a kategorizálás: kategóriába soroló fogalom, illetve az ezt leíró szó, vagy kifejezés¹⁶ használata.

A *szemantikus jellemző származhat* egy előre meghatározott körből (kategórialistából), vagy megválasztása lehet kötetlen. A címkézés – bár technikailag mindkét lehetőséget támogathatja – a gyakorlatban az utóbbi csoportba tartozik. Széleskörű elterjedésének egyik oka éppen az egyszerűségben és az egyéni szabadságban rejlik. A címkék formájában hozzárendelt

¹¹ Extensible Markup Language = kiterjeszhető jelölő nyelv.

¹² Pld. Napóleon ~ személy, konyak, helység, film, stb.

¹³ Pld. személy, szervezet, hely megnevezése, időpont, időtartam, mennyiség megadása stb. (lásd 'tulajdonnév' felismerés, Named Entity Recognition)

¹⁴ A címketípus hiányozhat is, ebben az esetben a 'tartalmat jellemző kategória' típus feltételezhető.

¹⁵ Lásd például a különböző Microsoft Office dokumentum tulajdonságokat: szerző, cím, tárgy, kulcsszavak, stb.

¹⁶ Kulcsszó, tárgyszó, index kifejezés.

szemantikus jellemzők alapvető sajátossága az is, hogy nem alkotnak hierarchikus kategória-rendszert (ez a kötetlen címkézés esetében nem is lenne lehetséges).

Az információegységet tartalmi szempontból leíró *szemantikus jellemző hozzárendelése* történhet manuális, vagy automatizált módon, informatikai megoldások segítségével. Előbbi esetben a jellemzőt hozzárendelheti az információegység előállítója (folyóiratcikkek, blog-bejegyzések, stb.), az információegység hozzáférésre, felhasználásra történő előkészítését végző szakember, esetleg az információegység felhasználója (felhasználói).

A megosztott információk szemantikus leírásának széles körben elterjedt megoldása a *közösségi címkézés* (folksonomy¹⁷), egyrészt egy folyamat, amely során több felhasználó rendel metaadatokat kulcsszavak formájában megosztott tartalomhoz [idézi 20, 153. o.], másrészt az információk 'demokratikus' leírása eredményeként kialakuló címke halmaz. Az alkalmazott címkék és alkalmazásuk gyakoriságának megismerését segíti a *címkefelhő* (tag cloud), amely az egyes címkék betűméretekkel és színekkel jellemzett vizuális megjelenítése.

A közösségi címkézés kategória listája a hagyományos osztályozás más megoldásaival (taxonómiák, teauruszok, stb.) szemben nem határoz meg kapcsolatokat a különböző elemek között és alkalmazásának módjából következően többértelmű, azonos értelmű, eltérő formátumú kifejezéseket tartalmazhat, ami az információkeresés szempontjából hátrányos. Ennek ellenére alkalmazása széles körben terjed, eredményessége vitathatatlan és kapcsolatrendszere más osztályozási megoldásokkal folyamatosan bővül.

Az egyedi szemantikus információk közös jellemzője, hogy az érintett információkhoz azok szándékolt jelentésére, tartalmára utaló kategóriákat, fogalmakat, pontosabban ezek megnevezését kapcsolja. Ezek a szemantikus jellemzők (címkék) csak megfelelő háttértudás, további információk birtokában segítik, biztosítják az értelmezést és a jelentésnek megfelelő felhasználást. Ehhez ugyanis szükség van a címkék jelentésének közösen elfogadott, egyeztetett értelmezésére, az általuk hordozott fogalmak ismeretére. Ezek pedig informatikai rendszerek, eszközök számára a fogalmakra, fogalomrendszerekre vonatkozó általános szemantikus információk formájában állnak rendelkezésre.

Általános szemantikus információk, formalizált fogalomrendszerek

Mint azt már korábban megfogalmaztuk, informatikai szempontból az adatok (külső információ-reprezentációk) jelentése az emberi tudatban hozzájuk kapcsolódó fogalmi alapú információ (mentális reprezentáció). Szemantikus információ alatt pedig az adatok és az általuk hordozott információk közötti kapcsolatra vonatkozó, más megfogalmazásban az adatok jelentéséhez kapcsolódó, az adatok értelmezését, jelentésmegőrző feldolgozását segítő kiegészítő információkat értünk. Az előző alpontban bemutatottuk, hogy az egyedi szemantikus információk lényegében az érintett adatokhoz rendelt szemantikus címkéket, fogalmakat jelentenek, amelyek önmagukban, jelentésük ismerete nélkül nem képesek betölteni rendeltetésüket.

Az egyedi szemantikus információk értelmezését, illetve az érintett információk feldolgozása során történő felhasználását az informatika a kapcsolódó fogalmak formalizált leírásával és rendelkezésre bocsátásával támogatja. Ehhez minimálisan szükség van az alkalmazott reprezentációk (megnevezések, kódok, stb.) és a megfelelő fogalmak meghatározásának leírására. Az informatikai eszközökkel is kezelhető formalizált fogalomrendszerek az általános szemantikus információk alapvető csoportját alkotják. Meg kell azonban jegyezni, hogy a valóságban a fogalomrendszerek egyrészt alkalmazási terület,

¹⁷ A folk (nép) és a taxonomy (taxonómia) szavak kombinációja, vagyis 'népi taxonómia', ami pontatlan leírás, mivel a kötetlen címkézés elemei nem rendeződnek hierarchikus rendszerbe.

szakterület, másrészt ezen belül is alkalmazói kör függőek, amelynek következményeivel jelen publikációban részletesebben nem foglalkozunk.

A *formalizált fogalomrendszerek* az emberi tudás alapvető részét képező fogalmak és sokrétű kapcsolataik külső reprezentációi, amelyek a tudásreprezentáció egy speciális típusát alkotják. Egy formalizált fogalomrendszer tartalmazza az érdeklődésre számot tartó fogalmakat és kapcsolataikat, a fogalmak meghatározását és a hozzájuk tartozó megjelöléseket, reprezentációkat (kifejezéseket, megnevezéseket, szimbólumokat/kódokat). Egy fogalomhoz különböző nyelveken (de akár egy nyelven belül is) különböző reprezentációk tartozhatnak. A fogalmak meghatározásai megjelenhetnek természetes nyelvi formában, de formalizált nyelven is (pld. matematikai fogalmak definíciói). A formalizált fogalomrendszerek két szintre tagolhatóak: az adott alkalmazási terület fogalmainak absztrakt, egyszerűsített modelljét, fogalmi leképezését (conceptualization) és annak valamilyen formában, nyelven történő reprezentációját (specification).

Témánk szempontjából a *formalizált fogalomrendszerek rendeltetése*, hogy elősegítse a bennük szereplő reprezentációk, kifejezések – szerepük szerint egyedi szemantikus információk – értelmezését, jelentésének meghatározását és támogassa ezen reprezentációk jelentésmegőrző felhasználását, feldolgozását. A kifejezések által hordozott jelentés, a reprezentált fogalom meghatározását egyrészt a definíciók, másrészt a köztük fennálló kapcsolatok segíthetik. Az informatikai eszközökkel történő jelentésmegőrző feldolgozás során viszont egyértelműen a formálisan reprezentált kapcsolatok kerülhetnek felhasználásra.

A *formalizált fogalomrendszerekben szereplő kapcsolatok* három alapvető csoportját az ekvivalencia kapcsolatok, a hierarchikus kapcsolatok és az asszociatív kapcsolatok alkotják (részletesebben lásd 22, 42-57. o.). Az *ekvivalencia kapcsolatok* az azonos (vagy hasonló) jelentésű kifejezések közötti kapcsolatok. Ilyen kapcsolatok állhatnak fent rokon értelmű kifejezések (szinonimák) és alakváltozatok¹⁸ között, de ez lehet a megoldás az alacsonyabb szintű fogalmak fölérendelt fogalmaik formájában történő összevonására (generic/upward posting).

A *hierarchikus kapcsolatok* alapját többféle kapcsolattípus képezheti. Ezek közé tartoznak többek között az általános fogalmak¹⁹ terjedelme közötti viszonyokra épülő alá/fölérendeltségi kapcsolatok; az általános és egyedi fogalmak (osztályok és egyedek) közötti viszonyt leíró kapcsolatok; valamint a fogalmak közötti rész/egész viszonyt leíró kapcsolatok.²⁰ Egy formalizált fogalomrendszer egy időben több hierarchikus kapcsolatot is tartalmazhat (vagyis struktúrája lehet polihierarchikus).

Végül az *asszociatív kapcsolatok* olyan további – fogalmak közötti – szemantikus kapcsolatok, amelyek nem sorolhatóak az előző két csoportba. Ezek fennállhatnak ugyanazon hierarchiában azonos fölérendelt fogalom alá sorolt fogalmak között, vagy különböző kategóriába sorolható fogalmak között. Ez utóbbiak köre szinte korlátlan és alkalmazási terület-függő (pld. folyamat-végrehajtó, tevékenység-eredmény, tevékenység-eszköz, ok-okozat, objektum-tulajdonság, stb.).

A könyvtárakban összegyűjtött információk rendszerezéséhez kapcsolódóan a formalizált fogalomrendszerek a *tudásszervezési rendszerek* fogalma formájában jelentek meg. Ebben a megközelítésben rendeltetésük információhordozó anyagok rendszerezése visszakeresés és a gyűjtemény kezelése céljából. A tudásszervezési rendszer képez hidat a felhasználók információigényei és a gyűjtemény anyagai között. Segítségével a felhasználó azonosítani tudja érdeklődésének tárgyát, anélkül hogy előzetesen tudomása lenne annak létezéséről. [21, 3. o.]

¹⁸ Eltérő írásmódok, tudományos és köznapi megnevezések, rövid és teljes alakok, rövidítések, stb.

¹⁹ Általános fogalom = terjedelmébe több dolog tartozik, egyedi fogalom = terjedelmébe egyetlen dolog tartozik.

²⁰ 'Is-a', 'instance-of' és 'has-a' kapcsolatok.

A formalizált fogalomrendszerek egyszerűbb változatai közé az *ellenőrzött szótárak*, *taxonómiák* és *tezauruszok* tartoznak. Az *ellenőrzött szótár* (controlled vocabulary) egy adott célra alkalmazott kifejezések²¹ és azok egyértelmű, nem redundáns meghatározásainak egy arra jogosult (regisztrációs) hatóság által felügyelt és kibocsátott listája.²² Rendeltetése, hogy biztosítsa a benne szereplő kifejezések egységes, következetes, azonos módon értelmezett használatát.

Amennyiben egy kifejezést különböző fogalmak megjelölésére is használnak, a többértelműség elkerülésére a megnevezéseket minősítő jelzések különböztetik meg. Ha több kifejezés használatos ugyanannak a fogalomnak a megjelölésére, akkor azok közül egy elsődlegesnek lesz jelölve, a többi pedig másodlagosként, szinonimaként szerepel. [22, 5. o.] Ennek megfelelően az ellenőrzött szótárakban szerepelnek ekvivalencia kapcsolatok.

Egy ellenőrzött szótár a nyelv fejlődésével, új fogalmak megjelenésével, használatuk változásával folyamatosan naprakészen kell tartani. Ennek megfelelően az elsődleges és az elfogadott (másodlagos) kifejezések mellett tartalmazhat nem javasolt, illetve elavult kifejezéseket is.²³

A *taxonómia* (taxonomy)²⁴ – más néven osztályozási rendszer – témánk szempontjából dolgok, fogalmak hierarchikus osztályokba sorolása, vagyis egy ellenőrzött szótár kifejezéseinek hierarchikus struktúrája, amelyben minden kifejezés egy, vagy több alá/fölérendelt fogalmi ('szülő/gyerek') viszonyban áll a taxonómia más kifejezéseivel. [22, 9. o.] A taxonómia tulajdonképpen egy alá/fölérendeltségi kapcsolatokkal kiegészített ellenőrzött szótár.

A *tezaurusz* (thesaurus) egy ellenőrzött szótár kifejezéseinek meghatározott rendbe sorolt összessége, amelyben a kifejezések közötti különböző kapcsolatokat egyértelmű, szabványos, egymással párokat alkotó kapcsolatjelzők azonosítják. [22, 9. o.] A tezaurusz tulajdonképpen egy további szemantikus kapcsolatokkal kiegészített taxonómia.

A *taxonómiák*, *tezauruszok szerepe* elsősorban az információk visszakeresésében jelentős. Ennek során felhasználhatóak arra, hogy egy keresett kifejezés esetén ne csak azokat az információegységeket bocsássák rendelkezésre, amelyekben ez a kifejezés szerepel, vagy amelyeket ezzel a kifejezéssel (mint egyedi szemantikus jellemzővel) írtak le, hanem azokat is, amelyekben a keresett kifejezéssel valamely szemantikus kapcsolatban álló kifejezések szerepelnek, vagy ilyenekkel vannak leírva. A taxonómiák, tezauruszok a fenti szerepük mellett a más fogalmakkal fennálló kapcsolataik dokumentálásával segítik a bennük szereplő fogalmak pontosabb értelmezését is.

A hagyományos információkezelésben az XIX. század második felében megjelent hierarchikus könyvtári osztályozási rendszereket a XX. század második felében követték az asszociatív kapcsolatokat is tartalmazó tezauruszok. A taxonómiák a múlt század közepétől megjelentek a különböző gazdasági szakterületeken, majd más állami szférákban, végül a szervezeti szintű alkalmazásban is. A megnövekedett szerep következtében aztán megjelentek a taxonómiák, tezauruszok informatikai eszközökkel történő kezelésének szabványos megoldásai is.²⁵

A formalizált fogalomrendszerek előzőekben bemutatott, elsősorban a könyvtári tudásszervezésben alkalmazott egyszerűbb változatai csak korlátozott mértékben voltak alkalmasak egy

²¹ Pld. dokumentumok leírásához használt tárgyszó lista, egy adatmező lehetséges értékei, vagy akár egy alkalmazási terület szakkifejezései.

²² Az ellenőrzött szótárak leegyszerűsített változatai meghatározásokat sem tartalmaznak, a bennük szereplő kifejezések értelmezését egyértelműnek feltételezik.

²³ Preferred, admitted, deprecated, obsolete terms.

²⁴ A kifejezés másik jelentése rendszertan.

²⁵ Pld. ISO 2788, ISO 5964, BS 8723, ISO 25964.

alkalmazási terület fogalmai közötti kapcsolatok reprezentálására. Az 1970-es évektől kezdődően először a mesterséges intelligencia kutatások során, majd a 90-es évek elejétől más alkalmazási területeken felmerült a tudásreprezentáció hatékonyabb, részletesebb leírást biztosító megoldásainak szükségessége és jelent meg az ontológiák, a szemantikus hálózatok és a fogalomtérképek fogalma.

A taxonómiák, tezauruszok, illetve a következőkben ismertetendő megoldások közötti különbség nem csak a leírás részletességében van, hanem elsődleges rendeltetésükben is. Míg az előbbieket elsősorban az osztályozást, rendszerezést és ehhez kapcsolódóan az információk visszakeresését szolgálják, addig az utóbbiak célja ennél szélesebb körű, a tudásreprezentáció, egy ismeretrendszer modellezése.

Az *ontológia*²⁶ témánk szempontjából egy közösen használt fogalomrendszer egyértelmű, formális specifikációja. Rendeltetése, hogy fogalmi alapot, egységesen elfogadott háttérismereteket biztosítson egy alkalmazási terület számára. Egy ontológia két szinten is egyéni leírása a világnak: először fogalmi szinten, a releváns dolgok (objektumok), tulajdonságok és kapcsolatok fogalmakkal történő leírásában, másrészt ezen fogalomrendszer reprezentációja szintjén.

Egy *ontológia alapvető tartalmi összetevői* közé entitások (dolgok, tulajdonságok és kapcsolatok) azonosítói (az ontológia 'szókészlete'), entitásokból összetettebb entitásokat létrehozó konstrukciós formulák (kifejezések), valamint entitásokra vonatkozó igaznak tekintett állítások (axiómák) tartoznak. A dolgok lehetnek az általános fogalmaknak megfelelő osztályok és az egyedi fogalmaknak megfelelő egyedek. A fenti összetevők mellett az ontológiák tartalmazhatnak a tartalmat nem befolyásoló, de az emberi értelmezést megkönnyítő természetes nyelvi megnevezéseket, megjegyzéseket (annotációkat).

Mivel az ontológia nemcsak általános fogalmak, hanem egyedi objektumok és ezekre vonatkozó állítások leírását is lehetővé teszi, így tulajdonképpen nem csak egy fogalomrendszer, hanem egy erre épülő *ismeretrendszer leírása is* lehet. Ez a bővebb tartalom is szolgálhatja az ontológia által meghatározott fogalomrendszer egységes értelmezését, mivel így a fogalmak meghatározása a más fogalmakkal fennálló viszonyuk és megkülönböztető jegyeik mellett történet a fogalom terjedelmének teljes, vagy részleges megadásával (felsorolásokkal, vagy példákkal).

Az ontológia formális reprezentációjának eszköze az *ontológia(leíró) nyelv*, amely meghatározza az alapvető ontológiai összetevők és a köztük fennálló kapcsolatok körét. Az ontológia(leíró) nyelvek rendeltetésükből következően az elsőrendű (predikátum), vagy a leíró logikai nyelvekre épülnek. Az 1980-as évektől kezdődően számos, eltérő tulajdonságokkal rendelkező ontológia(leíró) nyelv jelent meg, közülük a Szemantikus Web program keretében az RDFS és az OWL²⁷ játszanak kiemelt szerepet.

A szemantikus hálózatok és fogalomtérképek (semantic networks, concept maps), valamint különböző változataik²⁸ a tudásreprezentáció, a fogalomrendszerek leírásának, illetve szemléletes megjelenítésének megoldásai. Valamennyi változatuk fogalmak és fogalmak közötti kapcsolatok grafikus (gráf-alapú) leírására épül, amelyben a csúcsokat a fogalmak, az éleket a szemantikus kapcsolatok képezik. A különböző változatok egymással általában tartalmilag egyenértékűek és mivel a kapcsolatok reprezentációja viszonylag egyszerű, tartalmuk teljes körűen megjeleníthető ontológia formájában is. Egyes típusaiknak informatikai eszközökkel is

²⁶ Az ontológia eredeti értelmezése szerint lételmélet, a filozófiának a létezéssel, annak legalapvetőbb törvényszerűségeivel, a létező dolgokkal és alapvető kategóriáikkal foglalkozó része.

²⁷ RDF Szókészlet Leíró Nyelv RDF Séma (RDF Vocabulary Description Language : RDF Schema), OWL Web Ontológia Nyelv (OWL Web Ontology Language).

²⁸ Fogalmi gráfok (concept graphs), tématerképek (topic maps), gondolatterképek (mind maps), stb.

feldolgozható, szabványos reprezentáció²⁹ is vannak és jelentős ismeretrendszer tartalmazó 'adatbázisok'³⁰ épülnek rájuk.

Az *ontológiák felhasználási lehetőségei* a Szemantikus Web meghirdetésekor is megjelentek.³¹ A bevezető publikáció [7] az ontológiákat – a kiegészítő szemantikus információk és az egységes XML és RDF alapú reprezentáció mellett – a Szemantikus Web harmadik alapvető összetevőjeként határozta meg. Lényegüket a fogalmak taxonómiájában és az erre épülő következtetési szabályokban jelöli meg. A tapasztalatokat összegző következő publikációban [24, 98. o.] már az OWL ontológia(leíró) nyelv ismeretében fogalmazódik meg annak következtetéseket támogató lehetőségeit meghaladó, egységes szabályrendszer leírás (szabályleíró nyelv[ek]) szükségessége.

A formalizált fogalomrendszerek és ezek között kiemelten az ontológiák kiemelt szerepet játszanak a szemantikus informatikai megoldásokban. A szemantikus technológiákban megteremtik az egyedi szemantikus információk egységes alkalmazásának, értelmezésének lehetőségét, biztosítják közösen elfogadott szemantikus információk hozzákapcsolását elemi és összetett információ reprezentációkhoz és a fogalmak közötti kapcsolatok, következtetési szabályok segítségével alapját képezik új információk jelentés-alapú előállításának (ezen belül információegységek tartalmát leíró jellemzők meghatározásának).

ÖSSZEGZÉS

Összegzésképpen megállapíthatjuk, hogy a szemantika megkülönböztető alapját képező *jelentés a nyelvészetben* kialakult fogalom, amelynek értelmezése a szavak jelentéséről kiterjedt a szószerkezetek, mondatok, szövegek jelentésére is. A szavak jelentése több jelentéselemből tevődik össze, amelyek között alapvető szerepet az elsődleges, vagy fogalmi jelentés játszik. A nyelvészet értelmezése szerint a nyelvi jelek a jelentés (fogalom) közvetítésével vonatkoznak, utalnak a jelöltre (jeltárgyra). A szavak, kifejezések jelentésének meghatározása más, ismert jelentésű szavakra (fogalmakra) épülő definíciókkal lehetséges. Az összetett nyelvi formák jelentése összetevőik jelentésére és az összetétel módjára (jelentésére) épül.

A *jeltudomány* (szemiotika) megközelítése szerint a szemantika tágabb értelemben, a nyelvi mellett a társadalomban és a természetben előforduló más jelek jelentéséhez kapcsolódóan is értelmezhető. A jeltudomány alapvető megállapítása, hogy a jelek nem önmagukban, hanem meghatározott körülmények, viszonyok között töltik be szerepüket. A jelviszony összetevői közé a jelentést hordozó jelhordozó, a jeltárgy (amit a jel jelöl, vagy helyettesít) és az értelmező, jelfelhasználó tartoznak. A jelviszony keretében a jel mindig a jelértelmező tudatában megjelenő fogalom közvetítésével jelöli meg a jeltárgyat, a jelentés tehát első fokon a jel és a fogalom közötti kölcsönviszony.

Az *informatika* esetében az alapvető kérdés az információkat hordozó, reprezentáló adatok jelentésének értelmezése. Ez a jelentés nem önmagában rejlő tulajdonság, csak az adatokat kezelő (előállító, felhasználó) emberekhez kapcsolódóan értelmezhető. Ennek megfelelően szándékolt, illetve a felhasználás során értelmezett jelentésről kell beszélnünk. Az informatika szempontjából egy adat jelentése alatt általában az emberi tudatban ahhoz kapcsolódó információt (belső mentális reprezentációt) célszerű értenünk. Szemléletes információ-reprezentációk (álló- és mozgóképek, hangfelvételek) esetében tágabb értelemben a jelentésükhöz soroljuk a hozzájuk kapcsolódó fogalmi információk jelentését is (mit "tartalmaz", mit "ábrázol", mit "fejez ki").

²⁹ Fogalmi gráfok = ISO 24707, tématerképek = ISO 13250.

³⁰ Pld. WordNet = angol szavak adatbázisa (~155 ezer szó, ~200 ezer jelentés, ~110 ezer szemantikus reláció).

³¹ Részletesebben lásd [23, 361-420. o.].

Az adatok jelentésének szerepe az információtárolásban és az információtovábbításban másodlagos és nem érdemi a technikai eszközökkel történő információszerzésben (adatgyűjtésben) sem. Megkerülhetetlen viszont az információk adatokká alakításában, illetve az adatok felhasználásában, értelmezésében, ahol a jelentés korábban csak a személyek tudatában és hagyományos formában, az informatika hatókörén kívül állt rendelkezésre. Végül egyre jelentősebb ez a szerep az adatok feldolgozásában, átalakításában is, ahol korábban a jelentés a feldolgozó algoritmusokban, programokban elrejtve került felhasználásra.

Az adatok jelentés-alapú, közös értelmezésre épülő informatikai kezelése az adatok és az általuk hordozott jelentés közötti kapcsolatra vonatkozó kiegészítő információkat, adatokat igényel. Az ezen adatok által hordozott információkat nevezzük szemantikus információknak és az ilyen információkat (is) kezelő műveleteket tekintjük szemantikus műveleteknek.

A *szemantikus információk* értelmezése az ismeretek elsődleges és meta-szintjei fogalmára építhető. Egy adott szakterület ismeretrendszerének alapvető csoportját az érdeklődésre számot tartó objektumok tulajdonságaira, viszonyaira vonatkozó elsődleges, ún. objektum-szintű információk alkotják. Az ismeretek második szintjét az elsődleges információkra vonatkozó metainformációk képezik, amelyek közé a szakterület fogalomrendszerére vonatkozó, illetve az elsődleges információk értelmezését, feldolgozását, felhasználását befolyásoló kiegészítő információk tartoznak. A metainformációk az informatika számára metaadatok formájában állnak rendelkezésre.

Szemantikus információ alatt információ-reprezentációk, információs erőforrások értelmezését, jelentésmegőrző feldolgozását segítő, jelentéshez kapcsolódó metainformációt értünk. A szemantikus információk két nagy csoportját a szakterület fogalomrendszerére vonatkozó általános szemantikus információk, illetve az elsődleges információkat minősítő, azok értelmezését, feldolgozását, felhasználását befolyásoló egyedi szemantikus információk alkotják.

A fentiek alapján a *szemantikus műveletek* adatokkal végzett olyan műveletek, amelyek az érintett adatok tartalmára, jelentésére vonatkozó (szemantikus információkat hordozó) adatokat is felhasználnak, vagy eredményeznek. Amennyiben a szemantikus információkat és az ezeket hordozó (meta)adatokat a reprezentációs szinttől (mentális, illetve külső) elvonatkoztatva szemantikus jellemzőnek nevezzük, a szemantikus műveletek szemantikus jellemzők felhasználásával végrehajtott műveletek, amelyek eredménye lehet szemantikus, vagy 'hagyományos' jellemző.

Az *egyedi szemantikus információk* konkrét adatokhoz, információhordozókhoz kapcsolódó, vagy annak részét képező kiegészítő információk, amelyek rendeltetése, hogy elősegítsék a tartalom, jelentés megismerését, a jelentés-alapú feldolgozást. Két nagy csoportjukat az elemi adatok jelentéséhez kapcsolódó, illetve a dokumentumok tartalmát leíró jellemzők alkotják. Ez utóbbiakkal napjaink informatikájában elsősorban a címkézés formájában találkozhatunk. Az tartalmat leíró jellemzők hozzárendelése történhet manuális, vagy automatizált módon, informatikai megoldások segítségével.

Az *általános szemantikus információk* alapvető csoportját a formalizált fogalomrendszerek alkotják, amelyek többek között az egyedi szemantikus információk értelmezését, felhasználását biztosítják. A formalizált fogalomrendszerek egy adott témakör fogalmai és sokrétű kapcsolataik absztrakt modelljei és ezek valamilyen nyelven történő reprezentációi. Rendeltetésük, hogy elősegítsék a bennük szereplő kifejezések értelmezését, támogassák ezek jelentésmegőrző felhasználását, feldolgozását. A formalizált fogalomrendszerekben szereplő kapcsolatok három alapvető csoportját az ekvivalencia kapcsolatok, a hierarchikus kapcsolatok és az asszociatív kapcsolatok alkotják. Egyszerűbb változataik közé az ellenőrzött szótárak, taxonómiák és teauruszok, összetettebb kapcsolatokat is tartalmazó változataik közé az ontológiák, szemantikus hálózatok és fogalomtérképek, stb. tartoznak.

Felhasznált irodalom

- [1] SZÍVÓS Mihály: *A jeltől a kódig. Rendszeres szemiotika.* – Loisir Könyvkiadó, Budapest, 2012.
- [2] RIEMER, Nick: *Introducing semantics.* – Cambridge University Press, New York, 2010.
- [3] PETHŐ József: *Jelentéstan.* – Bölcsész Konzorcium, Budapest, 2006.
- [4] KÁLMÁN László-TRÓN Viktor: *Bevezetés a nyelvtudományba. Második bővített kiadás.* – Tinta Könyvkiadó, Budapest, 2007.
- [5] *ISO/IEC 2382-1:1993 (E/F), Information Technology – Vocabulary – Part 1: Fundamental terms. Third Edition.* – 1993.
- [6] MUNK Sándor: *Katonai informatikai rendszerek interoperabilitásának aktuális hadtudományi kérdései. MTA doktori értekezés.* – Budapest, 2007.
- [7] BERNERS-LEE, Tim-HENDLER, James-LASSILA, Ora: *The Semantic Web.* – Scientific American, 2001 (284)/5. (34-43. o.)
- [8] POLIKOFF, Irene-ALLEMANG, Dean: *Semantic technology. Version 1.1. TopQuadrant Technology Briefing.* – TopQuadrant Inc., 2003.
- [9] SHETH, Amit-RAMAKRISHNAN, Cartic: *Semantic (Web) Technology In Action: Ontology Driven Information Systems for Search, Integration and Analysis.* – *IEEE Data Engineering Bulletin*, 2003 (26.)/4. (40-48. o.)
- [10] DAVIS, Mills: *Semantic Wave 2006 Part 1: Executive Guide to Billion Dollar Markets.* – Project10X, Washington, 2006.
- [11] DAVIS, Mills: *Semantic Wave 2008: Industry Roadmap to Semantic Web 3.0 & Multibillion Dollar Market Opportunities. Executive Summary.* – Project10X, Washington, 2008.
- [12] HANSEN, Bjørn Jervel-GAGNES, Tommy-RASMUSSEN, Rolf-RUSTAD, Marianne-SLETTEN, Geir: *Semantic technologies.* – Norwegian Defence Research Establishment (FFI), Kjeller, 2007.
- [13] HANSEN, Bjørn Jervel-HALVORSEN, Jonas-KRISTIANSEN, Svein Ivar-RASMUSSEN, Rolf-RUSTAD, Marianne-SLETTEN, Geir: *Recommended application areas for semantic technologies.* – Norwegian Defence Research Establishment (FFI), Kjeller, 2010.
- [14] MATTHEWS, Brian: *Semantic Web Technologies. JISC Technology and Standards Watch Report.* – Joint Information Systems Committee, Bristol, 2005
- [15] *W3C Semantic Web Activity.* – World Wide Web Consortium, 1998.
<http://www.w3.org/2001/sw/> (letöltve: 2014.03.21.)
- [16] FEIGENBAUM, Lee: *Semantic Web vs. Semantic Technologies.* – Cambridge Semantics. Semantic University.
<http://www.cambridgesemantics.com/hu/semantic-university/semantic-web-vs-semantic-technologies> (letöltve: 2014.03.21.)
- [17] *ISO/IEC 11179-1:2004(E) Information Technology – Metadata Registries (MDR) – Part 1: Framework.* – International Organization for Standardization, Geneva, 2004.
- [18] *Understanding Metadata.* – National Information Standards Organization, Bethesda, 2004.

- [19] *ISO/IEC 2382-1:1993 (E/F) Information Technology – Vocabulary – Part 1: Fundamental Terms.* – International Organization for Standardization, Geneva, 1993.
- [20] PETERS, Isabella: *Folksonomies. Indexing and Retrieval in Web 2.0.* – Walter de Gruyter, Berlin, 2009.
- [21] HODGE, Gail: *Systems of Knowledge Organization for Digital Libraries. Beyond Traditional Authority Files.* – The Digital Library Federation, Washington, 2000.
- [22] *ANSI/NISO Z39.19-2005, Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies.* – National Information Standards Organization, Bethesda, 2005.
- [23] SZEREDI Péter-LUKÁCSY Gergely-BENKŐ Tamás: *A szemantikus világháló elmélete és gyakorlata.* – Typotex, Budapest, 2005.
- [24] SHADBOLT, Nigel-HALL, Wendy-BERNERS-LEE, Tim: The Semantic Web Revisited. – *IEEE Intelligent Systems*, 2006 (21)/3, 96-101. o.