

8. JELENTÉSALAPÚ TECHNOLÓGIÁK

Krauth Péter

A legkülönbözőbb formákban és helyeken tárolt információk egymáshoz, valamint az emberi „tudáshoz” való viszonyát leíró METAADATOK hatékony és sokoldalú kezelése ezen információk egyre automatikusabb és kifinomultabb feldolgozását teszi lehetővé.

1. Megnevezés és rövid leírás

Az Internet megjelenése nagyságrendekkel megnövelte az elérhető információk körét és változatosságát, ezért az Internet továbbfejlődéséhez elengedhetetlen, hogy az adatokat tömegesen és automatikusan lehessen kezelni. Az adattömegeket azonban csak akkor – és különösen akkor – lehet hatékonyan kezelni, ha pontos képünk (információnk) van a kezelendő adatokról, és nemcsak nekünk, hanem – ha áttételesen is – ezen keresztül a gépi rendszereknek is. Meghatározó ezért az adatokról szóló információk kezelésének, azaz a METAADAT-kezelés lehetőségeinek kiterjesztése és teljesítményének növelése: ti. a számítógép kizárólag az ilyen METAADATokon keresztül „tudhat” bármit is a feldolgozandó adatokról.

A ma KERESŐMOTORjai azért vannak lehetetlen helyzetben, mert a keresendő tartalmak nincsenek kellőképpen „felruházva” leíró, értelmező információkkal. Ennek kiküszöbölésére indították útjára az évezred elején a Szemantikus Web kezdeményezést: lássuk el fokozatosan a web-tartalmakat metainformációkkal, azaz olyan információkkal, amelyek megadják azok szerkezetét, belső összefüggéseiket, szabályaikat és külső kapcsolataikat. A kezdeményezés eredményeképp már megjelentek azok az alapszabványok és -technológiák, amelyek az ilyen leírások egységességét biztosítani tudják (XML, RDF, OWL).

2. Jelenlegi helyzet

A jelentésalapú technológiák jelenleg két fő területen az ember-gép ill. gép-gép közötti *kommunikáció*, valamint a tömeges, heterogén, illetve nem jól strukturált *adatok kinyerése és elemzése* területén mutatják meg az előrelépés irányát. A továbbiakban először ezt a két területet tekintjük át, majd a Szemantikus Web és az ehhez kapcsolódó formális logikai megközelítésekre térünk ki.

Kommunikáció

Az emberek azért „beszélnek el” időnként egymás mellett, mert adott dolgokról egészen másként vélekednek, máshogy képzelik el, mást tételeznek fel, más a természetes számukra: azaz a szituációról (öntudatlanul) alkotott modelljük jelentősen eltér egymástól. *A jelentésalapú technológiák egyik – gyakran ki nem mondott – hipotézise az, hogy minél közelebb van a kommunikáló felek „mentális” (tulajdonképpen belső) modellje a szituációról, annál könnyebben és hatékonyabban tudnak kommunikálni.* Ezért is van szükség tudásreprezentációs és következtetési képességekre: a tudásreprezentáció szintje meghatározza az ilyen belső modellek minőségét és részletezettségét, míg a következtetési képesség a származtatott (közvetett) információk kinyerésének mértékét – vagy egyáltalán lehetőségét – biztosítja.

Adatkinyerés és -elemzés

Az Internet és a mobiltechnológia elterjedése nagyságrendekkel megnövelte az elérhető információk körét. Azonban a fejlődés a hálózatok és a hardver területén messze meghaladja a segítségükkel létrehozott adatok kezelésének képességét. Az adatok még mindig túlságosan nyers állapotban vannak, emellett gyakran hozzáférhetetlenek, ellentmondásosak, széttagoltak, és kihasználtságuk rendkívül alacsony. Az információkeresés, adatabányászat, adatintegráció és -tisztítás javít ezen a helyzeten: javuló keresési és elemzési képességekkel egyre inkább és egyre hatékonyabban hatolnak be az adattömegekbe. A keresőgépek gyártói, statisztikai módszereket használó vállalkozások és üzletiintelligencia-rendszerek készítői rendszeresen átvizsgálják a vállalatok és intézmények adattárait. A weben pedig a Google-nak sikerült figyelemre méltó eredményeket elérnie. De önmagukban a keresési és elemzési funkcióknak mindig meglesznek a maguk korlátai. Bármennyire „okosak” is az algoritmusok, mindig is lesznek ellentmondások és kétértelműségek az adatokban.

Minden olyan esetben, amikor a keresési és elemzési funkciók nem működnek megfelelően, rendelkezésre áll az a lehetőség, hogy jobban meg kell szervezni az adatokat, és ki kell bővíteni azokat METAADATokkal (adatleíró adatokkal). *A METAADATok sokkal egyszerűbbé teszik a számítógépek számára az adatok automatikus feldolgozását és értelmezését.* Még pontosabban: éppen a METAADATok teszik ezt egyáltalán lehetővé. A METAADATok összegyűjtése azonban nem mindig egyszerű, és ebben is a jelentésalapú technológiák tudnak segíteni.

Szemantikus Web és formális logika

A Szemantikus Web koncepcióját Tim Berners-Lee-nek tulajdonítják (*Weaving the Web, 1999*). E koncepciónak két fő eleme van: egyrészt METAADATokat kell kapcsolni az egyes erőforrásokhoz a web-en, másrészt következtetni kell tudnunk ezeknek a METAADAToknak a segítségével. METAADAT az XML (eXtensible Markup Language) leíró nyelvvel megadható *egyszerű struktúrája* is az információnak, bonyolultabb esetekben azonban *információmodelleket* vagy *ontológiákat* kell használni.

A már említett XML a 90-es évek végén jelent meg. Ez lehetővé tette, hogy bármilyen adatáramot és adatforrást viszonylag egyszerűen leírjanak, szintaktikai szerkezetbe rendezzenek, ugyanakkor alapvető szemantikai jegyekkel is ellássanak (pl. az

adatelemek neve, keletkezésének körülményei). Az XML gyökerei a 60-as és 70-es évekre vezethetők vissza, amikor az 1986-ban megjelent ISO 8879 jelű dokumentumszabvány (Standard Generalized Markup Language – SGML) fő elvei megfogalmazódtak. Az XML lényegében ennek leegyszerűsített és „web-esített” változata. Nagy előnye, hogy nem kell feltétlenül külön (mintegy „kívülről”) megadni a szerkezetet, majd azt használva az adatokat megadni, hanem az adatmegadással együtt lehet a szerkezetet is leírni. Ez – egyszerű esetekben – jelentősen lecsökkentheti az adatkezeléshez szükséges előkészületek idejét.

A másik áttörés, amely ma a Szemantikus Web koncepciójában megtestesül, szintén jóval korábban történt – lényegében észrevétlenül a szakma nagy része számára. A 80-as években a formális logika különböző irányzatai – pl. keretalapú rendszerek (*frame-based systems*), Prolog-megvalósítások, leíró logikai (*description logics*) kísérletek – kutattak a mesterséges intelligencia területén az ismeretreprezentáció tökéletesítése céljából. A formális logika ezen irányzatai adják az elvi megalapozást az XML legutóbbi „leszármazottainak”, mint pl. a W3C (World Wide Web Consortium) ún. erőforrás-leíró keretnyelvének (Resource Description Framework – RDF) és a webontológiák nyelvének (Web Ontology Language – OWL).

Az OWL feltétlenül előremutató jelenség az internet fejlődésében: az RDF és az RDF Schema lehetőségeit tágítja ki az internet információs objektumainak leírásában, és egységesíteni képes a sok különböző, szerteágazó ontológianyelvet. Tulajdonképpen ontológia egy-egy adatbázis leírása (az ún. adatbázisséma): egy adott alkalmazási rendszer kontextusában bizonyos dolgokra (pontosabban az azokról alkotott fogalmainkra) hivatkozik, megadja ezen fogalmak felépítését (részfogalmak, tulajdonságok) valamint egymáshoz való kapcsolódásaikat – szokásosan táblák, attribútumok, relációk, megkötések stb. formájában.

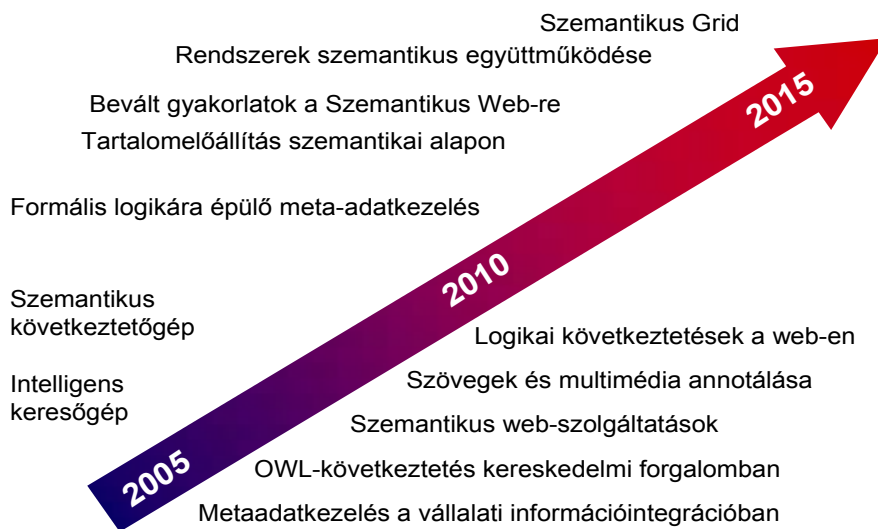
Az adatbázissémák azonban nagyon kötődnek egy-egy konkrét alkalmazási szituációhoz, ennél többre, általánosabbra van szükség. Egyes tudás-intenzív területeken (pl. orvostudomány) ezért az elmúlt egy-két évtizedben megjelentek olyan leírások, amelyek egy-egy konkrét alkalmazási szituációtól függetlenül, általánosan igyekeztek visszaadni az adott szakterület fogalmait, fogalmi rendszerét: ezeket nevezték ontológiáknak. A *szakterületi tudás fogalomrendszerének externalizálása és kodifikálása* az az irány, ahova az ontológiák kezdeti kiépítésével az informatika egyes területeken elmozdult.

Az OWL több ponton túllép a szokásos sémanyelvek képességein. Az egyik ilyen továbblépési pont, hogy ún. leíró logikai nyelvet (*Description Logic – DL*) is magába foglal, pontosabban arra épül. DL-nyelven 1) sokkal absztraktabb fogalmakat is meg lehet ragadni informatikai eszközökkel, mint amilyeneket egyébként a sémanyelvekkel lehet; 2) a fogalmakat egyes példányaiktól függetlenül lehet kezelni; és ami a legfontosabb, 3) logikai következtetések tárgyaként is szolgálhatnak az ilyen leírások. Egyszóval az OWL igyekszik ötvözni – az informatika jelenlegi fejlettségi szintjén – az ember két fontos mentális képességét: a tudásreprezentációs és a következtetési képességét. A Szemantikus Web-en tehát az információforrások olyan egységes ontológianyelven lehetnek leírva, amely bizonyos szempontból közel áll az emberi elme egyes fontos képességeihez.

3. A várható fejlődés eredményének jellemzése

Az ún. szemantikus (tulajdonképpen jelentésalapú) technológiák várhatóan már a vizsgált időszakban jelentős hatással lesznek a tartalomelőállításra. Ezen túlmenően azonban ezeket a technológiákat nemcsak az információkeresésben és adatelemzésben, hanem általánosabban informatikai rendszerek együttműködésének javításához is fel fogják használni (semantic interoperability), sőt az ember-gép kapcsolat magasabb szintjeinek eléréséhez is.

2010-re a formális logika válik (elsősorban a fent említett megnyilvánulási formáiban: RDF, OWL, DL) a legjelentősebb METAADATkezelési kezdeményezések többségének közös alapjává.



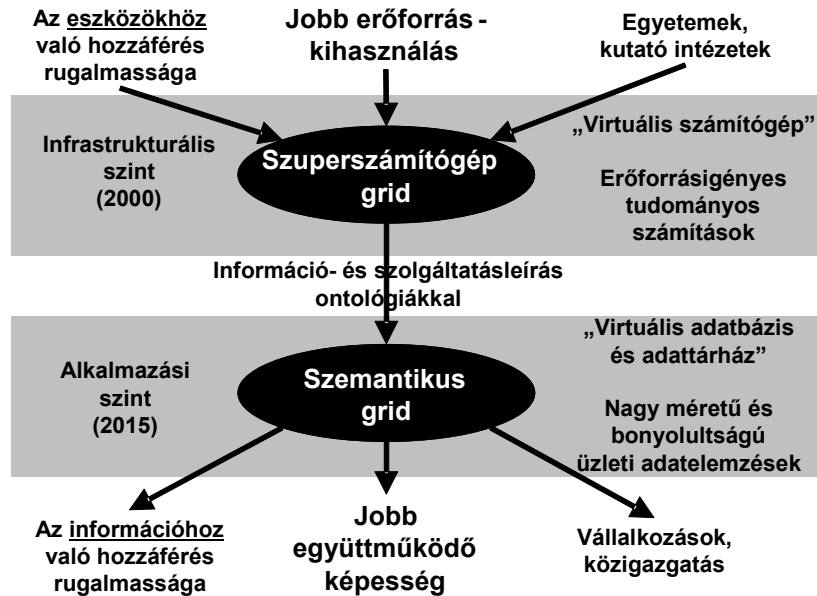
1. ábra: Mérföldkövek a jelentésalapú technológiák fejlődésében

A Szemantikus Web kontextusában várható, hogy a leglátványosabb, rövid távú sikerek az elektronikus kereskedelemben jelentkeznek. Sok időnek kell azonban még eltelnie, és inkább a 2010-15 közötti időszakra tehető, amikor a leszűrődött tapasztalatok világos útmutatást tudnak adni, hogy milyen esetekben igazolható a Szemantikus Web-hez kapcsolódó (és általában a jelentésalapú) technológiák költsége, és mikor nem.

Itt célszerű megemlíteni azt a tendenciát is, hogy a következő 10 évben a *GRID* technológia várhatóan megjelenik *alkalmazási szinten* is, amelynek eredményeképp „szuper-adatbázisokat”, azaz olyan virtuális adatbázisokat vagy adattárházakat lehet majd létrehozni hálózatba kötött, egyedi adatforrások integrálásával, amelyek nagy méretű és bonyolultságú üzleti adatelemzéseket tesznek lehetővé.

E jövőbeni GRID technológia alapegységei a hálózat olyan képességekkel felruházott csomópontjai lesznek, amelyeknek lényeges vonása lesz, hogy a hálózat különböző pontjain lévő információforrásokat ún. *ontológiákkal* írják le, képezik le egymásra,

azaz kapcsolják össze. A távolabbi jövőben különösen fontossá válik az ontológiatartalmak késleltetés nélküli átvitele (leképezése) az egyes csomópontok között, azaz az ún. *valós idejű ontológiák* szerepe megsokszorozódik a beágyazott (pl. mobil ill. környezeti) rendszerekben.



2. ábra: A szemantikus GRID kialakulása

4. Szükséges technológiai előfeltételek

Vannak olyan tényezők, amelyek egyelőre akadályozzák a jelentésalapú technológiák elterjedését rövid távon.

Megegyezés: A METAADAT-szabványok elterjedése közmegegyezést igényel, és ez önmagában hosszadalmas, lassú folyamat. *A szabványoknak a gyakorlatban is jól használható, népszerű ill. keresett eszközökben kell megjelenüniük.*

Felkészültség: A Szemantikus Web alkalmazásaihoz szükséges gyakorlati tapasztalatok még csak most kezdenek gyűlni. Kevés embernek van ma még olyan szakértelme, hogy meg tudná határozni: milyen és mennyi METAADATra van szükség egy adott üzleti alkalmazáshoz. *A METAADATkezeléssel kapcsolatos tevékenységek elismertségének növekedni kell.*

Bonyolultság: Egyes területeket közismerten nehéz strukturálni és formalizálni (például szervezetek együttműködése, politika és jog). Ez nem fog a közeljövőben sem megváltozni. Más területeken azonban jelentős előrelépések történtek – számítógéppel is kezelhető – terminológiakialakításban és fogalomalkotásban (pl. élettudományok, gyógyászat és számvitel). *Az ontológiaépítés és -kezelés eszközeinek be kell vonulniuk az informatika mindennapi gyakorlatába.*

SKÁLÁZHATÓSÁG: Nagy mennyiségű adat esetén a logikai következtetés elveinek alkalmazása lehet, hogy nem ad elfogadható eredményt (pl. a számítások túlságosan lassúak lesznek) még akkor is, ha ezt a szokásosnál könnyebben kezelhető, és gépi számításokra alkalmasabb logikák – pl. a leíró logikák (description logics – DL) – támogatják. *Szükség van olyan következtetési mechanizmusokra, amelyek az adott probléma jellegéhez illeszthetők vagy választhatók.*

Ellenőrizhetőség: Ma még nincsenek kialakult módszerek és eszközök az ontológiák tartalmának ellenőrzésére. Az adatbázissémák esetén a feladat egyszerűbb, mert mindig van egy alkalmazási rendszer, amelynek tesztelése a séma ellenőrzését is jelenti. *Az ontológiaépítő eszközöknek támogatniuk kell mind az automatikus (ellentmondás-kiszűrő és összehasonlító) elemzéseket az ontológiákon, mind az emberi szakértők kollaboratív bevonását az ellenőrzésbe.*

5. Folyamatban lévő kutatások, fejlesztések

Sok pénzügyi-számviteli folyamat még mindig lényegében manuális: az emberek kimásolnak adatokat egy alkalmazásból, és beteszik egy másikba, vagy – még rosszabb esetben – begépelik az adatot különböző alkalmazásokba. Egy nemrégiben kiadott specifikációsorozat (Extensible Business Reporting Language – XBRL) azt ígéri, hogy kodifikálni⁶ képes egy vállalat pénzügyi információit, és ez lehetővé fogja tenni, hogy a legtöbb adatátvitel teljesen automatizált legyen. Ez az automatizáltság mellett jobb adatminőséget eredményez, és együttesen gyorsabb pénzügyi jelentéskészítéshez, alacsonyabb költségekhez és nagyobb átláthatósághoz vezet. Ha elterjedne, az jelentős mérföldkő lenne a pénzügyi jelentéskészítésben – és ez jelzés értékű lenne sok más területnek is.

A METAADATOK a kereső motorokat és az üzletiintelligencia-szoftvereket is segítik. Olyan kiegészítő információkat hordozhatnak, amelyek a keresés során további „fogódzókat” jelentenek akár kulcsszóra, akár taxonómiára épül a keresés. A címkék elősegíthetik a jobb osztályozást és az előrejelző elemzést is. Néhány korai előnye már most látható az ún. helyalapú (location-based) keresési szolgáltatások megjelenésében, mint pl. a Yahoo Maps esetében, és szabványosított helycímkék új kereskedelmi lehetőségeket teremthetnek. A Google Froogle-ja a katalógusszolgáltatóktól megköveteli, hogy tartalmakat címkézzék fel a W3C RDF-je szerint. Vállalati kontextusban több olyan korai fejlesztési tevékenység is megfigyelhető, amely során a cégek javítják az információbeszállítási láncot – különösen a VÁLLALATI INFORMÁCIÓINTEGRÁCIÓ (ENTERPRISE INFORMATION INTEGRATION – EII), a vállalati tartalomkezelés (Enterprise Content Management –

⁶ Azaz informatikailag feldolgozható és vizsgálható módon formálisan leírni egy szervezet adatainak valamilyen jól meghatározott körét. Ez több, mint a szóba jövő informatikai rendszerek adatbázissémáinak összessége, mert olyan fogalmakat is leír, amelyek egyik sémában sem, vagy nem közvetlenül jelennek meg; viszont tartalmazza e fogalmak leképezéseit is az adatbázisokra. Pl. ide tartozhat annak leírása, hogy mit jelent és hogyan állítandó elő egy adott pénzügyi mutató értéke a szervezet adatbázisaiból; ez a gyakorlatban gyakran nem nyilvánvaló, és többféleképpen is történhet, ami különböző anomáliákat okozhat a szervezetek pénzügyi jelentéseiben.

ECM) és a javuló keresőgépek terén, és ezek mindegyike valamilyen formában használ jelentésalapú technológiákat.

Egyes esetekben a METAADATokat az adat – manuális vagy adatgyűjtő folyamaton keresztül történő – létrehozásakor kell inkább meghatározni és rögzíteni, és egyes alkalmazások a kimeneti adatáramuk METAADAT-elemekkel való kiegészítését jól el tudják végezni. A strukturálatlan adatok terén (például szöveges dokumentumok) bizonyos típusú METAADATok (például tárgyak, helyek, személyek, nevek és egyes kapcsolatok) *majdnem teljesen automatikusan kinyerhetők* nyelvi elemzés alkalmazásával – ezt a megközelítést információkinyerésnek is nevezik. Az ezen a területen működő gyártók közé tartozik az Attensity, a ClearForest, a Cymfony, az Insightful, az Inxight, a Lockheed Martin, a SAS, a Semagix, az SPSS, az SRA, a Temis és a Verity.

Számos szoftvergyártó ezen a területen már több mint 3 éve jelen van, de továbbra sem rendelkeznek meggyőző referenciapéldákkal. Ez változhat a következő 2 évben, ha a nemrégiben létrejött partnerkapcsolatok lökést adnak e piaci szegmensnek. 2004 elején az izraeli Unicorn lépett partneri kapcsolatba az IBM-mel, hogy egy integrált eszközínálatot alakítsanak ki az IBM DB2 Information Integratora köré. 2004 közepén a németországi Ontoprise állt össze a Software AG-vel, hogy OWL-következtetési funkcionalitást integráljon az XML-Server-ébe. A gyártók az élenjáró METAADAT-kezelési megközelítéseiket két területen alkalmazzák:

Információkeresés: Itt olyan cégek aktívak, mint az Aduna, az Empolis (Bertelsmann-érdekeltség), az Intelligent Views, az Intelligenxia, a Language & Computing, az Ontology Works, az Ontoprise és a Semagix.

Vállalati információintegráció: Az ezen a területen aktív cégek közé tartozik az Intellidimension, a MetaMatrix, a Network Inference, az Ontoprise (a Software AG-vel közösen) és az Unicorn (az IBM DB2 Information Integrator-ával).

NYÍLT FORRÁSKÓDÚ eszközök is kezdenek figyelmet kapni. A Protégé, az OntoEdit és a Chimera népszerű ontológiaépítő eszközök. A HP Jena nevű eszköze pedig egyike a legkedveltebb következtető motoroknak a szemantikus értelmezés, fordítás területén.

Az EU jelenleg több, mint 70 millió €-val 15 új projektet támogat a „jelentésalapú ismeretkezelő rendszerek” (semantic-based knowledge systems) stratégiai témakörben, amelyben több mint 200 szervezet vesz részt.

A hangsúly ezeknél a fejlesztéseknél a szemantikus web szabványain és ezek terjesztésén (KNOWLEDGE WEB, KB20), a szemantikus web-szolgáltatásokon (DIP), az ontológiákon és METAADATkezelésen (SEKT, METOKIS), az eddiginél intelligensebb keresőgépeken (ALVIS), a szövegek és MULTIMÉDIA-objektumok „rejtett” tartalmának feltárásán és kezelésén (ACEMEDIA, DIRECT-INFO, SIMAC, NEWS, MUSCLE, AIM@SHAPE), újfajta logikai formalizmusokon (ASPIC) és logikai következtetéseken (REVERSE) van.

6. Az IKT más területeire való hatások bemutatása

Jelentős átfedés van – és így kölcsönös egymásra hatás várható – a web-szolgáltatások területén a *szolgáltatás-orientált alkalmazáskészítéssel*. Emiatt egy-egy termék vagy projekt esetén nehezen dönthető el, hogy melyik technológiai kategóriába tartozik inkább. Általánosságban kijelenthető, hogy a jelentésalapú technológiák elsősorban az egyes rendszerkomponensek egymás közötti, illetve a felhasználóval folytatott *kommunikációjának* minél magasabb (értsd: az emberi fogalomrendszerekhez és logikai következtetési képességekhez közelítő) szintjét igyekeznek elérni, míg a szolgáltatás-orientált technológiák az új *rendszerarchitektúra* és az ezen működő alkalmazások létrehozásának és működtetésének *eszközeire* irányulnak.

A jelentésalapú technológiák igen erőforrásigényesek. Emiatt könnyen lehet, hogy a következő nagy húzóerőt jelentik a *számítási és tárolási kapacitások* további minőségi növelése irányában. A MULTIMÉDIA objektumok, majd a mobil, illetve környezeti (*ambient*) eszközök megjelenése és elterjedése elsősorban mennyiségi növekedést indukált. Az emberi befogadó képesség és a rendelkezésre álló idő végessége azonban természetes határt szab a meghallgatható zene, a megnézhető videók és a lebonyolítható telefonbeszélgetések mennyisége tekintetében. További kapacitásigény-növekedés csak az információfeldolgozás elmélyülésétől, bonyolultságának növekedésétől várható, és ebben a jelentésalapú technológiák fontos szerepet kaphatnak.

A jelentésalapú technológiák nagy hatással lehetnek az *ember-gép kapcsolat* fejlődésére. A jelentésalapú technológiák lényege, hogy tudásreprezentációs és következtetési képességekkel ruházza fel a számítógépet, mert az ember-gép kommunikáció a jövőben várhatóan ezeken a területeken tud csak további hatékonysági áttöréseket elérni. Egyre inkább érezni, hogy pusztán érzékszervi képességeinkre és ezek minél teljesebb kihasználására alapozó kommunikációs megközelítéssel áttörés nem érhető el az ember – fentebb említett – befogadási korlátai miatt. Az „okos” ember azonban több információt „lát meg” az érzékszervi információkban, mint amennyi közvetlenül azokban látszik. Több és mélyebb ismeretet tud leszűrni ezekből, mert jobban össze tudja kapcsolni korábbi ismereteivel. Így természetes módon vetődik fel a „*mentális*” *kommunikáció* lehetősége a rendszerekkel. Ahhoz, hogy egy adott – esetleg nagyon leszűkített – szituációban hatékony kommunikációt lehessen folytatni egy rendszerrel (pl. nagy mennyiségű átfogó információ időegység alatt; közvetett és részletes információk igény szerint) az kell, hogy magának a rendszernek is legyen egy megfelelő modellje az adott szituációról (időpont, témakör, előtörténet stb.).

7. Társadalmi-gazdasági hatások elemzése

A hatékony vállalati információ- és ismeretkezelés megteremtése és működtetése tudásigényes feladat és nem kis munka. Mindez látszólag ellentmondásban van azzal, hogy a számítógépek, informatikai rendszerek egyre több munkát végeznek el az ember helyett, és talán arra is lehet gondolni, hogy az embereknek nem marad munka. Azonban: „a munka nem vész el, csak átalakul”. Azaz a konkrét adatok szintjén az

informatikai rendszerek egyre több munkát végeznek ugyan el, ugyanakkor – ezzel egy időben – *az ember tevékenységének jellege átalakul, mert a munkája a METAADATok kezelésére kell, hogy fókuszáljon.* Ehhez a jelentésalapú technológiák segítségével meg kell teremteni és működtetni kell azt a szemantikus (tulajdonképpen értelmezési) környezetet, amely nélkül egyre növekvő mértékű és hatékonyságú gépi adatfeldolgozás nem lesz lehetséges. Jó hír, hogy ennek eredményeképp az emberek egyre növekvő mértékben kell, hogy foglalkozzanak *a szervezeti tudás kezelésével.*

A vállalati informatika illetve annak összhangja az üzletvitellel, a korszerűsége, az integráltsága és nem utolsósorban a várható változásokkal szembeni rugalmassága, jelentős mértékben elő tudja segíteni az ilyen vállalatok értékét hordozó „tudás” externalizációját, azaz *azt a folyamatot, amely egy vállalatban folyamatosan arra irányul, hogy az alkalmazotti és kapcsolati „tőke” – az adott lehetőségek mellett, amilyen mértékben lehet – független legyen e „tőke” természetes hordozóitól, azaz az egyénektől.*

A tudásalapú szervezetek irányításában és végső soron értékének megőrzésében – az informatikai rendszerekbe történő beépülésükön keresztül – a jelentésalapú technológiáknak megvan a maguk szerepe a szervezeti adatvagyon megfelelő szintű kezelésével és nem utolsósorban a METAADAT-kezelés és üzleti-informatikai kommunikáció fejlesztésével.

A tudásalapú társadalomban és gazdaságban várhatóan már nem az információhoz és ismeretekhez való hozzájutás lesz a fő kérdés, hanem a releváns információ, a használható ismeret kiszűrésének, és ezek beintegrálásának a képessége a meglévő keretekbe. A jelentésalapú technológiák természetüknél fogva elősegítik ennek megvalósulását.

8. Magyar vonatkozások

Az EU 6. keretprogramjában a jelentésalapú technológiák területén alig említésre méltó a magyar részvétel (egyedül a SZTAKI vesz részt egy ilyen projektben), bár 2000 óta folyik Magyarországon kutatás a szemantikai integráció és az ontológiakezelés határterületein. Az IQSOFT egyik korábbi, EU-finanszírozású projektjének eredményeire alapozva (SILK – Semantic Integration via Logic & Knowledge) az IQSYS vezetett K+F projektet (LOBO – LOGic-Based Ontologies) az IKTA támogatásával, majd nyert pályázatot a Nemzeti Kutatási-Fejlesztési (ma Jedlik Ányos) Program keretében (SINTAGMA – Szemantikai INtegrációs Technológia Alkalmazása GRID-alapú, Modell-vezérelt Architektúrákban). A SINTAGMA egy olyan kísérleti *technológiát* fog létrehozni, amely az *információintegrációt szemantikai szinten – azaz modellek,* illetve ontológiák segítségével, jelentésgazdag módon – képes megvalósítani, és amely a projekt során kísérleti *alkalmazásokban* felhasználásra kerül *GRID-jellegű architektúrába* szervezett hálózati elemeken keresztül.

A szemantikai területhez szorosan kapcsolódik a mély *szöveg-feldolgozási technológiák* létrehozása és használata. Ezeknek a technológiáknak ugyanis a szókészletek kezelése (lexikon) mellett a morfológiát, a szintaktikát valamint az általános és a szövegekörnyezetre specifikus szemantikát is fel kell tudni használniuk.

A magyar nyelvvel kapcsolatos számítógépes nyelvészeti kutatások mind akadémiai jelleggel (teljes grammatikai leírás, mintaszöveg-állomány), mind gyakorlati alkalmazásokban folynak (pl. nemzetközileg is terjesztett helyesíráellenőrzők és teljesnek tekinthető angol-magyar fordítóprogram). A legfontosabb fejlesztéseket a Morphologic Számítástechnikai Kft., a Scriptum Informatika Kft., a Dativus Translator Kft. és az Alkalmazott Logikai Laboratórium végzi, míg az akadémiai kutatások terén az MTA Nyelvtudományi Intézet mellett ma a Szegedi Tudományegyetem tevékenysége a meghatározó.⁷ Ezeket egészítik ki a BME (Szószablya-algoritmus) és a SZTAKI (kétértelműségek feloldási módszerei) fejlesztései.

Fontos megemlíteni a szintén az NKFP keretében megvalósuló *Magyar Egységes Ontológia* (MEO) projektet is, amely az előzők értelmében a magyar nyelv általános szemantikáját leíró (WordNet-hez hasonló) ontológiáját tervezi létrehozni. Érdekes további kutatási irány lehet majd az általános (pl. MEO által létrehozandó) és a szöveggörnyezet-specifikus (pl. SINTAGMA által kezelendő) ontológiák együttes alkalmazása a szövegértelmezésben.

9. Következtetések

A METAADATOK sokkal egyszerűbbé teszik a számítógépek számára, hogy az adatokat automatikusan feldolgozzák, és ez a képesség, ami egyéb előnyökkel is járhat: együttműködő képesség, költségcsökkenés, jobb adatminőség, átláthatóság, jobb döntéstámogatás és új üzleti lehetőségek.

A metadatok hatékony és sokoldalú kezelésére épülő jelentésalapú technológiák túlléptek azon a ponton, hogy csak figyelmet keltsenek. A vállalatoknak általában célszerű aktívan figyelemmel követniük az új kodifikációs szabványok kialakulását, és – ha lehetséges – elő is kell segíteniük az ilyen szabványok megjelenését a saját területükön. Magának a témakörnek a megértése és kezdeti tapasztalatok szerzése ezen a területen egyre inkább kritikussá válik.

Különös figyelmet kell fordítani az ember információ-feldolgozó tevékenységének megváltozására. Intelligens, automatikus rendszerek növekvő mértékű használatával együtt szinte észrevétlenül lemondunk napi döntéseink egy részéről, vagy legalábbis döntéseink egyre közvetettebbekké válnak. A szemantikus technológiák egyrészt lehetővé teszik e folyamatot, másrészt segíthetnek az automatikusan – emberi beavatkozás nélkül – hozott döntések ellenőrzésében és felügyeletében.

Van lehetőség, és érdemes bekapcsolni a hazai fejlesztéseket az európai vérkeringésbe. Továbbá célszerű lenne megvizsgálni, hogy a sikeres hazai K+F projektek eredményeinek egymást erősítő hatásait milyen finanszírozási keretek között lehet elősegíteni: pl. kontextusfüggetlen (általános) ill. kontextusfüggő (speciális, szakterületi) ontológiák együttes használata a nyelvi feldolgozásban ill. a jelentésalapú technológiákban.

⁷ Itt hozták létre és tartják karban a magyar nyelv legnagyobb mintaszöveg-állományát, az ún. „Szeged corpus”-t.