

Jelentésalapú technológiák

Krauth Péter

Az informatikai rendszerek a következő évtizedben olyan ún. szemantikus technológiák üzemszerű alkalmazásával egészülnek ki, amelyek az emberi elme jelentésadó és -kezelő mechanizmusaira emlékeztetnek, és áttörést fognak hozni e rendszerek lekérdezhetősége, karbantarthatósága és együttműködőképessége területén.

1. Témakör

A *kognitív informatika* (Cognitive Informatics, CI), a 2000-es években megjelent kutatási irány az informatikán belül, amely úgy tűnik, hogy felhagy a mesterséges intelligencia (Artificial Intelligence, AI) fél évszázados ábrándjának – már-már „Szent Grál”-jának – minden áron való hajszolásával. Helyette igyekszik módszeresen és interdiszciplináris¹ eszközökkel feltérképezni az emberi agy és az emberi elme szerkezetét, működését és belső információfeldolgozó mechanizmusait, és ezek minél mélyebb megértésére alapozva igyekszik javítani az informatikai rendszerek – általános értelemben vett – teljesítményén.

A *tudattartalmakat* az emberi elme nem úgy kezeli, mint egy (képzeletbeli) raktáruháza a polcrendszerén lévő árukat, hanem úgy, mintha bizonyos jelek hatására folyamatosan és automatikusan – időnként kifejezett tudati nyomás, koncentráció eredményeként – előállítaná, megkonstruálná, felszínre hozná azokat. Ennek alapján az emberi elme működése sem raktárhoz hasonlítható, ahonnan jól meghatározott dolgokat kívánságra készen elő lehet venni, hanem inkább egy *vevő-orientált információgyárhoz*, amely a kívánt dolgokat a raktárban lévő „árukból” (ha nincsenek még készen) automatikusan megkonstruálja, és – ha szükséges – a konstrukciós terv alapján legyártja, e terveket későbbi használatra megőrzi, és a raktárt, a gyártósorokat a feltételezett igények szerint folyamatosan átrendezi, átalakítja.

Az ennek mintájára létrejövő, ún. *kognitív rendszerek* célja az ember kognitív² képességeinek *nem beültetéssel* történő növelése és kiterjesztése. Az ilyen rendszerek tervezésekor elsősorban az ember érzékelési, észlelési és gondolkodási folyamatainak, az azokat meghatározó mechanizmusoknak pszichológiailag megbízható számítógépes reprezentációit veszik alapul. Kiindulási pontjuk, hogy az agy nem egyszerűen bizonyos bemeneteket átalakító, „számító” gép, hanem „világmodell” készítő, sokszintű memóriával rendelkező, rendkívül komplex információrendszer.

E rendszerek egyrészt kép-, arc-, gesztus- és általában *mintafelismerésre* képesek, másrészt *szabálykövető, asszociációs és tanulási képességeik* eredményeként jutnak el az észlelt információk feldolgozásában olyan (már szemantikainak³ nevezhető) szintekig, ahol *jelentéstartalmak* kezeléséről lehet beszélni, és állítanak elő ennek alapján újabb információt, hoznak, illetve javasolnak döntéseket akár gyorsan változó környezetben is.

¹ Például számítástudomány, kognitív tudomány, neuropszichológia, rendszerelmélet, kibernetika, szoftverfejlesztés és ismeretfejlesztés.

² A latin „cognitio” ismeretet, felfogást, megítélést jelent. Ilyen módon a „kognitív” megismerőt, megismerésre vonatkozót jelent.

³ A „szemantika” a *jelentés*tan görög eredetű elnevezése, amely eredetileg a nyelvtudománynak a nyelvi formák jelentésével és a jelentések változásával foglalkozó ága volt. Ma már kiterjed nemcsak a nyelvi formák, hanem más jellegű dolgok, jelenségek vizsgálatára is, például képek, zenei művek, filmek.

Egy információ *jelentése*⁴ mindig függ az adott *környezettől*, és valamilyen módon *összesíti* – ebben a környezetben – azokat a már ismert információkat, amelyek a szóban forgó információra vonatkoznak, azzal kapcsolatban állnak. Ahogy az informatika egyre komplexebb és absztraktabb dolgokkal foglalkozik, ahogy a rendszerek egyre bonyolultabb interakciókat folytatnak az emberekkel, és ahogy az emberek egyre inkább elvárják, hogy a számítógép „megértse” őket, úgy válik egyre fontosabbá az *információk jelentésének kezelése* és jelentésük szerkezetének ismerete. A *jelentésalapú technológiák* ezt biztosítják a kognitív (vagy ilyen jellegű) rendszerek számára.

A jelentésalapú (szemantikus) technológiák ilyen módon szoros kapcsolatban állnak a *meta- és háttér adatok kezelésével*. Fontos feladatuk, hogy az információ automatikus feldolgozásába minél hatékonyabban bevonják azokat az *általános asszociációkat és háttér ismereteket*, amelyeket az emberek fejében egy adott kontextusban egy-egy információ tipikusan kivált, és hogy maguk is képesek legyenek ilyen asszociációk, ismeretek létrehozására és kezelésére.

Ennek eredményeképp a szemantikus technológiák a szokásos adatbázistechnológiákhoz képest sokkal rugalmasabban kezelik a két alapvető adatműveletet az olvasást és az írást: 1) egymástól távolinak tűnő információkat *automatikusan kapcsolnak össze*, és származtatnak belőlük további információkat az információkérések minél teljesebb megválaszolásához, 2) új típusú információ hozzáadásakor *szükségtelemmé teszik az újratervezést* a rendszerben.

2. Jelenlegi helyzet

A jelentésalapú technológiák alkalmazása ma két fő kontextusban merül fel:

- a web szélesebb kontextusa (az ún. szemantikus web, ahogy azt Tim Berners-Lee, a W3C meghatározó személyisége, megálmodta, ld. alább); és
- a szervezetek információellátási tevékenységének szűkebb, de nem kevésbé fontos kontextusa.

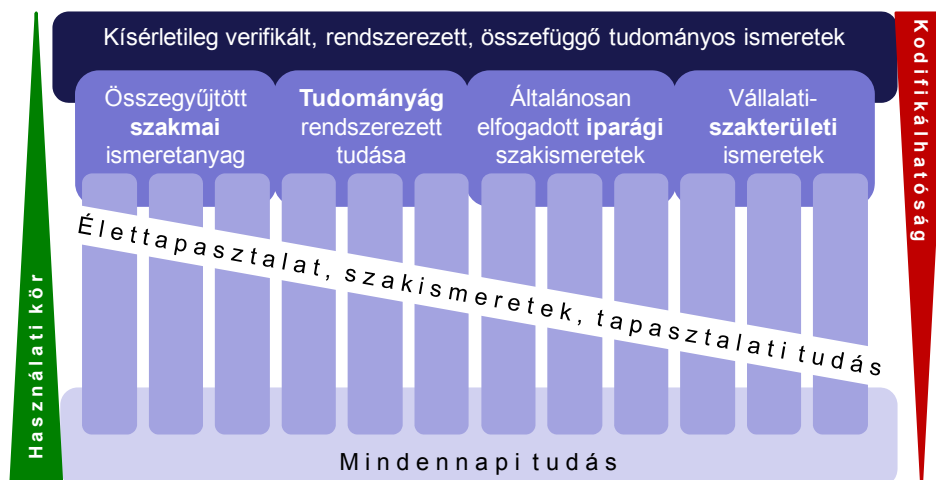
Az internet megjelenése, a technológiai fejlődés a hardverek területén nagyságrendekkel megnövelték az elérhető információk körét és változatosságát, azonban a létrejött adatok kellő hatékony kezelése még meghaladja a jelenleg tipikusan használt, hagyományosnak mondható technológiák lehetőségeit: az adatok még mindig túlságosan nyers állapotban vannak, emellett gyakran hozzáférhetetlenek, ellentmondásosak, széttagoltak, és kihasználtságuk rendkívül alacsony.

⁴ A jelentés az a gondolati, érzelmi, akarati (de mindenképpen: tudati) tartalom, amelyet valamely szó, nyelvi vagy egyéb jel *a tudatban* – esetleg sok áttételen keresztül, de – végsősoron *felidéz*. A tapasztalat szerint a szoros kapcsolatban, ugyanabban a közösségben vagy kultúrkörben élő emberek a szavakat vagy más jeleket *hasonló módon* értelmezik, azaz tudatukban nagyon hasonló tartalmakat idéznek fel. Ez a közös értelmezés adja az alapját e gondolati tartalmak *egyezményes megfogalmazásának*, és így például szótárakban, lexikonokban való rögzíthetőségének. Ettől függetlenül azonban természetesen egyéni szinten a jelentések között az egyének sajátos tulajdonságai, élményei, egyedi élettörténései stb. következtében azért továbbra is vannak – időnként lényegessé váló – különbségek. E különbségek elősegítik a gondolkodás változatosságát és megújulásának lehetőségét, de ezek adják a táptalajt az emberek közötti félreértéseknek, meg nem értésnek is. Túlhangsúlyozásuk jól hangzó, de félrevezető megállapításokhoz is vezethet, mint például „a szónak nincs jelentése, csak használata”. A szónak van jelentése (sőt: több is), de konkrét használata is van, amelyre – azaz arra a konkrét környezetre, ahol elhangzik vagy előfordul – figyelni kell, mert ez lényegesen módosíthatja az egyezményes alapjelentést, illetőleg hogy melyik jelentés érvényes az adott helyzetben. *A szó használati mintázatai rajzolják ki jelentésének szerkezetét.*

Emellett a vállalatok és intézmények egyre növekvő adatvagyonának mérete és bonyolultsága állandó kihívást jelent informatikai szervezeteik számára. Az elmúlt években például a nagyobb nemzetközi pénzügyi szervezetek már petabyte-nyi (mega-gigabyte-nyi) szervezeti tároló kapacitásokról számoltak be. Nyilvánvaló, hogy sokkal nagyobb mértékű és kiterjedtebb *automatizáltság* szükséges az ilyen „adathegyek” kezeléséhez. Az internet továbbfejlődéséhez is elengedhetetlen, hogy az adatokat tömegesen és automatikusan lehessen kezelni.

Paradox módon azonban a web kezdeti, emblemikus „kattintási” (közvetlen adatelérési) filozófiája is akadályozza az információhoz való minél sokrétűbb és hatékonyabb hozzáférést. A weblapok pókhálószerű, rendszertelen szövevényével szemben a ma keresőmotorjai lehetetlen helyzetben vannak. A keresőgépek gyártói, a statisztikai módszereket használó vállalkozások és az üzletiintelligencia-rendszerek készítői rendszeresen „átrostálják” 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. A keresendő tartalmak ugyanis nincsenek kellőképpen „felruházva” leíró, *értelmező információkkal*: nincsenek megfelelő *fogódzók* a kereséshez. Nagy adattömegeket ugyanis csak akkor – és különösen akkor – lehet hatékonyan kezelni, ha pontos képünk (információnk) van a kezelendő adatokról, és nemcsak az embereknek, hanem – ha áttételesen is – rajtuk keresztül a gépi rendszereknek is: jobban meg kell szervezni az adatokat, és ki kell bővíteni azokat *metaadatokkal* (adatléíró adatok). A metaadatok sokkal egyszerűbbé teszik a számítógépek számára az adatok automatikus feldolgozását és értelmezését, pontosabban: éppen a metaadatok teszik ezt egyáltalán lehetővé. A metaadatok összegyűjtése és rendszerezése azonban nem egyszerű feladat.

A jelentésalapú technológiák éppen itt, az adatokról szóló információk kezelésének, azaz a *metaadat-kezelés* lehetőségeinek kiterjesztése és teljesítményének növelése terén segíthetnek: a számítógép kizárólag az ilyen metaadatokon keresztül „tud” bármit is a keresendő információkról, és „tudja” megszervezni ezek alapján az információkeresést. Ennek jegyében indították útjára az új évezred elején a *szemantikus web* kezdeményezést is: lássuk el fokozatosan a webtartalmakat olyan információkkal, amelyek megadják azok szerkezetét, belső összefüggéseiket, szabályaikat és külső kapcsolataikat, azaz lássuk el őket *metainformációkkal*, és ezzel adjunk jelentést az információknak mindenütt a weben. A kezdeményezés eredményeképp már megjelentek azok az alapszabványok és -technológiák, amelyek az ilyen leírások egységességét biztosítani tudják (XML, RDF, OWL).



1. ábra: Az emberi tudás rétegei

A metainformációk tudást képviselnek, de az emberi tudás rendkívül sok részből és rétegből tevődik össze (1. ábra). A jelentésalapú technológiákkal nemcsak a lehetőségeket kell megteremteni a tudás szabványos kodifikálására⁵, hanem be is kell tudni mutatni, hogy az emberek *mindennapi tudását*, illetőleg a *szakterületi, szakmai ismereteik* kodifikálását konkrétan meg lehet valósítani. A legismertebb ilyen rendszer, amely ehhez a célhoz a legközelebb áll a több mint harminc év kutatás-fejlesztés eredményét összegző és magában foglaló Cyc rendszer.

2.1 Szemantikus web és formális logika

A szemantikus technológiák egyik legismertebb megnyilvánulása a *szemantikus web*, amelynek alapkonceptióját Tim Berners-Lee-nek tulajdonítják (1999), és amelynek két fő eleme, hogy: egyrészt metaadatokat kell tudni kapcsolni minden erőforráshoz a weben, másrészt következtetni kell tudni ezeknek a metaadatoknak a segítségével. Metaadat lehet az XML (eXtensible Markup Language) leíró nyelvvel megadható egyszerű struktúrája az információnak, bonyolultabb esetekben azonban információmodelleket (Resource Definition Framework, RDF) vagy ún. ontológiákat (Web Ontology Language, OWL) kell használni. E megközelítésnek már ma is jól kimutatható eredménye az a *szabványosodás*, amely a szoftvereszközök közötti információcserében (XML elterjedése) és a feldolgozáshoz szükséges háttérismeretek leírasi módjában (különböző, feladatspecifikus XML-származéknyelvek) megvalósult.

A már említett XML az 1990-es évek végén jelent meg a színen. Ez lehetővé tette, hogy bármilyen adatáramot és adatforrást viszonylag egyszerűen leírjanak, szintaktikai szerkezetbe rendezzenek, ugyanakkor alapvető szemantikai jegyekkel is ellássanak (például az adategységek neve, keletkezésének körülményei). Az XML gyökerei az 1960-as és 1970-es évekre vezethetők vissza, amikor az 1986-ban megjelent ISO 8879 jelű dokumentum-szabvány (*Standard Generalized Markup Language, SGML*) fő elvei megfogalmazódtak. Az XML lényegében ennek leegyszerűsített és „webesített”

⁵ Formális leírás, amely informatikailag feldolgozható és vizsgálható módon képes reprezentálni az emberi tudás valamelyik rétegének egyes részeit, beleértve egyes szakterületek általános, illetőleg egy vállalat vagy intézmény konkrét ismereteinek és információinak valamilyen jól meghatározott körét.

változata. Nagy előnye, hogy nem kell feltétlenül külön (mintegy „kívülről”) az adatszerkezetet, majd azt használva az adatokat megadni, ahogy az az adatbáziskezelés területén szokásos, 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. Az 1980-as években a formális logika különböző irányzatai – például *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 például a W3C (*World Wide Web Consortium*) ún. *erőforrás-leíró keretnyelvének* (Resource Description Framework, RDF) és a *webontológiák nyelvének* (Web Ontology Language, OWL).

Az OWL különösen 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 egyszerű ontológia egy-egy adatbázis leírása (az ún. adatbázisséma) is, amely 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 tudásintenzív területeken (például 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.

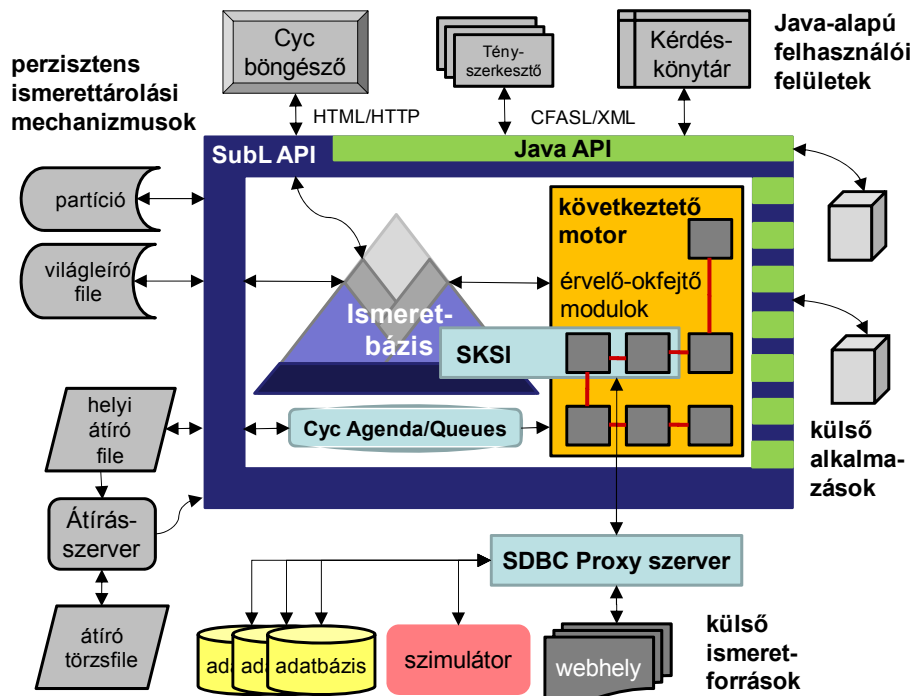
Az OWL több ponton túllép a szokásos sémanyelvek képességein. Az egyik ilyen továbblépési pont, hogy ún. *leíró logikai* nyelvet (Description Logic, DL) is magába foglal, pontosabban arra épül. DL-nyelven 1) sokkal *absztraktabb fogalmakat* is meg lehet ragadni informatikai eszközökkel, mint amilyeneket 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 képességet. A szemantikus weben tehát az információforrások olyan egységes ontológianyelven lehetnek leírva, amely bizonyos szempontból közel áll az emberi elme egyes fontos képességeihez.

2.2 Mindennapi tudás kezelése

A szemantikus technológiák spektrumának másik végén a teljesen egyedi megoldásokat alkalmazó, de a jelenleg létező legmélyebb szemantikai funkciókat nyújtó *Cyc rendszer* helyezkedik el, amely mögött huszonöt év kutatás-fejlesztési tevékenysége áll. Ez a világ legnagyobb, környezetfüggő ismereteket tartalmazó, mindennapi tudást, természetes

nyelvű hozzáférést, többfelhasználós működést és a legkorszerűbb következtető motort biztosító szemantikus rendszere.

Érdemes megismerkedni azokkal a legfontosabb megoldásokkal, amelynek segítségével a Cyc mindezt eléri, és egyben képet alkotni a jelentésalapú technológiák komplex világáról.



2. ábra: A Cyc rendszer architektúrája

A Cyc rendszer (2. ábra) egy ismeretalapú következtető programot⁶ (Cyc) tartalmaz, ehhez kapcsolódó perzisztens⁷ ismerettárolási mechanizmusokat, felhasználói felületeket és alkalmazásprogramozási felületeket (API). A rendszer működésében különösen fontos szerepet kapnak azok a mechanizmusok, amelyek lehetővé teszik az ún. állítások (assertions) formájában kifejezett ismeretek átvitelét a Cyc ismeretbázisába (knowledge base, KB), illetőleg a kivételét onnan. Ez lehetőséget biztosít a kommunikációra a Cyc rendszer komponensei és más programok között.

2.2.1 Az ismeretbázis tartalma

A Cyc középpontjában egy 2,5 millió logikai állítást (tény és szabály) tartalmazó és kb. 200 ezer fogalmat összekapcsoló ismeretbázis áll, amely a mindennapok tárgyaira és eseményeire vonatkozó, széles körű humán konszenzust, valamint számos alkalmazás-specifikus terület (mint például kémia, biológia, betegségek) szakértői ismereteit kodifikálja. Az ismeretbázis emellett nyelvtani és lexikai ismereteket is tartalmaz, amely

⁶ A Common Lisp egy változatában, a SubL-ben lett kifejlesztve.

⁷ Tartós, (az ismereteket) hosszabb távon is megőrző.

természetes nyelvű feldolgozási (szövegelemzési és -generálási) képességeket⁸ épít be a Cyc felhasználói felületébe.

2.2.2 A következtető motor

A Cyc érvelő-okfejtő (reasoning) képességét egy következtető motor (inference engine) biztosítja, amely *több száz* mintavezérelt, heurisztikus modult és rezolúció-alapú tételbizonyítót alkalmaz, hogy új következtetéseket vonjon le (dedukció) vagy új hipotéziseket állítson fel (abdukció) az ismeretbázisban tárolt állításokból.

A Cyc következtető motorja többszálú működésre képes, azaz több feladaton tud egyszerre dolgozni, mint például kérdés megválaszolása vagy problémamegoldás. A következtetések közbeni eredményeit a belső memória adatszerkezeteiben, az ún. problématarakba menti el. Ez a megoldás lehetővé teszi, hogy a következtető motor újrakezdje a munkát, például ha a felhasználó úgy dönt, hogy a Cyc számára további időt enged, hogy választ találjon – anélkül, hogy korábbi lépéseket meg kellene ismételni. A következtető motor képes mind előre- mind visszakövetkeztető módban működni, és teljes körű magyarázatot tud adni a válaszokra – beleértve azon források nevét (például emberek, kiadványok, webhelyek), ahonnan az információ származik. Még figyelmeztetni is képes a felhasználót azokra az esetekre, amikor bizonyos következtetések mellett és ellen is fel lehet sorakoztatni érveket – a különböző körülmények vagy a kontextusban bekövetkezett változások miatt. A felhasználók több tucat paramétert módosíthatnak, hogy a következtetési algoritmust finomhangolással optimalizálják. A következtető motor egyik komponense, az ún. tapasztalatvezérelt taktikai modul (experience based tactician) képes megtanulni az eredményesnek bizonyult keresési utakat, és ilyen módon a következtetés teljesítménye idővel automatikusan javul.

2.2.3 Az ismeretbázis referenciapéldánya

Az ún. világleíró file-ok (world file) a Cyc rendszer operatív memóriába betöltött és ott karbantartott ismeretbázisáról („világáról”) készített pillanatfelvételeket tartalmaznak. Ez biztosítja a leggyorsabb ismeretmentési és -betöltési lehetőséget.

2.2.4 Az átírási mechanizmus

Ha többen is használják a Cyc-ot, és az ismeretbázisainak különböző példányait gyakran módosítják, akkor ezeket a Cyc szinkronban tartja. Ez úgy történik, hogy mindegyik példány a saját műveleteit (KB-módosításait) ún. átíró file-ok (transcript file) sorozatában tárolja el. Ahogy az egyes műveletek hozzáíródnak a példány helyi átíró file-jához, rendszeres időközönként átkerülnek az átírási-szerverhez (transcript server), amely karbantartja az összes módosítás törzsfile-ját. Ha egy KB-példány úgy van (opcionálisan) konfigurálva, hogy dinamikusan fogadhat módosításokat, akkor rendszeres időközönként rákapcsolódik az átírási-szerverre, hogy az új módosításokat letöltse, és egy erre fenntartott sorba állítva végrehajtja azokat.

A Cyc átírási mechanizmusa jól működik az ismeretbázis különböző példányainak szinkronban tartására – néhány napig vagy hétig. De ahogy a műveletek száma a törzsfile-ban növekszik, úgy igényel egyre több időt, hogy egy újonnan indított Cyc-példány az összes addigi műveletet végrehajtsa. Ezért előbb-utóbb az egyik teljesen

⁸ További információkért vö. *Számítógépes szövegelemzés* elemzés.

naprakész Cyc-példányt, amely tehát a törzsfile-ban lévő összes műveletet feldolgozta, arra kell felhasználni, hogy egy új világleíró file-t hozzon létre. A továbbiakban ez lesz az új referencia-ismeretbázis, amelyet minden Cyc-példány induláskor betölt, és amelyre a későbbi műveleteket alkalmazzák.

2.2.5 A partíciók

Az ismeretbázisok közötti információátvitelre a Cyc a particionálás mechanizmusát használja. Ennek keretében egy konfigurációs file-ban meg lehet adni az átmozgatni kívánt részét (partition) egy ismeretbázisnak. Ennek alapján egy speciális, CFASL-típusú file készíthető el, amelyet aztán be lehet tölteni a cél-ismeretbázisba. Az eljárás gyors, de csak akkor működik igazán jól, ha a forrás és a cél elég hasonló egymáshoz.

2.2.6 Ismeretforrások szemantikus integrációja

A Cyc-technológia lehetővé teszi azt is, hogy az ismeretbázisán kívüli adatbázisokkal és webhelyekkel összekapcsolódjon, és következtetési funkcióit ezekre is kiterjessze. E célból egy speciális CycL nyelvet, következtető modulokat és adatkapcsolati modult (Semantic Data Base Connectivity, SDBC) alkalmaz, amelyek együtt alkotják a Cyc ezen részét (Semantic Knowledge Source Integration, SKSI). A CycL segítségével részletes szemantikus leírást (tulajdonképpen metaadatokat) lehet megadni az olyan külső információforrásokra, mint például adatbázisok és webhelyek. A következtető motor ezeket a metaadatokat használja az SKSI specializált következtető moduljaival együtt arra, hogy a forrásokat fizikailag elérje, és ezek adatain ugyanolyan módon végezzen következtetéseket, mintha azok az ismeretbázis szerves részei lennének. Ilyen módon a Cyc következtető gépe például SQL SELECT kéréseket képes intézni a megfelelő adatbázisokhoz, paramétereiket kérheti le webhelyekről vagy kereshet információkat magában a Cyc ismeretbázisában, majd a különböző forrásokból származó eredményeket összesítheti egy egységes válaszban.

Az SKSI lehetőséget ad arra is, hogy a következtető gép módosítsa a külső ismeretforrásokat, ami igen komplex átalakításokra ad lehetőséget.

2.2.7 OpenCyc

Az OpenCyc a Cyc-technológia nyílt forráskódú változata, amely tartalmazza a Cyc ismeretbázisának nyilvános (azaz nem védelmi célra használt) részét. Ez az ontológia akár OWL-formátumban is elérhető. A következtető motornak és a Cyc-böngészőnek azonban csak a kompilált változata érhető el. A legújabb 1.0.2-es változata már kapcsolatokat is tartalmaz a Cyc fogalmai és a Wordnet szinonimacsoportjai (synsets) között.

Két évvel ezelőtt alapították meg a Cyc Foundation-t a Cyc-technológia terjesztésére. A szervezet a Cyclify projektjében azt a célt tűzte ki, hogy „egy olyan világot teremtsen, amelyben mindenki szabad hozzáférést kap azokhoz a programokhoz, amelyek a teljes kodifikált emberi tudás felhasználásával képesek automatikus következtetésekre”.

3. Folyamatban lévő kutatások, fejlesztések

3.1 Ipari indíttatású fejlesztések

Sok pénzügyi-számviteli folyamat még mindig lényegében manuális: az emberek kimásolnak adatokat egy alkalmazásból, és beteszik egy másikba, vagy – még rosszabb

esetben – ismételten begépelik az adatot különböző alkalmazásokba. Egy nemrégiben kiadott specifikációsorozat (Extensible Business Reporting Language, XBRL) azt ígéri, hogy képes egy vállalat pénzügyi információinak kodifikálására.⁹ 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 elterjed, jelentős mérföldkő lesz a pénzügyi jelentéskészítésben – és ez jelzésértékű lenne sok más területnek is.

Az adatok kiegészítő felcímkézése XBRL-lel vagy más szabványos módon sok más előnnyel is jár; például *transzformációs szabályok* segíthetik az adatok átalakítását különböző olyan formátumokba, amelyek különböző típusú készülékeken való megjelenítésre teszik alkalmassá.

A metaadat-címkék a kereső motorokat és az üzletiintelligencia-szoftvereket is segíteni fogják. 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 például 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-ajánlása szerint. Vállalati kontextusban több olyan korai fejlesztési tevékenység is megfigyelhető, amely során a cégek javítják az információellátási láncot – különösen a vállalati információintegráció (Enterprise Information Integration, EII), a vállalati tartalomkezelés (Enterprise Content Management, ECM) és a javuló keresőgépek terén, és amelyek mindegyike valamilyen formában használ jelentésalapú technológiákat.

Sok esetben egy-egy szervezetnek már nagy mennyiségű, jól használható adata van az adatbázis-rendszereiben, amelyek nagyjából már úgy írják le a kezelt objektumokat, azok tulajdonságait és kapcsolataikat más objektumokhoz, hogy az elég közel áll a logikai formalizmushoz. Ezek formális logikára való leképezése gyakran ETL-eszközökkel (Extraction, Transformation and Loading) könnyen automatizálható.

Más 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 bizonyos kapcsolatok) majdnem teljesen automatikusan kinyerhetők nyelvi elemzés alkalmazásával – ezt a megközelítést információkinyerésnek is nevezik.¹⁰ 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.

⁹ Több mint a szóba jövő informatikai rendszerek adatbázissémáinak összessége, mert olyan fogalmakat is leír, ami egyik sémában sem, vagy nem közvetlenül jelenik meg; viszont tartalmazza e fogalmak leképezéseit is az adatbázisokra. Például 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 mindenféle anomáliákat okozhat a szervezetek pénzügyi jelentéseiben.

¹⁰ További információért ld. *Számítógépes szövegelemzés* elemzés.

Számos szoftvergyártó ezen a területen már több éve jelen van, de továbbra sem rendelkeznek meggyőző referenciapéldákkal. Ez változhat a következő években, amelyek során folytatódni fog az olyan piaci konszolidáció, mint például a 2004 elején az IBM által az izraeli Unicorn-nal létesített partneri kapcsolat, amelynek célja egy integrált eszközkínálat kialakítása az IBM *DB2 Information Integrator*-a köré. Egy másik példa, hogy 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.

3.2 Szemantikus rendszerek európai kezdeményezése

Az EU a kutatás-fejlesztés VI. keretprogramjában (FP6) több mint egy tucat projektet finanszírozott 2004-2008 között a „jelentésalapú ismeretkezelő rendszerek” (semantic-based knowledge systems) stratégiai témakörben.

A hangsúly ezeknél a fejlesztéseknél a szemantikus web *szabványain* és ezek terjesztésén (KNOWLEDGE WEB, KB20), a *szemantikus webszolgáltatásokon* (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), a *újfajta logikai formalizmusokon* (ASPIC) és *logikai következtetések* (REWERSE) volt.

A *Szemantikus Rendszerek Európai Kezdeményezése* (European Semantic Systems Initiative, ESSI) eredetileg három, korábbi (FP6) projekt együttműködésének az eredménye, amelynek célja a szemantikus webszolgáltatásokat és a jelentésalapú technológiával felszerelt rendszereket szemantikus szolgáltatás-orientált architektúrával ötvözze.

Ma ez a kezdeményezés ad keretet az STI (Semantic Technology Institute) International működésének, amelynek küldetése, hogy a szemantikus megoldásokat a *korszerű számítástechnika alappillérvé* tegye. Ma még sok társadalmi, gazdasági és technológiai kihívás és korlát állja annak útját, hogy a szemantikus technológia valós alkalmazási rendszereknek szerves és jól látható építőeleme legyen. Az STI International egyesületi formában működik, és keretet ad az e célt felvállaló, a világ bármely részén tevékenykedő, tudományos, ipari és kormányzati szervezetek együttműködésének. Az STI International vezető szerepet tölt be új üzleti modellek kifejlesztésében, valamint a dolgozók és a vállalatok kapcsolatának javításában. Az STI International jelenlegi legfontosabb projektjeit az alábbi pontok ismertetik.

COIN – Collaboration and INteroperability for networked enterprises

A projekt (<http://www.coin-ip.eu>) célja olyan IKT-megoldások tanulmányozása, tervezése, fejlesztése és prototípusok készítése, amelyek a hálózati kultúrát befogadó vállalatok közti üzleti együttműködést segítik elő. A tervezett rendszerek a vállalatok által felkínált szolgáltatásokból az üzleti kontextustól, a gazdasági szektortól, a szóban forgó vállalatok méretétől, valamint az együttműködés nyíltságától és dinamikájától függően *garantáltan a lehető legjobb kombinációt* fogják előállítani és összeintegrálni. Ennek érdekében képességértségi modelleket, üzleti szabályokat és önszabályozó döntéstámogatási megközelítéseket használnak fel.

LarKC– Large Knowledge Collider

A projekt (<http://www.larkc.eu>) célja egy olyan platform létrehozása, amely igyekszik megszüntetni a jelenleg létező, szemantikus webre készített *következtető gépeknél* fellépő skálázhatósági korlátokat *tömegméretekben elosztott rendszerek* segítségével. Az alapfeltételezés az, hogy egy megfelelő következtetési infrastruktúrának a szemantikus web számára túl kell lépnie a kizárólag a logikára építő jelenlegi megközelítéseken. A (logikai értelemben vett) *következtetés* (információkinyerés értelmében vett) *kereséssel való egyesítése*, és a *korlátozott racionalitás* elvének¹¹ követése szükséges a webméretű következtetési mechanizmusok megvalósításához.

OKKAM

A projekt (<http://fp7.okkam.org>) célja az ún. *létező dolgok webjének* (Web of Entities) a megteremtése, azaz egy globális virtuális tér létrehozása a létező dolgokra vonatkozó információk és ismeretek közzétételére és kezelésére. A létező dolgok webje az *elosztott információforrások zökkenőmentes integrációjának* lesz az eredménye a dolgok globális, egyedi azonosítóinak felhasználásával (és újrafelhasználásával!). Elvben minden dologra ugyanazzal az azonosítóval fognak hivatkozni a weben vagy nagy szervezetek hálózataiban.

Service Finder

A web statikus (nem változó) dokumentumok gyűjteményéből szolgáltatások gyűjteményévé válik. A szolgáltatások közötti, üzleti célú adatcsere megvalósulása érdekében általában a szolgáltatás-orientált architektúrákat és a webszolgáltatásokat tekintik a legígéretesebb megközelítésnek. Ennek következtében tekintélyes figyelem irányul az akadémia és az ipar felől egyaránt a webszolgáltatás- és egyéb kapcsolódó technológiák irányában.

Emellett általánosságban a web használatában is jelentős változás következett be. A webhasználó a hagyományos, passzív, fogyasztói szerepről fokozatosan egy aktív, beavatkozó szerepre vált át: ez az ún. „web 2.0” jelenség.¹² Bár mindkét megközelítést használják az alkalmazásfejlesztésben, mégis a két technológiát eddig elkülönülten használták.

A projekt (<http://www.service-finder.eu>) célja olyan *szolgáltatáslokalizáló platform* kialakítása, ahol a webszolgáltatások beágyazódnak egy web 2.0-ás környezetbe. A

¹¹ Herbert A. Simon and Associates: *Decision Making and Problem Solving*. In: *Report of the Research Briefing Panel on Decision Making and Problem Solving*. National Academy Press, Washington DC, 1986.

¹² Vö. *A web 2.0 (és ami mögötte van)* elemzés.

projekt növelni fogja a szolgáltatásnyújtás hatékonyságát, és kiterjeszti a szolgáltatás-orientált architektúrákat a weben elérhető szolgáltatások millióinak használatára.

SOA4All

A projekt (<http://www.soa4all.org>) célja olyan technológia megvalósítása, amellyel akár emberek milliárdjai is nyújthatnak és használhatnak szolgáltatásokat a legkorszerűbb webtechnológiák segítségével. A kialakítandó rendszer és infrastruktúra négy, egymást kiegészítő technikát egyesít egy koherens, szakterületfüggetlen szolgáltatási platformba:

- a *web* technológiája, mint infrastruktúra a világméretű szolgáltatásintegrációhoz,
- a *web 2.0* a hatékony ember-gép együttműködés érdekében,
- a *szemantikus web* technológiája a szolgáltatásmegtalálás intelligenssé tételéhez,
- a *kontextusmenedzsment*, amely lehetővé teszi a felhasználói igények gépi kezelését és így a meglévő szolgáltatások felhasználói igényekhez való automatikus hozzáillesztéséhez.

SHAPE – Semantically-enabled Heterogeneous service Architecture

A projekt (<http://www.shape-project.eu>) célja olyan vállalati rendszerek kifejlesztése és megvalósítása, amelyek *jelentésalapú, heterogén szolgáltatásarchitektúrára* (Semantically-enabled Heterogeneous service architecture, SHA) épülnek. Az SHA egy egységes megközelítésben terjeszti ki a szolgáltatás-orientált architektúrát szemantikus technikákkal és heterogén infrastruktúrákkal (webszolgáltatások, ágensek, szemantikus webszolgáltatások, p2p és grid). A projekt *modell-vezérelt fejlesztési* (Model-Driven Engineering, MDE) módszertant és ehhez eszköztámogatást alakít ki, és részt vesz az SHA metamodelljeinek és nyelveinek szabványosításában.

Service Web 3.0

A számítástechnika négy évtizednyi gyors fejlődés után is minden szinten forradalmi változásokat él át – beleértve a hardvert, a köztesszoftvert, a hálózati infrastruktúrát, de még inkább az intelligens alkalmazások terén. A szemantikus web és a webszolgáltatások az internetet információhálózatából *tudás- és szolgáltatáshálózat*tá alakítják át. Az interneten felkínált szolgáltatások számossága drámai mértékben fog növekedni a következő néhány évben.

A projekt (<http://www.serviceweb30.eu>) célja ennek az átalakulásnak a támogatása mind a közösségi kutatások, mind a technológiai fejlesztés oldalán.

ACTIVE

Európa a tudásgazdaságra való átmenet időszakát éli. Az eredményes tudásgazdaság ma már alapvető fontosságú a sikeres gazdasági tevékenységhez. Ennek ellenére a vállalatok csak nagy nehézségek árán tudják – ha tudják – meglévő, lényeges ismereteiket átadható, könnyen elérhető és bevethető ismeretkezelő eszközökké átalakítani. A szellemi dolgozók (knowledge worker) fejében lévő tudás rejtett módon van jelen a vállalatban – vagy azért, mert nincs jól *megfogalmazva* (tacit knowledge), vagy azért, mert közvetlenül ugyan (elvben) elérhető, de nincs kellően *széles körben megosztva* (például e-mailek, személyi PC-s mappák).

A projekt (<http://www.active-project.eu>) célja a szellemi dolgozók termelékenységének növelése proaktív, kontextusalapú, ugyanakkor egyszerű és észrevétlen módon.

Különböző területek fogalmainak és módszereinek ötvözésére törekszik innovatív alkalmazási rendszerekben: 1) közösségi szoftverek és web 2.0, 2) jelentésalapú technológiák, 3) kontextusmodellezés és -elemzés, kontextusérzékeny feladatkezelés, 4) ismeretkezelési folyamat modellezése, elemzése és proaktív támogatása.

A projekt olyan áttörést igyekszik elérni, amellyel például a tanácsadói, távközlési és mérnöki területeken a vállalatok az ismerettechnológiát hatásos eszközzé tudják tenni a vállalati tudás mainál sokkal nagyobb mértékű megosztására.

3.3 Open Mind Common Sense projekt

Sok kutató úgy véli, egyszerűen még elgondolni is túl *nagy* feladat, hogy a számítógépeket megtanítsák a mindennapi tudásra (tulajdonképpen a józan ész használatára). Mások azonban megpróbálták megbecsülni a mindennapi tudás *nagyságát*, és úgy találták, hogy *több százmillió ismeretelemből* állhat az a mindennapi tudás, amit az emberek használnak.

Az eddigi legnagyobb kísérlet a mindennapi tudás kodifikálására a Cyc rendszer, amelynek ismeretbázisa szakértők egy csoportjának több évtizedes munkájával, több tízmillió dollár elköltése után kb. két-hárommillió ismeretelemet (állítást) tartalmaz.

Maga Doug Lenat a Cyc-projekt vezetője is elismeri, hogy egy vagy inkább *két nagyságrenddel* még elmaradnak attól, amire szükség lenne. Mások azonban úgy vélik, hogy nem is tud ekkorra feladatot *egy szűk csoport* elvégezni, és ezért teljesen más megközelítést kell alkalmazni.

Az Open Mind Common Sense (OMCS) webhelyet

(<http://www.openmind.org/commonsense>) 2000-ben azon célból hozták létre, hogy a köznapi emberek teljes közösségét próbálják rávenni arra, hogy mindennapi tudásukat szervezett, számítógépes feldolgozásra alkalmas formában megfogalmazzák. Hét év alatt hétszázezer ismeretelemet (állítást) gyűjtöttek össze több mint tizennégyezer regisztrált, angol nyelvű közreműködő segítségével. Az összegyűjtött ismeretelemből százötvenezer fogalmat tartalmazó ismeretbázist, fogalmi hálót (ConceptNet) szerkesztettek össze. Ez azt jelenti, hogy a fogalmak szintjén már kezdi közelíteni a Cyc méretét, de az ismeretelemek tekintetében csak a negyede annak, amennyit a Cyc tartalmaz, de töredékidő alatt és töredékköltség mellett.

Az OMCS második generációs mindennapi tudást tartalmazó ismeretbázis, amely sokat tanult a Cyc megközelítéséből, de mégis máshogy közelíti meg a problémát.

Elgondolkodtató ugyanis, hogy miért nincs négy vagy öt Cyc méretű projekt ezen a területen?

Az OMCS keretében a résztvevők egyszerűen (vagy szabad szerkezetű, vagy strukturált) angol mondatokban fogalmazzák meg az ismereteiket. A Cyc ezzel szemben a CycL segítségével, egy precíz, de nehezen használható nyelven megfogalmazva várja az ismereteket.

Az egyik legfontosabb ok amiatt az OMCS vezetői úgy gondolják, hogy az angol jobb erre a célra az az, hogy ma már elég fejlett a természetes (angol) nyelvű szövegfeldolgozás ahhoz, hogy helyes fogalmakat és állításokat lehessen kinyerni a résztvevők által megadott mondatokból. Emellett a felhasználói felület legújabb változata (Open Mind Commons) lehetővé teszi a résztvevőknek, hogy a már megadott állításokat finomítsák – értékelve, hogy igaz, hasznos állítások-e vagy sem.

A mögöttes fogalmi háló, a ConceptNet legújabb, harmadik változata ma már biztosítja a többnyelvűséget is. Ez jelenleg az angol és a portugál nyelv együttes használatát jelenti a gyakorlatban. A ConceptNet 3 az OpenCyc-hoz hasonlóan össze van kapcsolva a WordNet-tel.

Emellett a ConceptNet új architektúrája alkalmasabb arra, hogy 1) fokozatosan legyen módosítva, 2) különböző forrásokból legyen feltöltve, és 3) olyan komplex kérdéseket lehessen megválaszolni, amelyek például az analógiák feltárásához szükségesek.

Az OMCS bizonyos értelemben „web 2.0” jelenség, hiszen a részvételre, közreműködésre épít, és hasonlít a Wikipédia megközelítésére is, mivel a kollektív tudás összegyűjtését eredményezi teljesen önkéntes és szabad hozzájárulással. Nagy különbség persze, hogy amíg a Wikipédiában összegyűjtött tudás csak az emberek számára használható, addig a ConceptNet tartalma további gépi feldolgozásra is alkalmas.

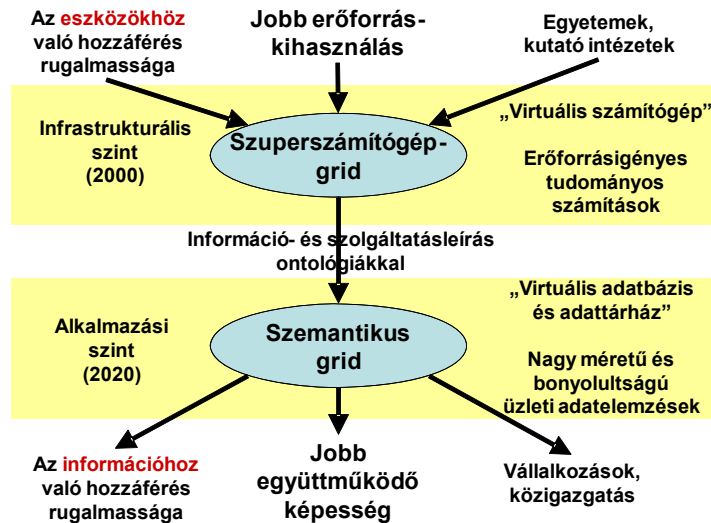
4. A várható fejlődés



3. ábra: Várható fejlődés (2008-2018)

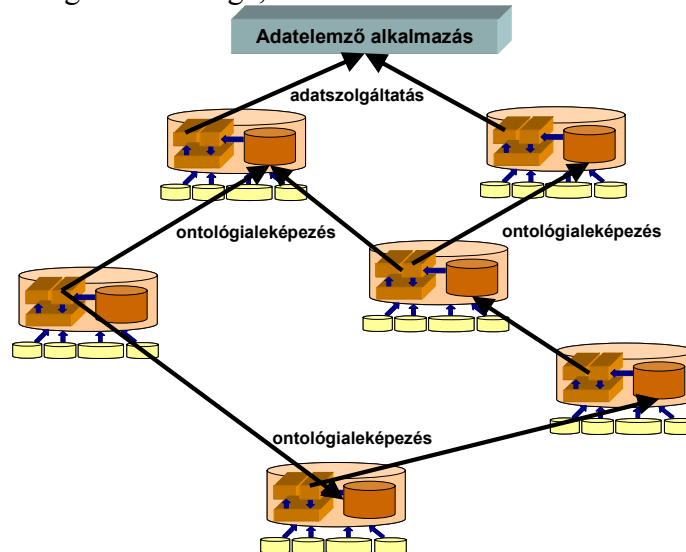
A szemantikus technológiák várhatóan már a következő évtizedben jelentős hatással lesznek a tartalomelőállításra. Ezen túlmenően azonban ezeket a technológiákat nemcsak az információkeresésben és adatelemzésben, hanem általánosabban *informatikai rendszerek együttműködésének javításához* is fel fogják használni (semantic interoperability), sőt az ember-gép kapcsolat magasabb szintjeinek eléréséhez is. A *szemantikus technológiák* a jövőben mélyen be fognak épülni az információfeldolgozó eszközökbe és folyamatokba. Várható, hogy a következő évtizedben az XML-alapú szabványosodás minden lényeges informatikai és alkalmazási területre kiterjed, és az évtized második felére stabil alapját képezi a szemantikus technológiák rutinszerű, mindennapi alkalmazásának.

Az évtized közepére a formális logika válik (elsősorban a fent említett megnyilvánulási formáiban: RDF, OWL, DL) a legjelentősebb metaadatkezelési kezdeményezések többségének közös alapjává.



4. ábra: A szemantikus grid várható kialakulása

A *szemantikus web* kontextusában várható, hogy a leglátványosabb, rövid távú alkalmazási sikerek az elektronikus kereskedelemben jelentkeznek. Sok időnek kell azonban még eltelnie, amíg a leszűródött tapasztalatok világos útmutatást tudnak adni, hogy milyen esetekben igazolható a szemantikus webhez kapcsolódó – és általában a jelentésalapú technológiák – költsége, és mikor nem.



5. ábra: Az ontológiák szerepe az alkalmazások adatokkal való ellátásában

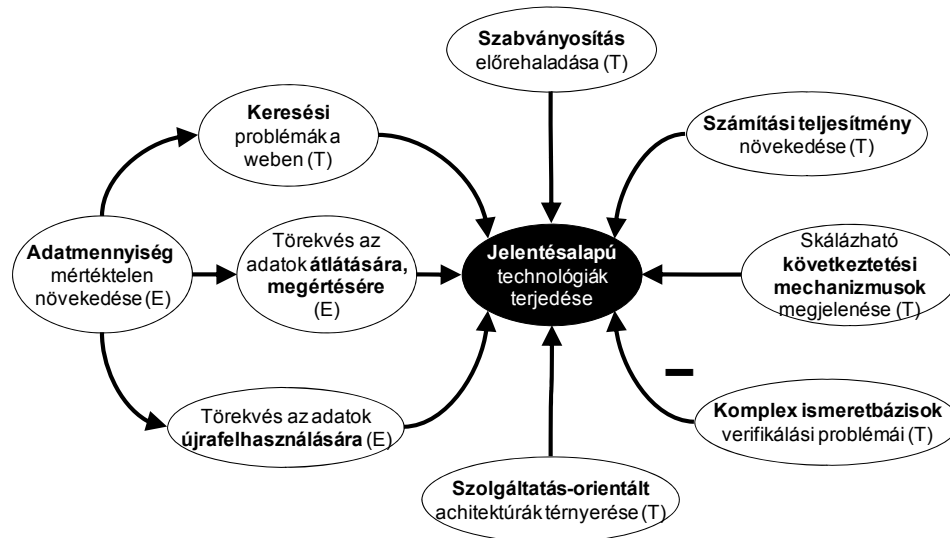
A *mindennapi tudást tartalmazó ismeretbázisok* használata a következő évtized végére igen gyakori lesz, és alkalmazásuk a vállalatirányítási rendszerektől kezdve a tudásmenedzsmenten keresztül a robotokig terjedően széles körű lesz.

Itt célszerű megemlíteni azt a tendenciát is, hogy a következő évtizedben a *gridtechnológia* 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 (ld. 4. ábra).

E jövőbeni gridtechnológia alapegységei a hálózat olyan képességekkel felruházott csomópontjai lesznek, amelyek lényeges vonása lesz, hogy a hálózat különböző pontjain lévő információforrásokat ún. *ontológiákkal* írják le, képezik le egymásra, azaz kapcsolják össze (ld. 5. ábra).

5. Befolyásoló tényezők



6. ábra: Befolyásoló tényezők

5.1 Gazdaság és társadalom

Az információs társadalom és gazdaság egyik kísérő jelensége, hogy *minden határon túl megnövekszik* a gépi úton – legalábbis elvben – kezelhető információ mennyisége. Egyre inkább „luxusnak”, felesleges pazarlásnak tűnnek a csak papírra vetetett információk és gondolatok, a (digitálisan) nem rögzített beszéd, beszélgetés. Egyre több embernél van mindig kéznél valamilyen információkezelő eszköz noteszgép, PDA, „okos” mobiltelefon stb. formájában, terjed a napi eseményeinek blogszerű rögzítése, kommentálása és esetleges közzététele a weben (a web 2.0 jelenség részeként).

Ez azonban csak az első lépés, mert ahogy növekszik a digitálisan rögzített adatok mennyisége, úgy válik igénnyé ezen adatok érdemi *újrafelhasználása*, amihez pedig elengedhetetlen annak *megértése*, hogy ezen adatok egy adott környezetben mit is mondanak, mit is jelentenek. Például az ügyfélkapcsolati rendszerekkel rögzített beszélgetések („Tájékoztatjuk, hogy az ügyintézőnkkel folytatott beszélgetés rögzítésre kerül.”) számos olyan ügyfélkérést, kiegészítő ügyfeladatot tartalmaznak, amelyek felhasználása már csak az ügyfelek minél jobb kiszolgálása miatt is célszerűnek mutatkozik. Mindez jelentős keresletet képez a jelentésalapú technológiák használatára.

5.2 Technológia

- A *metaadat-szabványok* elterjedése közmegegyezést igényel, és ez önmagában hosszadalmas, lassú folyamat. A szemantikus technológiák megfelelő együttműködéséhez azonban ennél többre van szükség: fontos, hogy ezek a szabványok a gyakorlatban is jól használható, könnyen beépíthető eszközökben testesüljenek meg.
- A metaadatkezelés eddig is igen nagy kihívást jelentett a *számítógépek teljesítményére*, és nem lehet ez másképp a jövőben sem. A szemantikus információkezelés a hagyományoshoz képest nagy mennyiségű többletfeladat elvégzését igényli, és ehhez többletkapacitásokra van szükség. Szerencsére egyelőre nem mutatkozik törés a számítási, tárolási és átviteli kapacitások eddigi ütemben történő növekedésében.

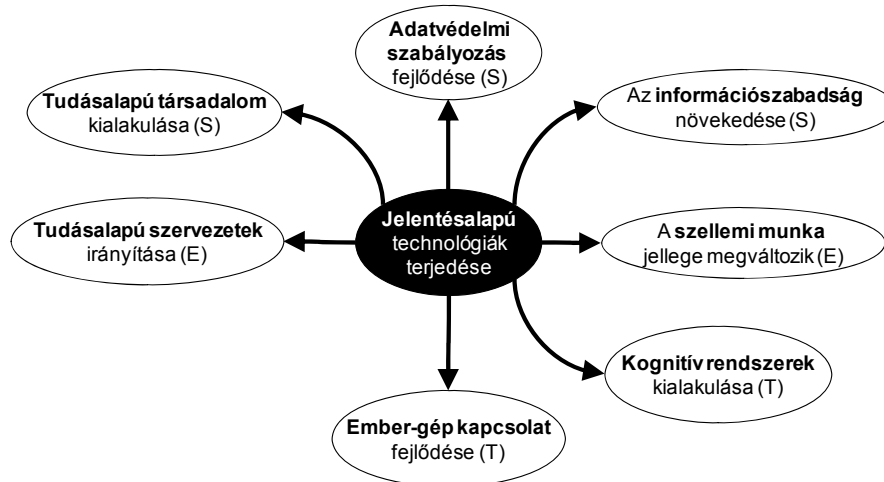
A multimédia-objektumok, majd a mobil, illetve környezeti (ambient) eszközök megjelenése és elterjedése elsősorban mennyiségi növekedést indukált. Az emberi befogadó képesség és a rendelkezésre álló idő végelessége azonban természetes határt szab a meghallgatható zene, a megnézhető videók és a lebonyolítható telefonbeszélgetések mennyisége tekintetében. További számítógépi kapacitásnövekedést az információfeldolgozás elmélyülése, bonyolultságának növekedése hozhat csak, és ebben a jelentésalapú technológiáknak kiemelkedő szerepe lehet.

- A szemantikus web alkalmazásaihoz szükséges gyakorlati tapasztalatok azonban még csak most kezdenek gyűlni. Még nem tipikus, hogy a vállalatok rendelkezzenek a szemantikusweb-szabványok kellő szintű ismeretével, és általában a *metaadatkezeléssel* kapcsolatos tevékenységek elismertségének növekedni kell.
- Egyes területeket közismerten nehéz strukturálni és formalizálni (például szervezetek együttműködése, politika és jog) az ismereteket. Ez nem fog a közeljövőben sem megváltozni. Más területeken azonban jelentős előrelépések történtek – számítógéppel is kezelhető – terminológiakialakításban és fogalomalkotásban (például élettudományok, gyógyászat és számvitel). Az *ontológiaépítés és -kezelés eszközeinek* be kell vonulniuk az informatika mindennapi gyakorlatába. Ez annál is inkább szükségszerű, mert a szemantikus technológiák által kezelt, esetenként igen *komplex ismeretbázisok verifikálása* (tesztelése és ellenőrzése) komoly nehézségekbe ütközhet: minél összetettebb és részletesebb az ismeretbázis annál több időt igényel és annál több felhasználót kell bevonni a munkába.
- Lényeges szerepet játszhat a szemantikus technológiák terjedésében a *weben való keresés egyre növekvő problémái*. Elvitathatatlan, hogy az új évezredben megjelent, webes keresőrendszerek átformálták a webről alkotott képet. Már nem hiperlinkek bejárásával a web mélyrétegeibe belebújva lehetett csak információt előbányászni, hanem az információigények keresőkérdésekben való megfogalmazásával „fekete dobozként” lehetett kezelni a webet. Azonban a jelenlegi technológia nem ad kellően fókuszált válaszokat, nem tudja figyelembe venni a körülményeket, a kérdező személyét stb., ezért a szemantikus technikákkal elérhető célzott kérdések, és az azokra adott pontos válaszok újabb lökést adhatnak a web produktív használatának.
- Ehhez azonban szükség van logikai következtetési mechanizmusokra, de ezek alkalmazása – nagy mennyiségű adat esetén – lehet, hogy nem ad elfogadható

eredményt (például a számítások túlságosan lassúak lesznek) még akkor sem, ha ezt a szokásosnál könnyebben kezelhető, és gépi számításokra alkalmasabb logikák – például a leíró logikák (description logics) – támogatják. Szükség van tehát olyan *következtetési mechanizmusokra*, amelyek az adott probléma jellegéhez illeszthetők és választhatók, valamint a feladat nagyságához *méretezhetőek*.

Végül fontos felhívni a figyelmet arra, hogy jelentős átfedés van – és így kölcsönös egymásra hatás várható – a szemantikus web és a webszolgáltatások közös területén a *szolgáltatás-orientált architektúrákkal*. Ennek eredményeként például egy-egy termék vagy projekt esetén időnként nehezen lehet eldönteni, hogy melyik technológiai kategóriába tartozik inkább. Annyi mégis kijelenthető, hogy a jelentésalapú technológiák elsősorban az egyes rendszerkomponensek egymás közötti, illetve a felhasználóval folytatott *kommunikációjának* minél magasabb (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.

6. Várható hatások



7. ábra: A jelentésalapú technológiák terjedésének várható hatásai

6.1 Technológia

6.2.1 Az ember-gép kapcsolat fejlődése

A jelentésalapú technológiák nagy hatással lehetnek az *ember-gép kapcsolat* fejlődésére. A jelentésalapú technológiák lényege, hogy tudásreprezentációs és következtetési képességekkel ruhazza 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 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 (például 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.).

6.2.2 Kognitív rendszerek kialakulása

Az emberi elme egyik alapjellegzetessége az asszociációs képesség. Nem nagyon tekinthető egy informatikai rendszer *kognitív rendszernek*, ha nem rendelkezik ilyen mechanizmusokkal. Szerencsére a jelentésalapú technológiák tulajdonképpen ennek teremtik meg a lehetőségét.

A szemantikus technológiák felhasználásával a legkülönbözőbb információs objektumok (programok, adatbázisok, ismeretbázisok, webhelyek, ágensek, szolgáltatások stb.) közötti kapcsolatok (asszociációk) könnyen és a mindenkori helyzet igényeitől függően létesülhetnek.

Természetesen a kognitív rendszerek működéséhez sok minden másra is szükség van, például a multimodális érzékelés képességére, az információt tároló memóriák jelenleginél hatékonyabb szervezésére.

6.2 Társadalom és gazdaság

6.2.1 Az ember információfeldolgozó tevékenységének átalakulása

A hatékony vállalati információ- és ismeretkezelés megteremtése és működtetése tudásigényes feladat, és nem kis munka. A szellemi dolgozók¹³ (knowledge worker) termelékenységének növelése a legnagyobb kihívás, amellyel a szervezetek jelenleg szembenéznek. Mindez csak látszólag van ellentmondásban azzal, hogy a számítógépek, informatikai rendszerek egyre több munkát végeznek el az ember helyett.

Talán arra is lehetne gondolni, hogy az embereknek nem marad munka, azonban: „a munka nem vész el, csak átalakul”. Azaz, a konkrét adatok szintjén ugyan az informatikai rendszerek egyre több munkát végeznek el és helyettesítik az emberi munkát, ugyanakkor – ezzel egy időben – az ember *szellemi tevékenységének* jellege a vállalati munkában megnövekszik, és átalakul azzal, hogy a metaadatok kezelésére kell, hogy fókuszáljon. Ezt támogatják a jelentésalapú technológiák, amelyek segítségével meg lehet teremteni és működtetni lehet azt a szemantikus (tulajdonképpen értelmezési) környezetet, amely nélkül egyre növekvő mértékű és hatékonyságú gépi adatfeldolgozás nem lehetséges. Az emberek pedig egyre növekvő mértékben kell, hogy inkább a *szervezeti tudás kezelésével* foglalkozzanak, semmint a jól technológizálható operatív adatkezeléssel. A repetitív, irodai munka egyre jobban és általánosabban alakul át kvalifikált, szellemi tevékenységgé.

6.2.2 Tudásalapú szervezetek irányítása és értékének megőrzése

Nyilvánvalónak tűnik, hogy a tudásalapú társadalmat zömében „tudásalapú szervezetek” kell, hogy alkossák. Azon túl azonban, hogy minden szervezet – valamilyen formában – az azt alkotó emberek személyes tudására, ismereteire, képességeire és kapcsolataira

¹³ Arányuk a teljes munkaerő-állomány kb. 40%-a az USA-ban.

épül, gyakorlatiasabbnak látszik például azokat a vállalatokat illetni ma ezzel a megjelöléssel, ahol a vállalat értékét (például tőzsdei értékét) meghatározó és döntő mértékben az alkalmazottai és kapcsolatrendszerük adják, azoknak tulajdonítható, és ehhez képest a vagyoneleltárban kimutatható értéke elenyésző.

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 állandóan arra irányul, hogy az alkalmazotti és kapcsolati „tőke” – az adott lehetőségek mellett és amilyen mértékben lehet – független legyen e „tőke” hordozóitól, azaz az egyénektől. Ez természetesen teljes mértékben sosem érhető el, de egy tudásalapú szervezetnek, amelyik adott pillanatban pénzügyileg sikeres például nem szabadna, hogy irreálisan nagy értéke legyen, ha az informatikai rendszere alacsony szervezettségű, nincs összhangban az üzleti tevékenységgel stb.¹⁴ Ilyen esetben az alkalmazottak – elbocsátásuk esetén – teljes mértékben „kiviszik” a tudást a szervezetből, tehát csökken az értéke bár se a vezetők, se a kontrollerek, se az auditorok nincsenek ennek tudatában, sőt gyakran pozitívnak tekintik, mivel a vállalat papíron kimutatott – rövidtávon mért – hatékonysága növekszik. Ez nyilvánvaló ellentmondás, amit idővel – minél előbb – fel kell oldani, mert különben tudásalapúnak mondott szervezetek vezetése ellenőrizhetetlenné válik.

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 – nem utolsósorban a metaadat-kezelés és üzleti-informatikai kommunikáció fejlesztésével.

6.2.3 Tudásalapú társadalom és gazdaság kialakulása

A „tudásalapú társadalom” kifejezést manapság gyakran átgondolatlanul használják, és általában is kétséges, hogy pontosan mit is kell rajta érteni. Két jellegzetesség azonban valószínűleg nem kérdőjelezhető meg a *modern tudásalapú társadalomban*:

1. Általában az *emberi tudás* és a *hasznosítható tudással rendelkező emberek* megbecsülése mind erkölcsi, mind anyagi értelemben
2. Az *emberek* és a *gépi rendszerek* egymást kiegészítő, harmonikus együttműködése az információ és a tudás megszerzésében és felhasználásában

A tudásalapú társadalomban és gazdaságban ugyanis 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 rendszerekbe és ezen keresztül az emberek tudásának bővítése. A *jelentésalapú technológiák* természetüknél fogva ennek megvalósulását segítik elő.

6.2.4 Az információszabadság növekedése

Az emberek számára továbbra is a szövegek, beszédek, képek, videók formájában való kommunikáció a legtermészetesebb és a leghatékonyabb, és nem a mesterségesen előre

¹⁴ Részben erre figyelmeztetnek az elmúlt évtized gazdasági válságai és botrányai is („dotkom lufi” kipukkanása, Elron, WorldCom, Arthur Andersen pénzügyi visszaélései és csődjei stb.) A Basel 2 követelmények és a Sarbanes-Oxley törvény is ebbe az irányba mozdítja el a vállalatok vezetését.

rögzített, gyakran érthetetlen tartalmú és célú bizonylatok, űrlapok kitöltése – legyenek azok papíron vagy akár képernyőn.

Az eredményes kommunikációban nagy szerepe van a be nem tervezett, kéretlen (unsolicited) információközlésnek, amit viszont többnyire csak utólagos feldolgozással lehet kinyerni. Például, mint ügyfél egyetlen telefonhívással szeretnénk minden problémánkat és kérésünket elmondani egy adott témával kapcsolatban, és elvárjuk, hogy azokat mind érdemben kezeljék. A rugalmatlan telefonos menürendszerek, a mechanikusan („robotszerűen”) végzett és a konkrét körülményeket figyelmen kívül hagyó ügyfélszolgálati munka, valamint a kapott információk teljes körű feldolgozásának elmaradása csak felesleges akadályokat gördít ennek megvalósulása elé.

Az állampolgárok szabadságát ilyen módon növeli, ha az információk közlése és fogadása természetes képességeikhez közel álló módon történik – azaz jelentésben gazdagon, a kontextushoz kapcsolódóan és személyesen. A szemantikus technológiák növelik a társadalomban ezt az *információs szabadságot*: komolyabb digitális írástudás nélkül¹⁵ is mindenki egyenlő eséllyel vehet részt a társadalom és a gazdaság alapvető működésében.

6.2.5 Adatvédelmi szabályozás fejlődése

A jelentésalapú technológiák is – mint annyi más új, hatékony, korszerű technológia¹⁶ – megnövelik a veszélyét a magánélet megzavarásának és ezen belül a személyes adatokkal való visszaélésnek.

A szemantikus megközelítés lényege ugyanis, hogy többletinformációt (metaadatokat) tárolnak az alapadatokhoz kapcsolódóan. Kicsit leegyszerűsítve: minél több és jobb minőségű leíró információ áll rendelkezésre, annál könnyebben és jobban lehet az adatokat kezelni, feldolgozni, más adatokkal összekapcsolni. És annál inkább lehet ezeket más célra is felhasználni...

„Számon tarthatják, mit telefonoztunk, és mikor, miért, kinek. Felírhatják, miről álmodoztunk és azt is, ki érti meg.” Ha az eddig csak emlékeinkben őrzött képek közszemlére kikerülnek a webre, és ha a képekhez hozzákapcsolódik, hogy mit tartalmaznak, mikor készültek, vagy ha az elmondott szövegekkel együtt a források és fő témájuk is eltárolódik, akkor „sejthetjük-e, mikor lesz elég ok, előkotorni egy adatrekordot, mely jogunk sérti meg?”¹⁷ Automatikusan és nagy teljesítménnyel.

A szemantikus technológiák sok hasznuk mellett a mai világháló korszakában újabb kihívásokat jelentenek a jogi szabályozás számára. Csak remélhető, hogy kellő bölcsességgel és előrelátással, de sor kerül az adatvédelem jogi szabályozásának megújítására.

7. Hazai helyzet

7.1 Jelenlegi helyzet

¹⁵ Természetesen egy minimális szintű digitális írástudásra mindig is szükség lesz, de ez inkább egyfajta affinitást vagy készséget igényel a gépi rendszerekkel való kommunikáció irányában. Ez tehát jóval inkább az idegenkedés megszűnését jelenti, és nem azt, hogy egyre mélyebb ismeretére lesz szükség az információtechnológiának. A „digitális írástudás”-ban ismét az *írástudáson* lesz a hangsúly, és nem a „digitális”-on.

¹⁶ Mint például a mobiltelefon, az adatbányászat, az e-mail, a noteszgép, az elektronikus bankkártyák.

¹⁷ József Attila: Levegőt! (részlet) nyomán szabadon.

Az EU VI. keretprogramjában a jelentésalapú technológiák területén alig említésre méltó a magyar részvétel: egyedül az MTA SZTAKI vett részt egy ilyen projektben. Bár általában alacsony a részvétel az újonnan csatlakozott országok részéről, figyelemre méltó, hogy Szlovénia és Bulgária kiugró részesedést tudott elérni (egyenként kb. 1.2 M€-ra becsülhető).

Magyarországon 2000 óta folyik ugyanis kutatás a szemantikai integráció és az ontológiakezelés határterületein. Az IQSYS (illetve jogelődje az IQSOFT) egyik korábbi, EU-finanszírozású projektjének eredményeire alapozva (Semantic Integration via Logic & Knowledge, SILK) vezetett K+F projektet a Nemzeti Kutatási-Fejlesztési Program keretében (SINTAGMA, Szemantikai Integrációs Technológia Alkalmazása Grid-alapú, Modell-vezérelt Architektúrákban). A SINTAGMA egy olyan kísérleti technológiát hozott létre, amely az *információintegrációt szemantikai szinten* – azaz modellek, illetve ontológiák segítségével, jelentésgazdag módon – képes megvalósítani, és amely a kísérleti alkalmazásokban felhasználásra kerül grid-jellegű architektúrába szervezett hálózati gépeken keresztül.

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

7.2 Kutatások, fejlesztések

A magyar nyelvvel kapcsolatos számítógépes nyelvészeti kutatások mind akadémiai jelleggel (teljes grammatikai leírás, mintaszöveg-állomány), mind gyakorlati alkalmazásokban folynak (például nemzetközileg is terjesztett helyesírásellenőrzők és fejlett angol-magyar/magyar-angol fordítóprogram). A legfontosabb fejlesztéseket a Morphologic Számítástechnikai Kft., a Scriptum Informatika Kft. és az Alkalmazott Logikai Laboratórium végzi, míg az akadémiai kutatások terén az MTA Nyelvtudományi Intézet mellett ma a Szegedi Tudományegyetem tevékenysége a meghatározó.¹⁸ Ezeket egészítik ki a BME (például szószablya-algoritmus) és az MTA SZTAKI (egyértelműsítési eljárás) nyelvi fejlesztései, a BME-ben folyó beszéd felismerési kutatások eredményei, és nem hagyható ki az az MTA SZTAKI-ban folyó kutatási irány sem, amely képek és videók tartalmi feldolgozásával, annotálásával foglalkozik. Fontos megemlíteni a szintén az NKFP keretében megvalósult Magyar Egységes Ontológia (MEO) projektet is, amely egy adott szakterület általános szemantikáját leíró, magyar nyelvű általános ontológiát hozott létre. Érdekes további kutatási irány lehet az általános (például MEO által létrehozott) és a szövegkörnyezet-specifikus (például SINTAGMA által kezelt) ontológiák együttes alkalmazása a szövegértelmezésben.

7.3 Várható fejlődés

Az elmúlt néhány év nem kedvezett a hazai informatikai kutatásoknak általában, és ezen belül az olyan szerteágazó, tudás- és munkaintenzív területeknek különösen nem, mint a jelentésalapú technológiák. Ez egyrészt a hazai vállalkozások (multikat is beleértve)

¹⁸Itt hozták létre és tartják karban a magyar nyelv legnagyobb mintaszöveg-állományát, az ún. Szeged-korpuszt.

óvatos, alapvetően *átvevő-követő* jellegű innovációs megközelítésére, másrészt a magyar gazdaság innovációs rendszerének általános fej(1)etlenségére vezethető vissza. Megújulás jelenleg csak a 2008-ban beindult ún. technológiai platformok és klaszterek¹⁹ sikeres tevékenységétől várható. Remélhető, hogy a *számítógépes nyelvészet* és a *mobil eszközök/multimédia* területeken már létrejött technológiai platformok együttműködése – kiegészülve más kapcsolódó területeken szerveződő platformokkal (például üzleti intelligencia) – a szemantikus technológiák terén is valós, megvalósítható, gyakorlatias, ugyanakkor előremutató célokat tud kijelölni, amelyek mentén a klaszterekbe szerveződő hazai vállalatok sikeres projekteket tudnak végigvinni.

8. Összefoglalás

Az emberi elme szerkezetének és működésének egyes elemeit felhasználó, *kognitív technológiák* és rendszerek megjelenése a következő évtizedek egyre meghatározóbb tendenciája lesz. E technológiák közül az elme *rendszerező, asszociációs és következtető képességeit* mintául vevő, a metaadatok hatékony és sokoldalú kezelésére épülő, jelentésalapú technológiák ma már túlléptek azon a ponton, hogy csak figyelmet keltsenek.

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

A szemantikus technológiáknak az az ága, amely a *mindennapi és szakterületi tudás* megragadásával, kodifikálásával foglalkozik szemantikus *hálóknak, ontológiáknak és ismeretbázisoknak* formájában több évtizedes múlttal rendelkezik, és olyan egyedi tervezésű, évtizedes aprómunkával és kiemelkedően innovatív megoldásokkal létrehozott rendszerekben csúcsosodik ki, mint amilyen a Cyc.

Egy évtizeddel ezelőtt kezdődött a szemantikus technológiák kiterjedése a webre. Először a *szemantikus web* koncepciója fogalmazódott meg, amit hamarosan követett a vonatkozó szabványok megjelenése és elterjedése (XML, RDF). A két irányzat (ontológiakészítés és szemantikus web) az OWL ontológialeíró nyelv (Web Ontology Language) létrejöttével közös fejlődés előtt áll.

A weben keresztül azonban a szemantikus technológiák egy másik fejlődésvonalba is bekapcsolódtak, a webszolgáltatások területébe. A *szemantikus webszolgáltatások* a szolgáltatás-orientált architektúrák szemantikai szintre való emelését igyekeznek megvalósítani, azaz hogy a webszolgáltatások leoptimalisabb összekapcsolását és együttműködését (semantic interoperability) azok leírásainak és egyéb metaadatainak felhasználásával *logikai következtetések* és *mintaillesztések* útján önállóan is képesek legyenek elérni.

A következő évtizedben a már eddig kialakult (fent említett) megoldások *konzolidációja, integrálódása* és a *mindennapi használatba* való bevonulása várható. Az eszközpiacon a helyzetet egyik oldalról a túl sok kiemelkedő, de egyedi és drága, másoldalról a túl kevés általános, de széles körben elterjedt és ezért olcsó megoldás jellemzi.

¹⁹ A technológiai platformok feladata, hogy tíz éves távlatban meghatározzák egy-egy IT-szakterület javasolt hazai, stratégiai céljait, míg a klaszterek jóval rövidebb időtávban gondolkodnak, és már néhány éven belül megvalósítható, közvetlenül bevezethető fejlesztéseket hajtanak végre.

A vállalatoknak általában célszerű aktívan figyelemmel követniük az új – saját működési területüket érintő – kodifikációs szabványok kialakulását, és célszerű aktívan elő is segíteniük az ilyen szabványok megjelenését. Magának a témakörnek a megértése, és bizonyos kezdeti tapasztalatok szerzése ezen a területen, egyre inkább kritikussá válik. Van lehetőség, és érdemes bekapcsolni a hazai fejlesztéseket az európai vérkeringésbe. Továbbá célszerű lenne megvizsgálni, hogy a sikeres hazai kutatások és K+F projektek eredményei hogyan erősíthetnék egymás hatását: például kontextus-független (általános), illetve kontextusfüggő (speciális) ontológiák együttes használata a nyelvi feldolgozásban, illetve a jelentésalapú technológiákban.

Ajánlott irodalom

- Berners-Lee, T. – Hendler, J. – Lassila, O.: *The Semantic Web*. Scientific American, May 2001.
- Chira, Ovidiu: *The Semantic Web*. IDIMS Report, 2003.
(<http://ce.sharif.edu/courses/84-85/1/ce926/resources/root/SWIntro.pdf>)
- Davis, M.: The Business Value of Semantic Technologies. TopQuadrant Special report, 2004. (<http://web-services.gov/Davis.pdf>)
- Ding, Ying – Fensel, Dieter – Stork, Hans-Georg: *The Semantic Web: from Concept to Percept*. OGAI, 2003. <http://www.requirementsengineering.net/eleadership/content/The Semantic Web from Concept to Percept.pdf>
- Fensel, D. – Horrocks, I. – Van Harmelen, F. – McGuinness, D. – Patel-Schneider, P. F.: *OIL: Ontology infrastructure to enable the Semantic Web*. IEEE Intelligent System, 16(2), 2001. (www.cs.vu.nl/~frankh/postscript/IEEE-IS01.pdf)
- Havasi, Catherine – Speer, Robert – Alonso, Jason B.: *ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge*.
(<http://conceptnet.media.mit.edu/cnet3.pdf>)
- Hendler, J.: *The Dark Side of the Semantic Web*. IEEE Intelligent Systems, Vol. 22, No. 1, 2007.
- Lenat, Doug: *Cyc: A large-scale investment in knowledge infrastructure*. Communications of the ACM, 11:33–38, 1995.
(<http://www.cs.umbc.edu/471/papers/cyc95.pdf>)
- Liu, H. – Singh, P.: *ConceptNet — a practical commonsense reasoning tool-kit*. BT Technology Journal, Vol 22 No 4, October 2004.
(<http://web.media.mit.edu/~push/ConceptNet.pdf>)
- Minsky, Marvin: *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Scuster, 2005.
- Pollock, J.T. – Hodgson, R.: *Adaptive Information: Improving Business through Semantic Interoperability, Grid Computing, and Enterprise Integration*. John Wiley & Sons, 2004.
- Shadbolt, N., Berners-Lee, T., Hall, W. *The Semantic Web Revisited*. IEEE Intelligent Systems pp. 96-101, May 2006.
- Simon, Herbert A. and Associates: *Decision Making and Problem Solving*. In: *Report of the Research Briefing Panel on Decision Making and Problem Solving*. 1986, National

Academy of Sciences. National Academy Press, Washington, DC.

(<http://dieoff.org/page163.htm>)

- Sure, York et al. *On-To-Knowledge: Semantic Web Enabled Knowledge Management*, 2003. (<http://www.cs.vu.nl/~frankh/postscript/WI-book03.pdf>)

- Wang, Y.: *The Theoretical Framework and Cognitive Process of Learning*. Proc. 6th International Conference on Cognitive Informatics (ICCI'07), IEEE CS Press, August 2007, CA.