

## MAGYAR EGYSÉGES ONTOLÓGIA

### összefoglaló szakmai jelentés

NKFP-2/042/04

1. munkaszakasz

2004. december 1.-

2005. május 30.

projektvezető:

Szakadát István

<http://ontologia.hu>

BME MOKK

BME TMIT

Morphologic Kft.

Scriptum Rt.

ALL Kft.

MTA NYTI

SZTE SZTCS

#### Az 1. munkaszakasz részfeladatai:

- 
- 1.1. közönségszolgálati tevékenység szakirodalom-feldolgozása
  - 1.2. távközlési közönségszolgálati tevékenység felmérése
  - 2.1. ontológiákkal kapcsolatos szakirodalom feldolgozása
  - 2.2. ontológiaépítési módszertan kiválasztása
  - 2.3. ontológiaszerkesztő kiválasztása, használati útmutató készítése
  - 2.4. logikai nyelv, következtetési rendszer kiválasztása
  - 3.1. ontológiai előzmények a távközlés, közönségszolgálat terén

## Tartalom

Az 1. szakasz összefoglaló jelentése.....	2
Az ügyfélszolgálati tevékenység felmérése (1.1. és 1.2.).....	2
Az ontológiákkal kapcsolatos szakirodalom feldolgozása (2.1.) .....	3
Az ontológiamodell kérdései.....	3
Ontológiai-logikai rétegek.....	6
A rétegmodell áttekintése .....	6
A jelentésrepresentáció kérdései .....	9
Az ontológiaépítési módszertanok értékelése (2.2.).....	11
A logikai nyelv, következtetési rendszer kiválasztása (2.4.).....	11
Az ontológiaszerkesztő kiválasztása (2.3.).....	12
Az ontológiai előzmények felmérése a távközlés, közönségszolgálat terén (3.1.).....	14
A projekt egyéb tevékenységei.....	14
A csúcsontológiába tartozó adatok.....	14
Kooperációs lehetőségek .....	15

## Az 1. szakasz összefoglaló jelentése

A MEO projekt – 2004. december 1. és 2005. május 30. közti – első munkaszakaszában az alábbi részfeladatokat kellett elvégeznünk:

- 1.1. közönségszolgálati tevékenység szakirodalom-feldolgozása
- 1.2. távközlési közönségszolgálati tevékenység felmérése
- 2.1. ontológiákkal kapcsolatos szakirodalom feldolgozása
- 2.2. ontológiaépítési módszertan kiválasztása
- 2.3. ontológiaszerkesztő kiválasztása, használati útmutató készítése
- 2.4. logikai nyelv, következtetési rendszer kiválasztása
- 3.1. ontológiai előzmények a távközlés, közönségszolgálat terén

Az alábbiakban röviden összegezzük, hogy a sikeresen lezárult első projektfázisban milyen eredményeket értük el.

## Az ügyfélszolgálati tevékenység felmérése (1.1. és 1.2.)

Mivel a pályázati anyagban nem rögzítettük, hogy melyik távközlési vállalat milyen közönségszolgálati tevékenységéhez akarunk támogató alkalmazást fejleszteni, ezért a projekt elindulása után úgy döntöttünk, hogy két szálon is elindulunk (párhuzamosan). E célból felvettük a kapcsolatot egyfelől a Magyar Telekom/Matáv, másfelől az Invitel távközlési cégekkel, és interjúk során felmértük a terepet, hogy lássuk, milyen konkrét tevékenységet folytatnak ezek a vállalatok, illetve, hogy tudjuk, elméletileg milyen fejlesztési lehetőségeink adódnak ezekben az irányokban? Ezzel párhuzamosan természetesen feldolgoztuk a közönségszolgálati tevékenységgel kapcsolatos szakirodalmat is. Hogy a majdani fejlesztési feladat jól megfogható, és behatárolható legyen, ezért úgy döntöttünk, hogy a projekt számára korlátozzuk a közönségszolgálati tevékenység fogalmának terjedelmét, és csak egy leszűkített tartományra fókuszálva fogjuk a fejlesztéseket elvégezni. A kiválasztott terület egyik ágon a hibakezelés lett, míg a másik ágon a CAII Centerben dolgozók munkáját segítő tudásbázis szemantikai támogatását tervezzük elvégezni.

A Call center tevékenység leírására alkalmas általános keretrendszer felállítása után a részletes szakmai tanulmányunkban bemutatjuk a kommunikáció fontosságát, jellegzetességeit az ügyfélkapcsolati folyamaton belül. Általánosságban is összegezzük a sikeres kommunikáció feltételeit, majd alaposabban bemutatjuk, hogy milyen teendők vannak a hibakezelés területén, milyen feladatokat, hogyan láthatnak el a Call centerek ügyfélszolgálati munkatársai. Ehhez először tisztázni kell az ilyen szervezetek belső felépítését, majd fel kell tárni (illetve tudatosan kell tudni alakítani!) a Call centerekben dolgozó munkatársak rendelkezésére álló tudás minőségét, mennyiségét, szerkezetét, mélységét, illetve azt is tudni kell, hogy az itt folyó tudáskezeléssel szemben milyen elvárásokat kell érvényesíteni. Az ügyfélközpontú tudáskezelési modell legfontosabb jellemzői:

- a tudás átláthatósága
- A tudás szétosztása, elérhetővé tétele
- A tudás fejlesztése
- A tudás hatékonyságának fenntartása

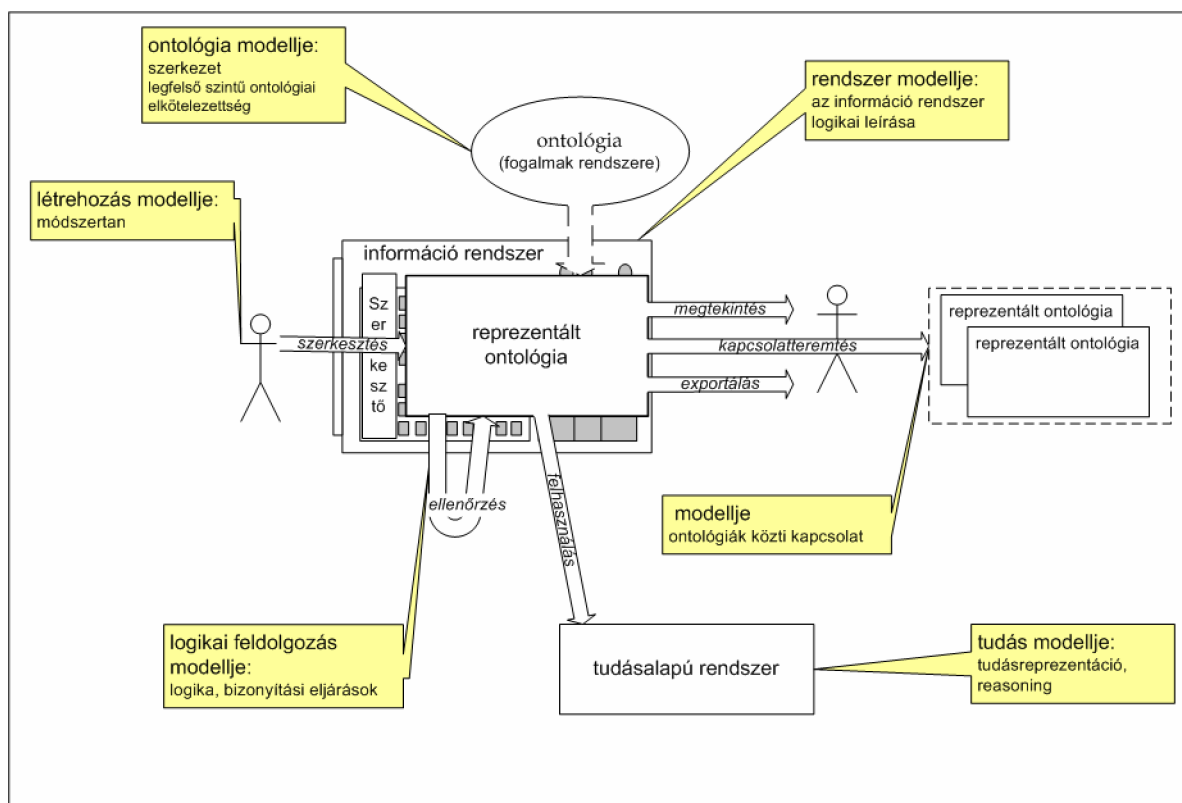
A közönségszolgálati feladatok általános jellegű áttekintése után elemeztük a hibakezeléssel kapcsolatos folyamatokat is. Áttekintettük a hatályos törvényi szabályozást, a konkrét terepfelmérés eredményeire támaszkodva megpróbáltunk felállítani egy általánosabb modellt a hibakezelés vállalati szabályozására vonatkozóan, majd röviden felvázoltuk, milyen lehetőségei vannak az ontológia alkalmazásának ezen a területen. Végül összeállítottunk egy kisebb glosszáriumot is, amely a közönségszolgálati tevékenységgel, s különösen a hibakezelés területével kapcsolatos legfontosabb fogalmak meghatározását tartalmazza.

## Az ontológiákkal kapcsolatos szakirodalom feldolgozása (2.1.)

Szakmai szempontból az ontológiai szakirodalom feldolgozása, az ontológia általános modelljének kidolgozása volt a projekt legfontosabb (és legnehezebb) feladata. Ebben a feladategyüttesben sokfajta szakember sokféle szaktudását, eltérő szempontrendszerét kellett valahogy egyszerre érvényesíteni, kihasználni. A feladatot tovább részfeladatokra bontva kisebb csoportokban folytattuk a munkát. Ennek során elkészült az ontológia általános modelljére vonatkozó javaslat, melyben fontos módszertani elveket is lerögzítettünk. Ugyancsak tisztáztuk a fogalmak jelentésrepresentációjával kapcsolatos kérdéseket, illetve az ontológia egészén belül elkülönítettük a különböző szintű és tartalmú logikai-ontológiai rétegeket egymástól.

### Az ontológiamodell kérdései

Első lépésként tisztáztuk az ontológiával kapcsolatos mindenféle modell egymáshoz való viszonyát, melyet az alábbi ábrával szemléltethetünk:



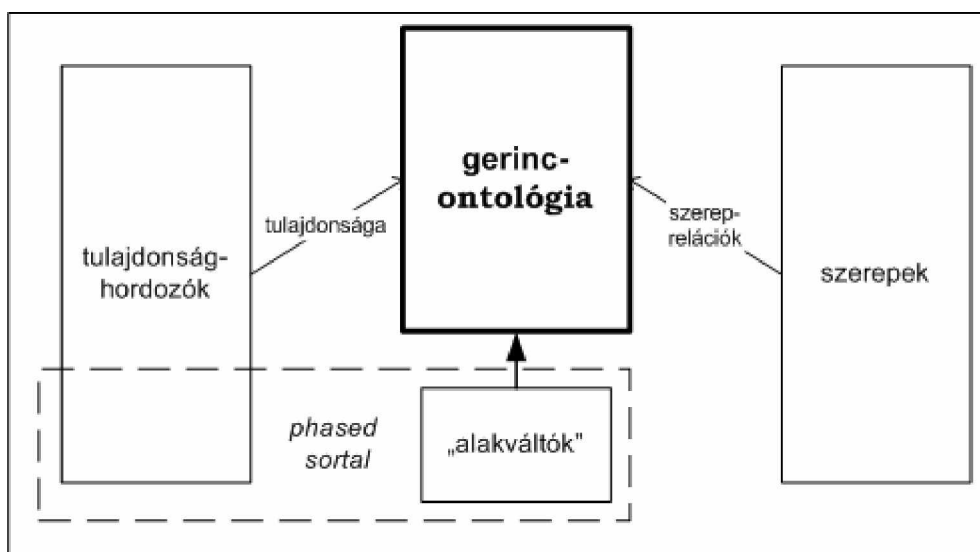
A modell alapfogalmainak (elsődleges, másodlagos reláció stb.) meghatározása után rövid, de rögzítettük a projekt metaontológiai, módszertani elképzeléseit, elkötelezettségeit, melynek során kiváltképp igazodtunk az ontológiákkal foglalkozó, meghatározó nemzetközi szakmai közösségek körében magasra értékelt OntoClean módszertan előírásaihoz. Ebben a módszertani keretrendszerben az ontológiai fogalmak metatulajdonságait vizsgálva, ellenőrizve lehet az ontológia egészen magas fokú konzisztenciáját biztosítani. Az OntoClean módszertan az alábbi kulcsfogalmak vizsgálatát írja elő:

- Rigiditás (*RIGIDITY*)
- Azonosság (*IDENTITY*)
- Egység (*UNITY*)
- Függés (*DEPENDENCE*)

Ezek alapján (a részletes szakmai jelentésben kifejtettek értelmében) a következő – jól definiált – fogalomkészlet áll rendelkezésünkre:

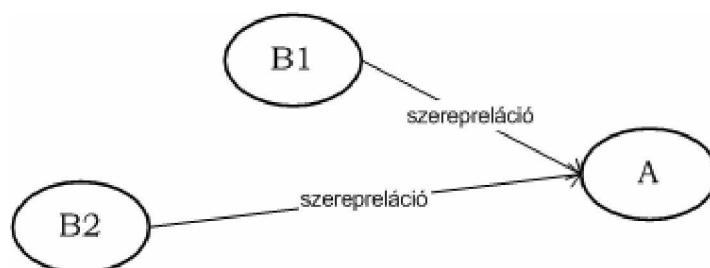
+O	+I	+R	±D	típus (type)	„sortal”
		+R	±D	kvázi-típus (quasi-type)	
~R	+D	materiális szerep (material type)			
~R	-D	(phased sortal)			
-R	±D	keverék (mixin)			
-O	-I	+R	±D	kategória	
		~R	+D	formális szerep (formal role)	
		~R	-D	tulajdonsághordozó (attribution)	
		-R	±D		

A fenti kategorizálás szerint a következő javaslat fogalmazódott meg az *ONTOLÓGIA* áttekintő struktúrájára vonatkozóan:

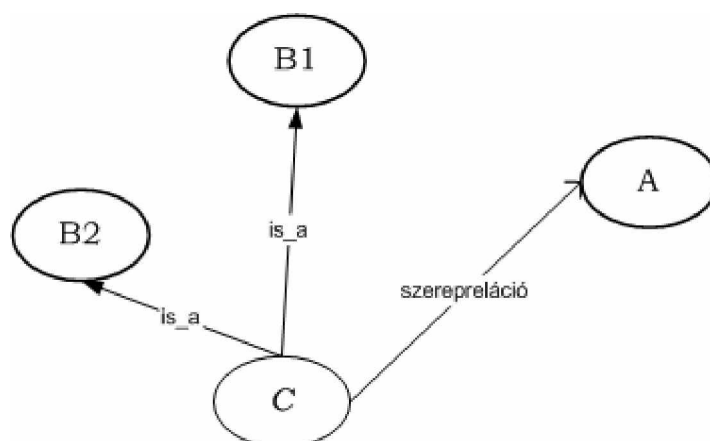


A modell az *ONTOLÓGIA*építés folyamatában is leírja az *ONTOLÓGIA* szerkezetét. Nyilvánvalóan az építés iteratívan történik, itt egy iterációs lépést írunk le.

- 1.) Először a gerinc*ONTOLÓGIA* készül el – a fogalmak és a köztük lévő relációk, és az axiómák.
- 2.) Azokat a fogalmakat kell sorba venni, amelyek relációként fogunk fel (általában a tevékenységeket jelentőeket (ajándékozás), de nem kizárólagosan (házasság)). A fogalom jelentésének megfelelő szereprelációk értelmezési tartományát alkotó fogalmakat kell megkeresni. Ha a szerepet betöltő egyedek halmazára nem kívánunk külön fogalmat alkotni, a megfelelő szereprelációkat feltüntetjük. Példaként legyen *A* a relációt képviselő fogalom, és a *B1*, *B2* fogalmak előfordulásai játszhatják a szerepreláció által kijelölt szerepeket. Ekkor az *ONTOLÓGIA* részletet a következő ábra illusztrálja:



Ha kívánunk szerepet betöltő egyedek halmazára külön fogalmat alkotni, ezt a már megtalált fogalmak fajtájaként tüntetjük fel, és ehhez kötjük a szereprelációt:



3.) Hasonlóan járunk el a fogalomként definiálni kívánt tulajdonsághordozókkal:

- kiválasztjuk azt a gerinc*ONTOLÓGI*Abeli fogalmat, amelyből származtatni kívánjuk, és mint fajtáját felvesszük;
- a megfelelő tulajdonsággal összekötjük a megfelelő tulajdonság relációval.

Közben módszeresen építeni kell a relációhierarchiákat és felírni az *ONTOLÓGI*A bővítéssel együtt járó axiómákat.

### Ontológiai-logikai rétegek

Az ontológia építéséhez (és használatához) nyilvánvalóan szükség van valamilyen logikai nyelvre, amelynek kifejezőerejét egyre több és több ontológiai elkötelezettség vállalásával növelhetjük. Mint minden axiomatikusan építkező, formális rendszer, a MEO is komponensekre, rétegekre bontható. Mivel a projekt jelen fázisában az ontológiaépítés nem minden kérdésére tudunk egyértelmű választ adni, ezért a réteges építkezés gyakorlatát részben arra használjuk a projekt következő szakaszaiban, hogy a különböző szinteken (és különböző csoportokban) folyó munkák hatással tudjanak lenni a többi szintekre (csoportokra). Azt az alapkérdést, hogy a két lehetséges ontológiatervezési és -építkezési módszertan, vagyis az 'alulról fölfelé' vagy a 'fölről lefelé haladás' metodológiai közül melyiket válasszuk magunknak, a projekt eddigi menete során nem tudtuk eldönteni. A jelen helyzetben – a szükségből erényt kovácsolva – megpróbálkozhatunk azzal, hogy egyfajta iteratív megoldást választunk. A SUO ontológiaépítő projektben erre a kérdésre az ún. *reflektív egyensúlykeresés* ('reflective equilibrium') módszertanának alkalmazását javasolták, ami annyit jelentett, hogy az alsószintű ontológiai egységeket folyamatosan sorolták be a felső szintű fogalmak alá, és bizonyos feltételek teljesülése esetén revíziót tartottak, s szükség esetén módosították a felső szintű fogalmakat. Ezt a módszert hasznosíthatónak tartjuk a MEO-projekt során is – a következők szerint. A különböző rétegekben adott ideig „szabadon” folyik az építkezési munka, de úgy, hogy az egyes rétegek közti kapcsolatokat, 'interfészeket' folyamatos kölcsönös kontroll alatt tartjuk, és ha valamelyik rétegben folyó munkák (és döntések) eredményeként szükségessé válik egy másik réteg tartalmának, szerkezetének, struktúrájának megváltoztatása, akkor azt megtehetjük az adott rétegen belül. Ebben a rendszerben a változtatási igény mindkét irányból kezdeményezhető (fentről lefelé vagy fordítva).

### A rétegmodell áttekintése

#### ± *alaplogikai réteg*

Ebbe a rétegbe valamely elsőrendű logika (DL, OWL) tartozik, de bizonyos pontokon szükségesnek látszik másodrendű, illetve modális logika alkalmazása is.

Ebben a rétegben a logikai konstansokon túl a reláció, függvény és műveletek tulajdonságaira, típusaira, a köztük megállapítható összefüggésekre vonatkozó terminusokat, definíciókat és formulákat írjuk le. Ebben a rétegben logikai konstansok mellett csak a halmaz terminusára építkező relációs fogalmak szerepelnek mint nem logikai konstansok.

Ez a réteg csak exoszkéletális egységeket tartalmaz.

± *elemi relációs réteg*

Az alaplogikai réteg terminusaira és formuláira támaszkodva újabb ontológiai elköteleződések mentén definiáljuk a legalapvetőbb, legfontosabb ontológiaképző relációtípusokat. A projektben eddig két relációtípus, a generikus, illetve a partitív aláförendeltje relációk elméletét vettük fel a rendszerbe. Jelen tudásunk szerint ez az elemi relációs réteg nem kell, hogy több, mint 8-10 relációból álljon. Iránymutatásként (hipotetikus alátámasztásként) hivatkozhatunk a nemzetközi (és magyar) szabványokban régóta rögzített és fenntartott gyakorlatra, amely ennyi relációtípust különít el és kezel.

Ez a réteg csak exoszkéletális egységeket tartalmaz.

± *relációs raszter*

A relációs raszter az elemi relációk alá sorolja be – egyre magasabb szinten, vagyis egyre több tulajdonság és kapcsolat felvételével – a különböző relációs fogalmakat. A raszter célja az, hogy „fentről” elérjük vele az ontológiai rendszerek, illetve taxonómiák elemeit. Vagyis ez a réteg köti össze a logikai és az ontológiai szinteket.

Ez a réteg nagyobb részben exoszkéletális, kisebb részben csúcsontológiai, elenyésző mértékben szakontológiai egységeket tartalmaz.

± *duáltaxonómia*

Ez a réteg a MEO csúcsontológia vázát alkotó fogalmak közti kapcsolatrendszert, tehát a MEO taxonómiáját írja le. A fő hangsúly itt az ontológiai egységek közti relációkon van. Azért duáltaxonómia, mert az egységek osztályba sorolása mellett az osztályképző tulajdonságok összefüggő rendszerével szeretnénk kiegészíteni a „hagyományos” taxonómiát.

Ez a réteg lényegét tekintve a csúcsontológia elemeiből áll, de tartalmazhat a relációs raszterből, esetleg a szakontológiából is néhány egységet.

± *kategoriatan*

A kategoriatan azoknak az ontológiai egységeknek a rendszere, amely a létezés, a létezők legmagasabb szintű, legegységesebb egységeit, fogalmait írja le. Ez a kérdés foglalkoztatja évezredek óta a metafizikával, ontológiával foglalkozó filozófusokat, és a történelem természetesen egymással nem összeegyeztethető, „versengő” rendszereket termelt ki mindeddig. Ez azt jelenti, hogy a projekt során döntenünk kell abban a kérdésben, hogy mely csúcsszintű elkötelezettségeket tesszük meg, vagyis melyik legfelső kategóriákat választjuk. Ezekre a kategóriákra a duáltaxonómia építéskor van szükségünk, de használnunk kell majd akkor is, amikor a fogalmainkat komponensekre bontjuk és egyes komponensekre vonatkozóan ontológiai kényszereket, kategoriális megszorításokat fogalmazunk majd meg.

Ez a réteg döntően exoszkéletális fogalmakat tartalmaz.

± *ontológiai egységek*

Ez a réteg tartalmazza az ontológia elsődleges tartalmát. Attól függően, hogy milyen célú és tartalmú ontológiáról van éppen szó, természetesen más és más lesz az ezen a szinten tárolt ontológiai egységek halmaza. Projektünkben többféle minősítést adhatunk az ontológiai egységeinknek attól függően, hogy mely ontológiákba tartoznak? Jelenlegi elképzeléseink szerint a csúcsontológiába, a középszintű és a

szakontológiába vagy pedig az exoszkeletális egységek közé lehet besorolni a fogalmainkat, melyek közül az első három, illetve a negyedik halmaz között kölcsönös kizárási reláció van. Az elképzelhető (bár nem túl gyakori) eset lehet, hogy egy-egy egységet mindhárom, tehát a felső-, a közép- és az alsó szintű ontológiába be lehessen sorolni.

± *fogalomreprezentációs módszertan*

Az egyes ontológiai egységeket úgy írjuk le, hogy a jelentésük alapján komponensekre bontjuk őket, melynek során messzemenően figyelembe vesszük az elmúlt évtizedekben a nemzetközi szakirodalomban megjelent, – legalább részben – bevált, hasznosítható nyelvészeti, szemantikai eredményeit.

± *ontológiai vázépítő módszertan*

Minden tudásszervezési rendszer, minden teaurusz, minden ontológia vázát a generikus (és kisebb részben a partitív) reláció adja, ezért az ontológiaépítés során mindvégig kiemelt figyelemmel kell lennünk erre a két relációra (illetve az általuk összekötött elemekre). A nemzetközi ontológiaépítő gyakorlatban egyre több műhely fogadja el, veszi át, implementálja az ún. OntoClean „ontológiatisztító”, konzisztenciaellenőrző módszertant, ami mi is megteszünk. A projekt során alkalmazni fogjuk e módszertan legfontosabb összetevőit, megoldásait.

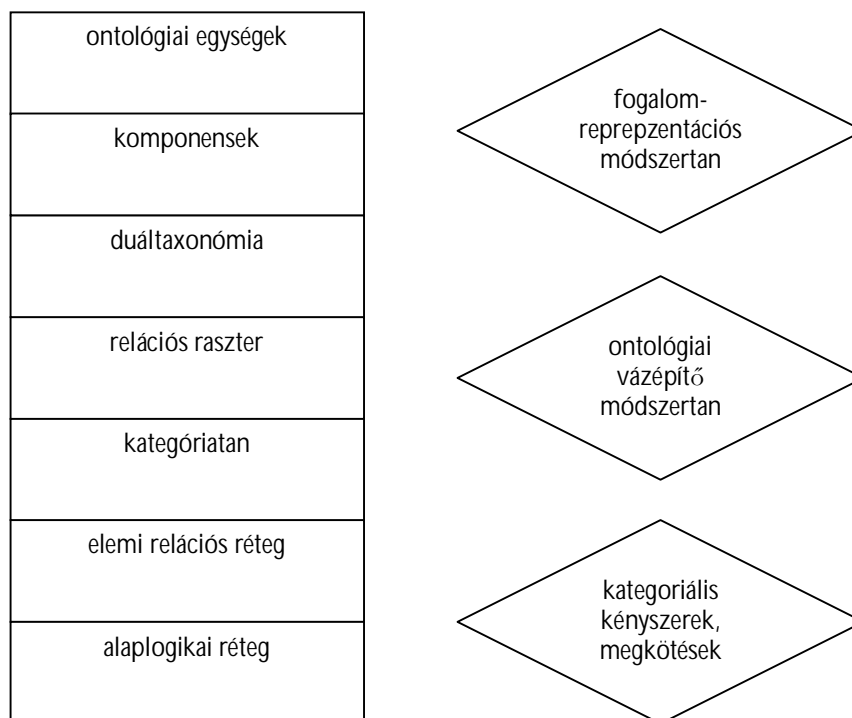
± *komponensek*

A fogalomreprezentációs módszertan legfontosabb eleme az, hogy a fogalmainkat komponensekre bontjuk. Ez a művelet várhatóan kis számosságú, még kezelhető nagyságrendű komponenstípust eredményez, és reményeink szerint az ontológiai egységekhez idővel egy – a munka során lassan kialakuló, véglegesülő – komponenstípus-tárból rendelhetjük hozzá a komponenseket.

± *kategoriális megkötések*

A relációs raszter és a duáltaxonómia felépítésében, illetve az ontológiai egységek leírásában, komponensre bontásában gyakran szükségünk lesz arra, hogy kategoriális megszorításokat tegyünk bizonyos komponensek tulajdonságaira, kapcsolataira vonatkozóan (pl. adott esetben ágens szerepben csak emberek szerepelhetnek). A relációs fogalmak esetében természetesen ide tartoznak a relátumok értelmezési tartományával kapcsolatos kategoriális kényszerek, de az ordinalitás, kardinalitás megkötései is.





## A jelentésreprezentáció kérdései

Fontos kérdésként vetődött fel a projektben, hogy egy általános, a természetes nyelvre orientált gépi alkalmazások céljaira is megfelelő tudásreprezentáció voltaképpen *milyen* tudás reprezentálását is tűzze ki célul. Emellett azonban éppoly fontos volt a *hogyan* tisztázása is. Utóbbi kérdés kezelése a következő „peremfeltételek” mellett képzelhető el.

A keresett reprezentációs keret:

- legyen totálisan explicit, azaz ne hivatkozzon „odaértett” információkra; ezt a célt egy pontos interpretációval rendelkező formális nyelv segítségével lehet biztosítani;
- a nyelv mögötti absztrakt fogalmi rendszert ábrázolja, azaz ne kötődjön a természetes nyelv esetlegességeihez, ám
- jól definiálható módon kapcsolódjon a természetes nyelvhez, azaz tegye lehetővé a természetes nyelvben releváns grammatikai (szintaktikai, morfológiai) funkciók és a jelentések közötti viszonylag problémamentes kapcsolatteremtést, ezen kívül pedig
- kapcsolódjon a szorosan vett ontológiához, azaz tartalmazzon olyan pontokat, amelyeken keresztül elérhetők az ontológiában tárolt generikus (és egyéb) információk.

A jelen feladat igényeihez jobban illeszkedik egy olyan megközelítés, ami a szavak jelentését mint részleges struktúraleírásokat fogja fel. Más szóval, az *ajtó* szó<sup>1</sup>, jelentésén keresztül egy egész kis struktúrát szakít ki a valóságból (nem pusztán objektumokat jelöl ki) és az *ajtó* szó helyes alkalmazása egy adott kijelentésben megköveteli, hogy az *ajtó* által leírt struktúra *beágyazható* legyen a kijelentés által ábrázolt szituációba. Ez a beágyazási reláció természetesen két irányban is bejárható. Egyrészt *ellenőrizhetem* a segítségével, hogy valaki egy adott szituációban helyesen (értelmesen) használja-e ezt a kifejezést (adott kijelentésben), másrészt pedig legalábbis

<sup>1</sup> A továbbiakban a ``szó” szót csak a ``tartalmas/lexikai szó” értelmében használjuk, hiszen a grammatikai szavak kezelése nyilvánvalóan ettől eltérő módon kell, hogy történjen.

részlegesen *megkonstruálhatom* azt a szituációt (modellt), amelyet a szóbanforgó, az *ajtó* szót tartalmazó kijelentés ábrázol. Egy szövegértési alkalmazás során nyilván az utóbbi iránynak van kitüntetett jelentősége. A rendszer a szójelentésekben tárolt struktúrák unifikációjával képes előállítani a kijelentésben ábrázolt szituáció egy olyan reprezentációját, amelyben aztán különböző, expliciten nem mondott tényekre, dolgok létezésére tud következtetni. Ennek részleteit természetesen majd a rendszert használatba vevő felhasználónak kell kidolgoznia a saját igényei szerint. A jelen projektben a cél csak egy általános felhasználó érdekeinek szem előtt tartása lehet.

A fent említett beágyazási szemantikának jól kidolgozott elmélete van az ún. DRT (*Discourse Representation Theory* -- Diszkurzusreprezentációs Elmélet) keretén belül, illetve annak későbbi változataiban. A fentieknek az az előnye is megvan továbbá, hogy a DRT-t, mint a neve is mutatja, (többmondatos) diszkurzusok kezelésére dolgozták ki, így annak a felhasználónak, aki a diszkurzusok/dialógusok gépi feldolgozása felé (is) kacsingat, ez a szójelentés-reprezentáció zökkenőmentes átmenetet biztosíthat abba az irányba is.

A jelentésleírásokban a szó által megjelenített prototipikus tudás leírására törekszünk. Ez annak a minimális szituációnak a maximálisan explicit leírását jelenti, amelyben az adott szó alkalmazható. Másképpen megfogalmazva, a jelentésleírás az adott szót tartalmazó mondatok által leírt szituációk metszetében található szituáció legrészletesebb leírása kell, hogy legyen (ideális esetben), mert egy szó jelentésimportja azzal a struktúrával jellemezhető, amit a szó minden olyan mondatba bevisz, ami tartalmazza azt. Persze, ez a meghatározás több szempontból is pontosításra szorul, hiszen pl. sokszor semmi nem garantálja, hogy ez a metszet nem lesz üres vagy éppen igen kevésbé informatív (túl sematikus). Ez a kérdés közvetlenül kapcsolódik a homonímia vs. poliszmia kérdésköréhez, és túl összetett ahhoz, hogy egyszerű megoldásokra számíthassunk. Itt csak annyit lehet megjegyezni, hogy ez esetben elkülönítendőek a szót tartalmazó mondatok olyan alosztályai, ahol a fenti feltétel már teljesül. Egy másik pontosítási irány pedig a grammatikai szavak szerepének figyelembe vételét jelenti, hiszen pl. a tagadással sok esetben éppenséggel olyan szituációt ábrázolunk, ahol a szó által képviselt szituációelem nincs vagy nem lehet jelen.

A fenti elképzelés egy az elsőrendű logika kereteit túl nem lépő formális keretben is megvalósítható. Ennek az az ára, hogy eseményekről, szituációkról, stb. képesek legyünk entitásként beszélni, azaz az elsőrendűség keretein belül maradás a reifikáció széles körű alkalmazását követeli meg. Ez azonban nem probléma, hiszen az ontológiai komponens többek között éppen az ilyen reifikált entitásokra vonatkozó tudást tárolja.

Pontosabban fogalmazva, a kívánt ] formális nyelv egy szortált (azaz típusos) változókkal rendelkező elsőrendű nyelv, amelyben a változók értéktartományára vonatkozó tartalmazási („generikus”) kikötéseket az ontológia tárolja. Más szóval, ] nem-logikai konstansai típusos argumentumhelyeket tartalmaznak, és tetszőleges formula interpretációja csak abban az esetben helyes, ha a formulában található változók értékeiket a megfelelő ontológiai tartományból veszik. A természetes nyelvhez való kapcsolódást megkönnyítendő a reprezentációk a tematikus szerepek használata mellett elkötelezett neo-davidsoniánus keretet feltételeznek. Ez a keret lehetővé teszi a későbbi felhasználók részére, hogy az egyes, mondatbeli grammatikai szerepeket (alany, tárgy, stb.) viszonylag egyszerűen köthessék a jelentésreprezentáció megfelelő elemeihez. Ebből következően ] nem-logikai konstansai a következő csoportokba sorolhatók:

- a tartalmaz szavakhoz rendelt konstansok, pl. **ajtó**, **csenget**;
- a leendő felhasználó számára a természetes nyelvvel történő kapcsolatteremtést segítő „interfész fogalmak” konstansai, pl. tematikus szerepek mint **AG(ENT)**, **PAT(IENT)**, **INS(RUMENT)**, **GOAL**, **BENEF(ICIARY)**, **EXP(ERIENS)**;

- a tartalmas szavak igen nagy körének jelentésében szerepet játszó igen általános fogalmakhoz rendelt konstansok, pl. az artefaktumokat jelentő szavak esetében a **Function**, vagy a szintén alapvető **Cause** (ezekre a félreértést elkerülendő angolul hivatkozunk, mert nem azonosak a magyar „funkció”, illetve „ok” tartalmas szavak reprezentációival).

## Az ontológiaépítési módszertanok értékelése (2.2.)

A projekt fontos módszertani feladatként – a nemzetközi szakirodalom segítségével – értékeltük a legelterjedtebb, legismertebb ontológiaépítési módszertanokat. Ezek az alábbiak voltak:

- Uschold & King
- Grüninger & Fox
- KACTUS projekt
- SENSUS
- On-To-Knowledge
- Methontology

Az értékelés során az alábbi szempontokat vettük figyelembe:

- Az ontológiaépítési folyamat kidolgozottsága, a folyamatleírás részletessége
- Definiált fogalomkészlet: a módszertan alapfogalmainak meghatározása
- Ontológia-életciklus: a kezdeti specifikációktól a fenntartásig, ill. üzemeltetésig terjedő folyamat.
- Öszeilleszhetőség.
- Kiegészíthetőség.
- Integrálási képesség: már létező ontológiák beépítésének lehetősége.
- A fogalmi rendszer felépítésének módja:
- A fogalmak rendszerezésének módja, fogalmak közötti kapcsolatfajta.
- Tanulhatóság.
- Ontológiaértékelő eszközök kidolgozottsága.
- Az alkalmazott ontológianyelv.
- Méret, dimenzió, terjedelem a gyakorlati megvalósulások alapján.
- Az ontológiáról általában (szemléletmód, ontológiafogalom, főbb jellemzők).

Bár az önmagában vett értékelések természetesen megteremtették a különböző módszertanok összehasonlításának lehetőségét, és ez alapján dönthettünk volna arról is, melyik módszertant választjuk a projekt számára követendőnek, de ezt a választást tudatosan a projekt második szakaszára toltuk el. Menet közben ugyanis kiderült, hogy amíg nem jutunk konszenzusra az ontológiával kapcsolatos alapvető modellezési kérdésekben, addig nincs értelme sem a módszertan kiválasztásának, sem a gyakorlati munka elindításának.

## A logikai nyelv, következtetési rendszer kiválasztása (2.4.)

A számítógépek számára érthető és kezelhető ontológiák formális nyelven vannak leírva. Az ontológianyelvek azonban „csak” az eszközt biztosítják a tartalom kifejezéséhez, önmagukban nem hordoznak információt. Az egyes ontológianyelvek közötti különbség éppen azáltal ragadható meg, hogy megvizsgáljuk, milyen lehetőségeket biztosítanak a tényleges tudás, illetve

tartalom rögzítésére, kifejezésére. Az ontológia alapvető jellemzőit jelentősen befolyásolhatja a formális leírás során használt nyelv.

A MEO projekt keretében készülő ontológiák kifejezőerejét, hasznosítási lehetőségeit is meghatározza a majdan alkalmazott ontológianyelv. A projekt számára megfelelő ontológianyelv kiválasztását természetesen befolyásolja, illetve meghatározza a használt ontológiaépítési módszertan és az ontológiaszerkesztő is. A projekt első szakaszában elvégeztük az ontológianyelvek elemzését és összehasonlítását. az alábbi nyelveket vizsgáltuk meg:

- o 1.1 XOL
- o 1.2 SHOE
- o 1.3 RFML
- o 1.4 RDFS
- o 1.5 DAML+OIL
- o 1.6 OWL
- o 1.6.1 OWL Lite
- o 1.6.2 OWL DL
- o 1.6.3 OWL Full
- o 1.7 SWRL
- o 1.8 SWRL-FOL
- o 1.9 OCML
- o 1.10 Ontolingua
- o 1.11 FLogic
- o 1.12 CycL
- o 1.13 Loom
- o 1.14 PowerLoom
- o 1.15 KM
- o 1.16 EPILOG
- o 1.17 SNePS

Külön fejezetet szenteltünk a Leíró logika (Description Logic) alapos bemutatásának, mert – úgy tűnik – ez a formális nyelv biztosítja a leghasználhatóbb ontológia-nyelvek közös alapját.

A nyelvek jellemzésén, elemzésén túl a fontosabb nyelvek esetében konkrét példákat biztosítottunk a nyelvek sajátosságainak szemléltetésére, összeszerkesztettünk még egy – az ontológianyelvekkel kapcsolatos fogalomkészletet felölelő – rövid glosszáriumot is, végül táblázatos formában is összefoglaltuk a különböző ontológianyelvek legfontosabb jellemzőit, továbbá megadtunk egy „áttérési táblát” a különböző ontológianyelvek között.

Az összehasonlító elemzés eredményeként az az ajánlás fogalmazódott meg, hogy érdemesnek tűnik több nyelven is leírni a csúcsontológiánkat, mert ezáltal többféle célra is lehetne használni.

Egy OWL DL szintű formális leírás alkalmazható lenne a modell konzisztenciájának ellenőrzésére, illetve az osztályok közötti kapcsolatok ellenőrzésére. Az OWL DL-es leírás előnye, hogy XML-es szintaxisa miatt az ontológia egy másik nyelvre átalakítása viszonylag egyszerű. Az SWRL egy ilyen továbblépés lehet, ha nem elegendő a DL-es kifejezőerő. Az OWL DL további előnye, hogy több szerkesztő, illetve konzisztencia ellenőrző rendszer készült hozzá.

A távközlési ontológia létrehozásakor valószínűleg nagy szerepük lesz az eseményeknek, relációk közti kapcsolatoknak. Amennyiben ezekről bonyolultabb kijelentéseket szeretnénk tenni és a következtetésekhez ezeket alkalmazni szeretnénk, akkor célszerű olyan nyelvet választani, amelyik ezt valamelyest támogatja. (Erre legmegfelelőbbnek a frame alapú nyelvek tűnnek, bár kétségtelen, hogy ezekben megfogalmazni bonyolultabb dolgokat szinte már programozásnak minősül.) Amennyiben különböző forgatókönyveket is modellezni kell, akkor néhány speciális nyelv erre is alkalmas. (KM, EPILOG, SNePS)

## Az ontológiaszerkesztő kiválasztása (2.3.)

Az ontológiaépítés és –fenntartás egyik kulcskérdése az, hogyan milyen eszközökkel, milyen informatikai támogatás mentén lehet az ontológia tartalmát, vagyis a fogalmakat, állításokat, kényszereket megfogalmazni, módosítani? Egyértelmű, hogy ehhez ma már szükség van valamilyen ontológiaszerkesztő alkalmazás rendszerbe állítására. Ebben a részfeladatban összehasonlítottuk az elérhető ontológiaszerkesztőket, értékeltük azokat, majd döntöttünk arról, hogy – jelenlegi tudásunk alapján – a MEO projekt számára melyik tűnik a legalkalmasabbnak.

A döntési folyamatban a feladat által megkívánt metódust választottunk, az ún. analitikus hierarchikus folyamatet (Analytic Hierarchy Process, AHP), amely egy intuitív és hatékony

módszer a többtényezős döntési alkalmazásokban (a részletes tanulmányunkban röviden összefoglaltuk az AHP-eljárás matematika modelljét).

Az összehasonlítás alapjául egy szakmai tanulmányban összehasonlított 96 szerkesztő közül azokat választottuk, amelyek a majdani felhasználók preferenciáinak legjobban megfelelnek, valamint van elérhető ingyenes verziójuk, amelyekkel a tesztek elvégezhetők. A programok rangsorát kicsi belső felmérésre alapoztuk. A rangsort a megkérdezettek (a projektben résztvevő szakemberek, illetve a potenciális felhasználók) preferenciáit aggregáló AHP módszer segítségével állítottuk fel, mellyel a szerkesztők legfontosabb jellemzőit fontossági sorrendbe tudtuk rendezni. A sorrend felállításánál tekintetbe vett, legfontosabbnak ítélt jellemzők a következők voltak:

- Támogatja-e a nemzetközi szabványnak tekinthető adatrepresentációs modellt, az OWL/RDF formátumot?
- Rendelkezik-e grafikus megjelenítő modullal, amely a fogalmi hálót képes kirajzolni, esetleg szerkeszteni is grafikus úton? (Ha rendelkezik ilyennel, mennyire fejlett szolgáltatást nyújt?)
- Rendelkezik-e a szerkesztő beépített konzisztencia ellenőrző rutinokkal, melyek a végeredmény minőségbiztosításában, korrigálásában nyújtanak segítséget?
- Támogatja-e több felhasználó egyidejű munkáját ugyanazon az adatbázison (konfliktuskezelés, verziókövetés, ...)?
- Képes-e más lexikai adatbázisokból elemek konvertálására és beillesztésére az éppen szerkesztés alatt álló ontológiába?
- Nyújt-e szótárszerű, lexikai támogatást (elemek címkézése, szűrések, keresés, csoportosítások)?

A felsorolt hat jellemző szerinti, a felhasználók preferenciáin alapuló rangsorból a következő szerkesztők kerültek kiválasztásra:

- Protégé (<http://protege.stanford.edu/index.html>), OWL pluginnal
- Semtalk ([www.semtalk.com](http://www.semtalk.com))
- WebKB ([www.webkb.org](http://www.webkb.org))
- WebODE (<http://webode.dia.fi.upm.es/WebODEWeb/index.html>)
- DL-Workbench (<http://projects.opencascade.org/dl-workbench/>)
- OntoEdit (<http://www.ontoprise.com>)

Az értékelések célja a MEO projekt fejlesztése során fellépő ontológiaépítési feladatok elvégzésére különböző ontológiaszerkesztő környezetek tanulmányozása, valamint a célokhoz megfelelő eszköz kiválasztása volt. Vizsgálataink során szem előtt tartottuk, hogy – amennyiben nem sikerül minden szempontból alkalmas eszközt találni – ajánlásokat tudjunk tenni, hogy a projekt tagjai által fejlesztendő szerkesztő eszköz pontosan milyen megoldásokat tartalmazzon, milyen kritériumoknak feleljen meg. Emiatt a részletes elemzésben vizsgált szerkesztő eszközöknél feltüntettük azok gyakorlati használata során megállapított kedvező, és hátrányos tulajdonságait.

Az elvégzett vizsgálatok alapján azt találtuk, hogy a megvizsgált szerkesztők közül a Protégé 3.0 program felel meg legjobban az általunk támasztott követelményeknek. Tulajdonképpen az összes lényegesnek tartott jelvárást magas színvonalon megvalósítja, vagy az alaprogram részeként, vagy csatolható pluginként. Emellett kezeli a legmodernebb, nemzetközi szabványoknak megfelelő formátumokat, így alkalmas lehet a MEO projekt során fejlesztendő ontológiához felhasználni.

## Az ontológiai előzmények felmérése a távközlés, közönségszolgálat terén (3.1.)

Bár az ontológia fogalmának a filozófiában többezer éves hagyománya van, az ipari világba csak a mesterséges intelligencia kutatások egyik kísérő fogalmaként került be évtizedekkel ezelőtt. A jelenkori kiemelt (és folyamatosan növekedő) népszerűségét azonban vélhetőleg a szemantikus web kezdeményezésnek köszönheti, amely az utóbbi években már az ontológialeíró nyelv szabványát (OWL) is elkészítette. A fogalom divatos mivolta ellenére tényleges (és még inkább: működőképes) ontológia-alkalmazásokról meg nagyon keveset hallani (nem csak nálunk nem, de sehol máshol a világban).

A projekt részfeladata az volt, hogy felmérjük, vannak-e szemantikával, ontológiai tudással támogatott fejlesztések, alkalmazások a távközlési közönségszolgálat területén, és ha igen, ezek milyen fajta erőforrásokat, megoldásokat, szabványokat használnak (ha használnak)? A vizsgálat lényegében megerősítette az előzetes feltételezéseinket, bár nagyon hasznos kezdeményezésekre bukkantunk. Az általunk megismert hazai távközlési vállalati szegmensben (az egész Telekom-csoport, Invitel) gyakorlatilag nem használnak ontológiai tudást sem a közönségszolgálat tevékenység, sem másfajta tevékenységek támogatására. Ennek ellenére találtunk olyan korábbi kezdeményezéseket, illetve ezek eredményeként létrejött olyan erőforrásokat, amelyeket hasznosíthatónak vélünk a projekt későbbi mentében.

Ilyen hasznosítható erőforrásnak kell minősítenünk azt a távközlési teauruszt, amely eredetileg már több, mint egy évtizede készült, s emiatt az eredeti állomány (amit könyv alakban is kiadtak) érvényessége, frissessége, megbízhatósága megkopott, a váza mégis olyan alapot jelenthet, amelyre „rá lehet aggatni” az elmúlt időszakban megjelent új fogalmakat. Ezt egyébként – legalább részben – elvégezték a távközlési teaurusz állományával is, de a projektnek nyilván még sok feladat lesz ezzel, mert a fejlesztendő konkrét alkalmazás által megkívánt logika mentén biztosan sok „mára már hiányos”, kiegészítenivaló részt találunk majd a teauruszban.

A felmérések során egyébként találtunk még másfajta, később hasznosítható erőforrásokat, melyek alapján a munka közben kialakított vállalati együttműködés keretében elkezdtünk egy vállalati korpusz felépítését, illetve egy kétnyelvű értelmező szótár összeszerkesztését a vállalati dokumentumokban található információk feldolgozásával. Mind a – sok területen kétnyelvű – vállalati szövegtörzs, mind a kétnyelvű értelmező szótár komoly segítséget jelenthet majd az szakontológiába beillesztendő fogalmak jelentésének, illetve relevanciájának feltárásakor.

### A projekt egyéb tevékenységei

A projekttervben vállalt feladatok eredeti ütemezését nem tartottuk be minden esetben, ami mögött egyaránt akadtak elméleti és gyakorlati megfontolások. Az előzetesen nem tervezett (illetve nem erre a projektszakaszra tervezett) feladatok közül jelentős előrelépés történt a csúcsontológia elemeinek összeállításában, az adatmodellezés és előzetes adatbázis-építés terén, illetve a kommunikációt és kooperációt biztosító ontológia portál kiépítésében.

### A csúcsontológiába tartozó adatok

A projekt kezdetén – a csúcsontológia tartalmáról folytatott vitával párhuzamosan létrehoztunk egy olyan több nyelvű „szótári állományt”, amely a csúcsontológiánk alapját jelentette. Kiindulásként a Longman Dictionary of Contemporary English (LDOCE) szótárt vettük, amely a durván 50 ezer angol szó meghatározását tartalmazza. Mivel az LDOCE szerkesztői tudatosan törekedtek arra, hogy minden a szótárban szereplő szó egy 2000 szavas ‘core vocabulary’ segítségével legyen definiálva. Ezt az ún. „magszókinccset” vettük a munka alapjául. Ezt kiegészítettük egy másik – igen hasonló – lista, a Whitney-féle szanszkrit tövek elemeivel. Bár Whitney kevesebb, mint 900 elemet tartalmaz, de ezek mind igék, ami azért fontos a projekt

számára, mert általában az igék (hiánya) jelenti(k) az ontológiák leggyengébb pontját. Ezáltal tehát előállt egy 2500-3000 szavas szótárunk, és bár az igazi feladat ez ezek mögött álló koncepciók kidolgozása, mégis nagy előrelépésnek kell tekintenünk ezt az eredményt.

Ezután ennek az – angol nyelvű – állománynak elkészült a magyar, majd a latin, végül a lengyel nyelvű fordítása, vagyis előállt egy olyan durván 3 ezres fogalom-állományunk, melynek négy nyelven állt rendelkezésre az ún. nyelvi kötése. Ekkor természetesen csak a szavak – különböző nyelvű – listái voltak meg, ebben a szakaszban a fogalmak formalizálásának még nem láttunk neki. A magzókincs létrejötte azonnal felvetette az a problémát, hogy miként lehet majd ekkora nagyságrendű szóállományt az osztott elvű működésen, a kooperáción alapuló praxisban konzisztens és rugalmas módon kezelni, amire nyilván csak az adatbázisban tárolás (és kezelés) jelentheti az igazán jó megoldást. Meg kellett tehát tervezni az ontológiai és szótári egységek tárolására alkalmas adatmodellt. A modelltől elvárt legfőbb funkció ebben a fázisban csak annyi, hogy a modell alapján felépítendő adatbázissal segíteni tudjuk a projekt kooperációs tevékenységét.

A projekt következő jelentős eredményeként értékelhetjük, hogy az összeállt magzókincsből a fenti adatmodellre támaszkodva elkészítettük a MEO-csúcsontológia első – kísérleti – adatbázisát. Az adatbázist Microsoft Access adatbázis-kezelő programban készítettük el, de a projekt második fázisában ezt az adatbázis-tartalmat át fogjuk tenni olyan adatbázis-kezelő platformra, amely egyfelől hálózati környezetben működik, másfelől nyílt forráskódú termék (valószínűleg PostgreSQL-platformra). Az Access-állományt kitettük a portálra is, ahonnan szabadon letölthető.

## Kooperációs lehetőségek

A projekt elsődleges célja egy csúcsontológia és egy szakontológia felépítése, egy általános célú ontológia-infrastruktúra, azaz valamiféle ontológiamenedzsment képesség, módszertan kialakítása, valamint egy olyan konkrét alkalmazás elkészítése, amelyben szemantikai/ontológiai tudást lehet hasznosítani. Mindezekon túl azonban megfogalmaztunk egy olyan célt is, amely azokat az elveket, feltételeket, gyakorlati lépéseket, elvárásokat, megoldásokat fogja össze egy közös módszertanba, amelyre bármikor szükség lehet, amikor ontológiák fejlesztéséhez, alkalmazásához eltérő tudású, eltérő érdekeltségű, eltérő motivációjú emberek, szervezetek közötti kooperációra van szükség.

Fontos igazodási szempont még a projekt számára, hogy vállalásainkat a lehető leghozzáértőbb hozzáférés szellemében tettük meg, vagyis azt ígértük, hogy a projekt valamennyi fontos eredményét szabadon hozzáférhetővé tesszük.

A fenti két projektcélt szem előtt tartva úgy szerveztük meg a projekt kommunikációját, hogy folyamatosan eleget tegyünk a magunk számára felállított elvárásoknak. Ezért a projekt elején egy önálló portált hoztunk léte, ahol egyfelől folyamatosan elérhetővé tettük a projektmunka során keletkező dokumentumokat, erőforrásokat, másfelől ahol a nyilvánosság előtt folytattuk a szakmai vitáink jelentős részét. Természetesen a kiscsoportos szakmai munka minden részlete, a munkaközi anyagok nem kerültek bele ebbe a szakmai nyilvánosságba, de a „teljes nyilvánosság” elvét lényegileg mindvégig betartottuk a projekt eddigi szakaszában. A portálon egyébként minden érdeklődő szabadon regisztrálhatja magát, miután a projektagokkal közel azonos jogosultságokat szerezhet magának.

A portál az alábbi címen érhető el:

<http://ontologia.hu>