

MIEO

MAGYAR EGYSÉGES ONTOLÓGIA

összefoglaló szakmai jelentés

NKFP-2/042/04.
2. munkaszakasz
2005. június 1.-
2005. november 30.
projektvezető:
Szakadát István
<http://ontologia.hu>

BME MOKK
BME TMIT
Morphologic Kft.
Scriptum Rt.
ALL Kft.
MTA NYTI
SZTE SZTCS

A 2. munkaszakasz részfeladatai:

- 1.3. távközlési közönségzolgálat gépi támogatásának lehetőségei
- 1.4. távközlési közönségzolgálati tudásbázis konszolidálása
- 2.5. következtetési rendszer implementálása
- 2.6. az OntoClean módszertan feldolgozása, implementálása
- 2.7. konzisztencia-menedzsment képesség kialakítása
- 2.8. ontológia-formalizmus kialakítása, fejlesztői útmutató
- 3.2. a távközlési teaurusz felújítása, transzformálása
- 4.1. a hasznosítható csúcsontológiák előkészítése
- 4.3. igei vonzatkeret-tár felépítése

Budapest, 2005. december 5.

Bevezető

A 2. számú projektjelentésben beszámolunk arról, hogy a Magyar Egységes Ontológia (MEO) projekt 2. munkaszakaszában milyen feladatokat végeztek el a konzorcium tagjai. A jelentés tagolása igazodik a 2. fázis részfeladataihoz (melyeket a jelentés 1. oldalán soroltunk fel).

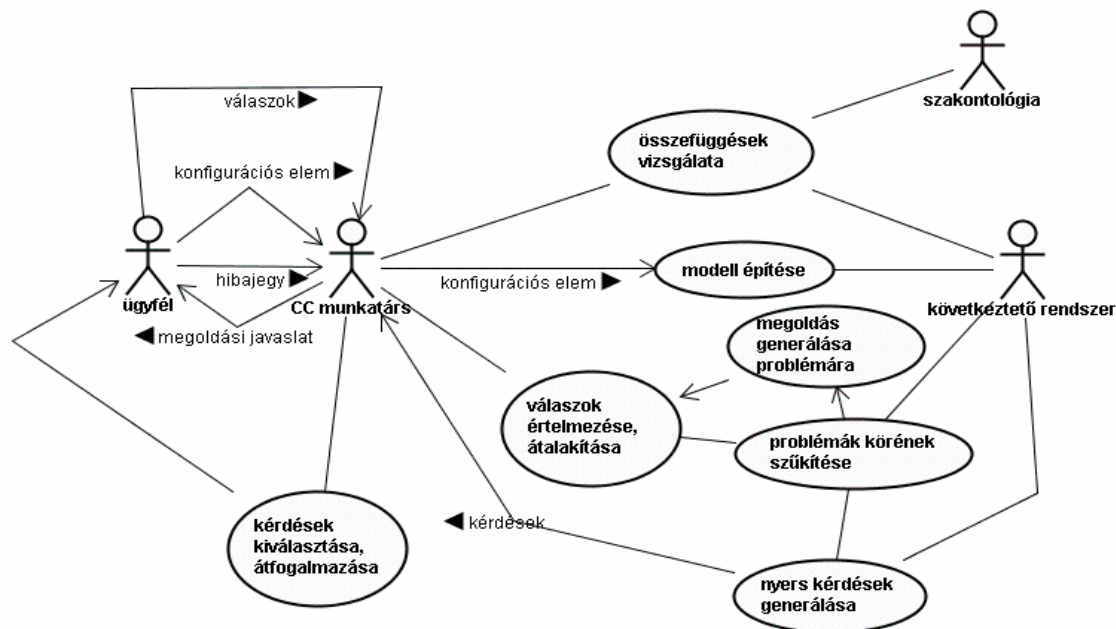
A közönségszolgálati tevékenység

A projekt rendszerfejlesztési feladatai között szerepel a távközlési közönségszolgálati tevékenység támogatására alkalmas informatikai rendszer megvalósítása. A projekt második szakaszában az 1.3. és 1.4. részfeladatokkal kellett foglalkoznunk. Egyrészt funkcionális specifikációt kellett készítenünk a közönségszolgálati tevékenység támogatására, másrészt fel kellett mérni, milyen – használhatónak ígérkező – tudásbázisok, részrendszerek állnak rendelkezésre a konkrét vállalati gyakorlatban, és ezeket a használható erőforrásokat konszolidálni kellett a majdani célok és felhasználási módok ismeretében és érdekében.

A közönségszolgálat gépi támogatásának lehetőségei

A projekt első szakaszában pontosítottuk a 'távközlési közönségszolgálati tevékenység' tartalmát, és a hibakezelés problémakörét választottuk ki magunknak a közönségkapcsolati munkatársak által végzett munkafajták széles spektrumából. A második szakaszban tovább pontosítottuk az elvégzendő feladat megfogalmazását, és az egyéni ügyfeleknél felmerülő, az internethasználattal kapcsolatos hibák megoldására, elhárítására alkalmas rendszer felépítését tűztük ki célul. Ehhez elkészítettük a távközlési (hibakezelő) szakontológia és a hozzá kapcsolódó hibakezelő rendszer specifikációját.

A rendszer modelljében a tipikus szerepek a következő ábra segítségével mutathatók be:



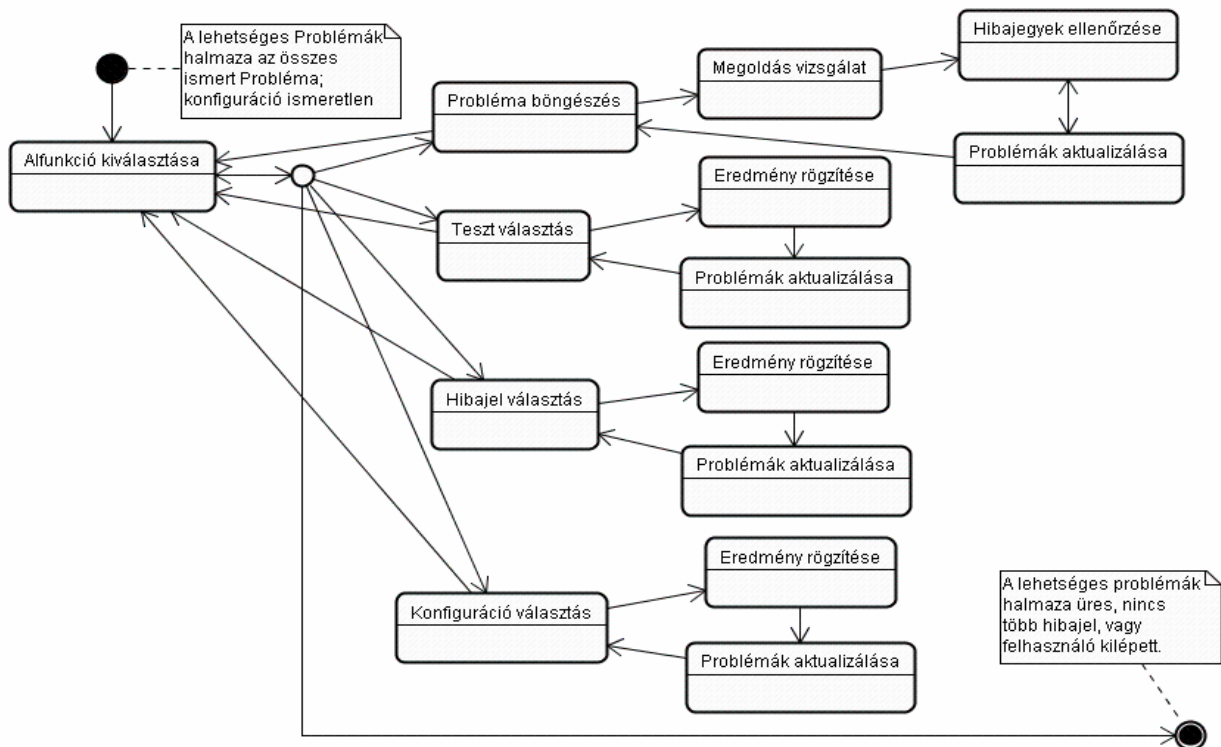
A rendszerfejlesztési feladat egyszerű: az *ügyfél* saját *problémája* megoldása érdekében segítséget kér a *CC munkatárstól*, aki igénybe veszi az ontológiával és következtető képességgel támogatott *rendszer*t. A rendszer az alábbi, nagyobb részekből áll:

- ontológiaszerkesztő,
- következtető rendszer,
- ontológia,
- segédalgoritmusok,

- felhasználói felület a CC munkatárs és a rendszer között.

Az alkalmazás használatának elképzelt forgatókönyve a következő:

- I. az ügyfél elmondja a CC munkatársnak, hogy milyen hibajelenségeket tapasztalt (ezeket a felhasználó átfordítja Hibajelekké) és valamennyit a saját rendszerének konfigurációjáról (melyekből Konfiguráció komponenseket készít a felhasználó)
 - II. a rendszer az információk alapján leszűkíti a Problémák lehetséges körét (első cél: a Probléma minél pontosabb felismerése)
 - III. ha a Probléma kellően meghatározott, létezik hozzá (ismert) Megoldás, a felhasználó kiválaszt egy ilyet, akkor a program kiajánlja a Megoldásokat (ha valamelyik Megoldást kipróbálta az ügyfél (természetesen a felhasználó fordításában és az ő javaslatára), akkor az eddig tapasztalt Hibajegyekre a IV. a.-nak megfelelően újra rákérdezőnk); egyébként folytassuk a IV. ponttal
 - IV. a program a további szűkítés céljából az alábbi dolgok valamelyikét teszi:
 - a. rákérdezhet egy olyan Hibajelre, ami legalább egy potenciális Probléma esetében felmerül, és még nem ismert, hogy fellépett-e a Hibajel. Amennyiben ez a Hibajel nincs, úgy az adott Problémákat ki lehet zárni; ha van, akkor „erősödik a gyanú”, hogy ezen Problémák valamelyikével állunk szemben.
 - b. kérheti egy bizonyos Teszt végrehajtását a felhasználótól; a tesztek sajátossága, hogy egy-egy Probléma vagy Problémaosztályra vonatkozóan sikeres végrehajtásuk egyértelműen megállapítja, hogy az adott Probléma(osztály) valamely eleme fellép-e vagy sem.
 - c. rákérdezhet a felhasználó az ügyfél Konfigurációjának valamely tulajdonságára, mely tulajdonság szükséges egy nem-kizárt probléma fennállásához.
 - V. Ezek után menjünk vissza a II. pontra.
- A főbb átmeneti lépéseket mutatja a következő ábra:



A felépítendő rendszer kulcseleme még a *Metatár*, melynek leírása megtalálható a projektjelentés egyik részdokumentumában.

Bővebben: [CC-spec]

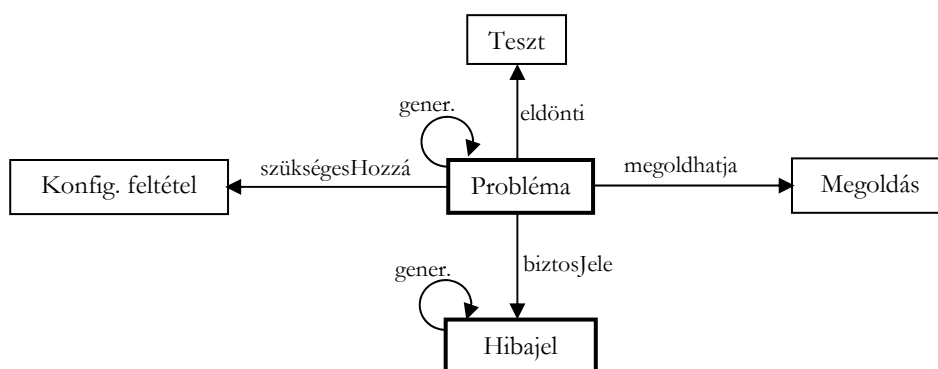
A közönségszolgálati tudásbázis

Az 1.4. részfeladatban az ügyfélszolgálati tevékenység támogatására már használatban levő adatbázisok, passzív informatikai komponensek újrahasznosításának lehetőségeit kellett megvizsgálunk. Kiderült, hogy az együttműködő ipari partnernél nem állnak rendelkezésre ilyen jellegű erőforrások, tehát a projektnek a semmiből kellett elkezdenie az építkezést. Ez a nehézség egy szempontból a projekt előnyére fordítható: nem kell igazodnunk a már meglévő rendszer használati gyakorlatából, rutinjaiból, elkötelezettségeiből fakadó kényszerekhez.

Az internethasználat során jelentkező hibák kezelésének folyamata kétfázisú: az első fázisban információgyűjtéssel és tesztek végrehajtásával leszűkítjük a lehetséges hibák halmazát egy elegendően kis méretre (azaz minél pontosabban behatároljuk a probléma okát). A második fázisban pedig olyan megoldásokat próbálunk ki, melyek képesek megszüntetni a lehetséges problémák valamelyikét. A gyakorlatban ez a két fázis összemosódik: időnként egy – pl. a tapasztalat szerint gyakori – lehetséges hibára még azelőtt kipróbálunk egy megoldási módot, hogy eléggé leszűkítettük volna a gócpontot; ilyen pl. az „otthoni” megoldás, miszerint „indítsuk újra a számítógépet”. Az információgyűjtés kétféle információ beszerzését takarja: az egyik fajta az ügyfél (aki problémát tapasztalt a rendszerében) rendszerének hardware- és software-kiépítésére vonatkozik (sokszor a problémák okai hardware- vagy software-inkompatibilitásra vezethetők vissza, pl. egy protokoll nem ismeretével). A másik információ – ami pedig a probléma detektálásához elengedhetetlen – a tapasztalt hibajelenségek (rendellenes működések) minél részletesebb leírása. Ezek alapján a rendszer elvárt funkcionálisai:

- be lehessen vinni az ügyfél rendszerének konfiguráció-leírását, valamilyen részletességgel, amit később pontosítani is lehetséges;
- be lehessen vinni a tapasztalt hibajeleket, melyeket később pontosítani, vagy esetleg (egy megoldás hatására) megszüntetni is lehetséges;
- a rendszer ezen inputok hatására tartsa nyilván a problémák egy osztályát, melyek kiváltó okai lehetnek a tapasztalt jelenség(ek)nek; ezt a problémaosztályt kérésre sorolja fel;
- kérésre javasoljon olyan teszteseteket, ellenőrzéseket, melyeket végrehajtva egyes potenciális hibák esetleg kizárhatóak lesznek;
- kérésre javasoljon olyan megoldásokat, melyek kijavíthatnak egy vagy több potenciális hibát.

A hibakezelési folyamat támogatásához felépítendő szakontológiának az egyik kulcskérdése, hogy milyen osztály-, egyed- és relációfogalmakat kell felvennünk? Ezt az alábbi ábrán mutatjuk be:



Az osztályként funkcionáló fogalmak vastagabb keretben szerepelnek. A Probléma és a Hibajel fogalmak *osztályok*; ezek kapcsolódnak Megoldás, Teszt és Feltétel *egyedek*hez. Az ügyfél oldalán pedig van egy adott konfiguráció, amire vagy *illeszkedik* egy konfigurációs feltétel, vagy nem; továbbá a felhasználói oldal konkrét hibajeleket és konkrét problémákat tartalmaz, melyek a megfelelő osztályok valamelyik alosztályának példányai, ám hogy egy-egy konkrét alosztálynak

példányai-e vagy sem, az a kérdés-válasz iteráció során módosulhat. A fogalmak leírását a projektjelentéshez mellékelte résztanulmányban adjuk meg: [CC-tb].

A rendszerben használandó **ontológianyelvnek** a következő funkcionalitásokat kell biztosítania:

- támogatni az osztály/egyed fogalmakat, többszörös öröklődést, példányosítást
- az osztályokat a generikus kapcsolaton kívül más kapcsolatba is hozhassa egyedekkel (így tudjuk biztosítani a problémaosztályokhoz a tesztek, feltételek ill. megoldások kapcsolását – pl. statikus adattagokon keresztül)
- sőt, osztályok között is kijelenthessünk benne tetszőleges kapcsolatot (így tudunk problémaosztályhoz hibajel-osztályt rendelni)

A **következtető rendszernek** a szükséges algoritmusok megvalósításához nem szükséges bonyolultnak lennie, azonban a képesnek kell lennie olyan kérdések megválaszolására, melyre a válasz egy felsorolás. Az ellentmondásosságot nem szükséges észrevennie, hisz a felhasználó – remélhetőleg – ezt elkerüli. Így a következtető rendszer működéséhez elegendő egyfajta mintaillesztő algoritmus futtatása egy jól meghatározott struktúrával rendelkező címkézett gráfban és a találatok listázása. Ez a valós idejű működés igénye miatt egy szerencsés (és nem véletlen) eredmény: egy ilyen algoritmus várhatóan gyorsan fut.

Bővebben: [CC-tb]

Ontológia-infrastruktúra építés

A projekt egyik kiemelt célkitűzése egy többretegű ontológia-építési módszertan megtervezése és kifejlesztése. Ennek érdekében egyrészt megvizsgáltuk azokat a következtető rendszereket, amelyek az ontológiák passzív tudásbázisainak aktív hasznosítására képesek. Másfelől folyamatosan gyűjtjük és útmutatókba foglaljuk össze az ontológiák építésére, az ontológiák konzisztenciájának kialakítására, fenntartására vonatkozó szabályokat, kényszereket. A tényleges ontológiaépítési munkák támogatására pedig a megfelelő tartalommenedzsment eszközöket keresünk és javaslunk.

A következtető rendszerek vizsgálata

A projekt 2.5. részfeladata a következtetési rendszerekre vonatkozó adatok összegyűjtése, illetve az elérhető és használhatónak tűnő rendszerek összehasonlító elemzése volt. A részfeladat teljesítése során összegyűjtöttük a következtetési tevékenységgel kapcsolatos legfontosabb logikai fogalmakat, melyeket a következő szakaszban beépítünk a projekt egységes wiki-szótárába. A projektszakasz legfontosabb feladata az következtető rendszerek összegyűjtése, tesztelése és összehasonlító lemezése volt.

A projektszakasz legfontosabb tanulsága az volt, hogy a következtetési rendszerek fejlesztésének még nagyon az elején van a világ, ami miatt inkább arra kell felkészülnünk, hogy a konkrét célfeladat nagyon konkrét igényeinek megfelelő, a szakterület sajátosságait messzemenően kihasználni képes, nagyon speciális következtető rendszert kell majd használnunk. Nyomban ebből az is következik, hogy a következtető rendszer végleges kiválasztása (esetleg kifejlesztése?) csak akkor lehetséges, amikor a tényleges fejlesztésben már magasabb szintre jutunk (vagyis amikor már alaposabban ismerjük a felhasználható ismeretbázis, szakterületi ontológia tényleges tartalmát).

Az elemzés legelső körében egy előre felállított, egységes, „elméleti“ szempontrendszer alapján kiszűrtük az előzetesen összegyűjtött következtető motorok egy jelentős részét. Ekkor még tényleges tesztelést nem végeztünk. Az alábbi következtető motorok „maradtak versenyben“:

- Carine (v 0.72)
- E (v 0.82)
- EPILOG (v EPILOG-2005-JUN-22)
- Gandalf (2.6 r1)

- JTP
- Otter (v 3.3f)
- SNePS, SNePSLOG (v 2.6.1)
- SPASS (v 2.1)

Az első szűrőn átment rendszereket lefuttattuk a teszttállományainkon. Minden terméket ugyanazon a szerveren futtattunk, így az eredmények objektíven összemérhetőek voltak. A **tesztkonfiguráció** az alábbi volt:

- CPU: Intel(R) Pentium(R) 4 CPU 2.80GHz, 1MB cache
- Memória: 1,5 GB
- OS: RedHat Linux - 2.4.21-4.EL
- Fordító: a C-ben készült, forráskód alakban terjesztett programokat gcc-vel (3.2.3) fordítottuk.

A felmérés eredményeként a következőképpen összegezhetjük. Az általános célú ontológia nyelvekhez kapcsolódó következő rendszerek (EPILOG, SNePS) kevésbé voltak sikeresek a válasz megkeresésében. Azonban az EPILOG több kiegészítő funkcióval rendelkezik, magasabb kifejezőerővel bír, mint a többi nyelv. Ezt a rendszert nem biztos, hogy objektíven ítélnénk meg elsőrendű logikai próbáink alapján (különösen, ha természetes nyelv feldolgozás az ontológia használat célja).

A tételbizonyítók egész jól szerepeltek, sokféle tesztet vettek sikeresen, viszonylag alacsony idővel. Kiemelkedő volt a vizsgált rendszerek közül az E teljesítménye az OWL-es tesztek esetében. Többnyire ez volt a leggyorsabb. Azonban a SPASS CNF-re alakító részét – a FLOTTER-t – is gyakran dicsérték. Úgy tűnik, hogy mind az általunk készített, mind a többi CNF-re alakító rendszerrel jobban teljesít. Érdemes lehet az E-t választani a SPASS FLOTTER részével együtt, bár valószínű, hogy az E is kellően hatékonyan rendelkezik.

A végső sorrend a következőnek tűnik: E, SPASS, Gandalf, Otter, JTP, Carine, SNePS.

Ez a sorrend nem teljesen valós, az első négy hasonló sebességűnek tűnik, bár az Otter kissé lemarad az első háromtól. Az EPILOG nem igazán azonos célú, az támogat modális kijelentéseket, valószínűségeket, időt, valamint a tudás egy jelentős részét hatékonyabban kezeli, ha a specialistái kapják meg, nem pedig elsőrendű formában kapja meg. Az utolsó három pedig az egyenlőség hiánya miatt nem ajánlott (persze ahol nincs szükség egyenlőségre, ott elegendő lehet a Carine, a másik kettő kevésbé megbízható, ám ekkor is érdemes lehet egyenlőséget támogató rendszert használni, mivel a Carine dokumentációja igen szegényes).

Bővebben: [kovrend], [kovrendfunkc], [kovrendtptp], [kovrendsajaj].

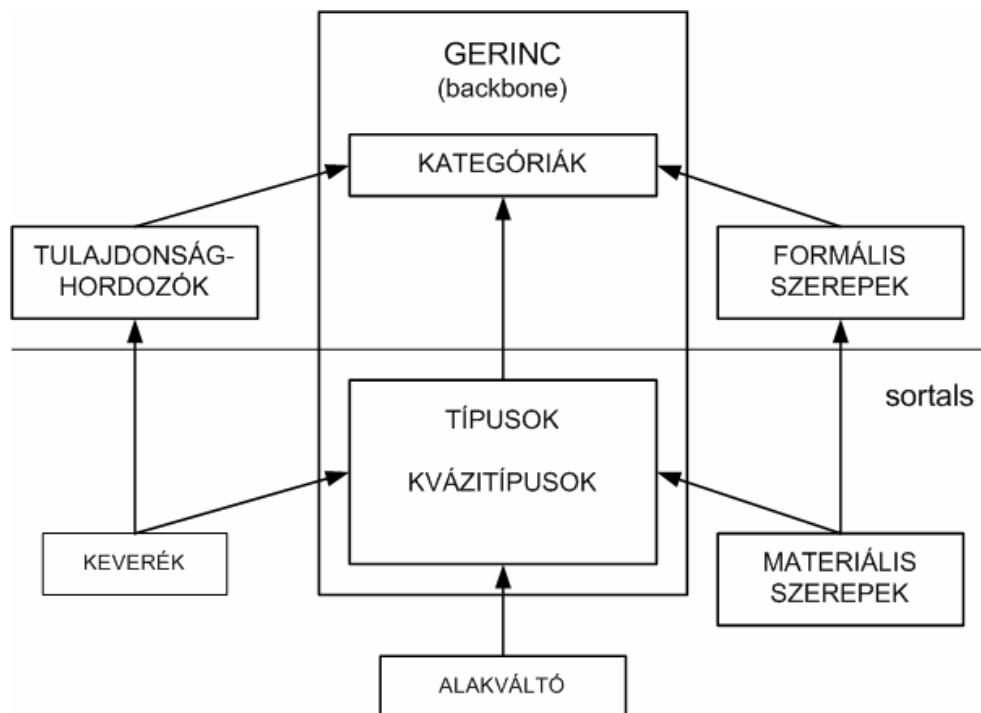
Az ontológiaépítési módszertan kérdései

A 2.8. projektszakaszban az ontológia-formalizmus kialakítását kellett megkezdenünk, illetve fejlesztői útmutatót kellett készítenünk. Ez részben az OntoClean módszertant ismertető tanulmány útmutatóvá formálásával, részben egy ontológiaszerkesztő alkalmazásra vonatkozó követelmény- és funkcionális specifikáció kidolgozásával valósítottuk meg.

Az OntoClean módszertan

A 2.6. részfeladat az OntoClean módszertan feldolgozását írta elő. Az első projektjelentésben foglalkoztunk a módszertan bemutatásával, ezért itt nem érezzük szükségesnek annak ismételt, részletes tárgyalását. Az OntoClean módszertan céljáról, összetevőjéről, elveiről, megoldásairól, szabályairól egy alapos összefoglaló tanulmány készült, mely már bevezető, útmutató anyagként is használható. A projekt második szakaszában sikerült mélyebben megérteni a módszertan filozófiáját, fogalomkészletét. Ez azért különösen fontos, mert az elkövetkező ontológiaépítési feladataink során nagy szükségünk lesz az OntoClean módszertan által kínált metafoglalmak pontos alkalmazására. A csúcsontológiába előzetesen besorolt fogalomhalmaz elemeire elkészült generikus struktúra és a csúcselemek közt megállapított egyéb relációkból fakadó külső

összefüggések, illetve a jelentésrepresentációs munka során megállapított belső összefüggések egymáshoz kapcsolása során jelentkező ellentmondások, illeszkedési, integrációs nehézségek feloldásában komoly segítséget remélhetünk a pontos metafogalom-használatnak. Utóbbi mutatja az alábbi ábra:



Bár az első projektjelentésünkben már foglalkoztunk az OntoClean szabályaival is, mégsem árt itt újra kitérni ezekre. Az OntoClean metatulajdonságok (rigiditás, azonosság, teljesség) jellemzi a fogalmakat, és a különböző tulajdonságértékek eltérő megoszlásai alapján különíti el a különféle fogalomtípusokat (lásd a fenti ábrát). A metatulajdonságok definíciójából következnek olyan összefüggések, amelyeknek állniuk kell az **ontológiákra** (ezek az OntoClean szabályok).

Legyen **P** és **Q** fogalom, amelyek közt fennáll a **generikus** reláció (**P Q**-nak alosztálya, **P Q**-nak fajtája, **P is-a Q**, **P is subsumed by Q**). Ekkor igaznak kell lennie az alábbiaknak.

- 1.) ha **Q** antirigid, **P** is antirigid,
- 2.) ha **Q** hordoz azonosságot, **P** is hordoz,
 - 2.1.) ha **Q** hordoz mereológiai azonosságot, **P** is hordoz,
- 3.) ha **Q** hordoz teljességkritériumot, **P** is hordoz,
 - 3.1.) ha **Q** hordoz topológiai teljességkritériumot, **P** is hordoz,
 - 3.2.) ha **Q** hordoz morfológiai teljességkritériumot, **P** is hordoz,
 - 3.3.) ha **Q** hordoz funkcionális teljességkritériumot, **P** is hordoz,
- 4.) ha **Q** hordoz anti-teljességet, **P** is hordoz,
- 5.) ha **Q** függő, **P** is az,
- 6.) ha két fogalom IC-je vagy UC-je összeegyeztethetetlen, a két fogalom diszjunkt.

A következő szabályok egész általános ontológiai elkötelezettségek:

- 7.) Minden egyed tartozik egy olyan fogalomhoz, amely hordoz IC-t.
- 8.) Ha két egyed megegyezik, van egy olyan fogalom, amely mindkettő azonossági feltételét hordozza.

A szabályokat értékelhetjük is. Az öröklődésre vonatkozó szabályok nyilvánvalóan következnek a definícióból, a 7.) és 8.) elég nyilvánvalóak; a 7.) alapvető feltétel a fogalmak osztályozásához. Azonban a 6.) szabály értelmetlen, ugyanis nem tudni, mit jelent az, hogy két kritérium összeegyeztethetetlen. Vannak, akik az összeegyeztethetőséget a két azonossági feltétel

ekvivalenciájával definiálják. Ez az ekvivalencia azonban logikailag csak akkor látható be, ha a fogalmak részletes axiomatizálása rendelkezésre áll. A különböző azonossági feltételek ekvivalenciának kérdése akkor merül fel, ha teljesen különböző ismertetőjelekre vonatkoznak (pl. az ember DNS struktúrája és a tér-idő kontinuumban elfoglalt helye egyaránt az **ember** fogalom azonossági feltétele lehet). Azonban nincs is szükség a definícióra. A 6.) szabályt így kell érteni: *ha egy A fogalomnak az azonossági feltétele (vagy teljességkritériuma) nem alkalmazható B előfordulásaira, a két fogalom diszjunkt.*

Az OntoClean módszertan értelmezését megpróbáltuk azzal is segíteni, hogy esettanulmányokat készítettünk a legkülönbözőbb fogalmak kapcsán, melyekben a vizsgált fogalmak leírása mellett szövegesen is kommentáltuk a problémát. Az alábbi tárgyköröket vizsgáltuk meg:

- elmélkedés az anyagról és az anyagöszeletről
- mi az építőanyag? eledel-e a rántott csirkecomb?
- vajon a „társadalmi egység” (social entity) emberek egy csoportja-e?
- hogyan nemzet az angol, hogyan faj az ember?
- az ember ágens-e, ember-e a tanár és mi a kalapács?
- a lepke és a hernyó esete
- mitől (kitől) függ a férj?
- munkanapok és ünnepek, azaz az időfogalmak

Amenyiben a projekt további menetében még hasonló anyagok készülnek, akkor érdemes lesz majd azokból külön „oktató anyagot” létrehozni. Egyelőre a módszertani ismertető anyag mellékletében mutatjuk be ezeket az esettanulmányokat.

A módszertan feldolgozása során természetesen rengeteg olyan problémával szembesültünk, amelyek jelenleg még nincsenek (jól) megoldva az OntoClean módszertanon belül sem, ezért a projekt számára továbbfejlesztési célokat is kitűztünk (amiről a jelentés másik fejezetében adunk számot).

Az ontológiaszerkesztők értékelésekor kiderült az is, hogy a projekt kiválasztott szerkesztőjébe, a Protégébe már implementálták az OntoClean módszertan szabályait, kényszereit, így ezzel nekünk már nem kellett külön foglalkoznunk.

Bővebben: [OntoClean]

A konzisztenciavizsgálat szempontjai

A 2.7. szakaszban a konzisztencia-menedzsment képesség kialakítását kellett elvégeznünk. A feladat megvalósítását két részre osztottuk. Egyrészt az ontológiai vázszerkezet konzisztens felépítésének biztosítékaként az OntoClean módszertan előírásainak betartását jelöltük meg. Ezt a projektjelentés előző fejezetében bemutattuk. Másrészt további konzisztenciafeltételeket, konzisztenciakényszereket kezdtünk el összegyűjteni, leírni az ontológiaszerkesztőkre vonatkozó követelmény- és funkcionális specifikációban, melyet a jelentés későbbi részén ismertetünk.

Módszertani továbbfejlesztési lehetőségek

A konzisztenciavizsgálatok, a módszertani elvek kidolgozása területén több irányban is látunk előrelépési lehetőségeket, hiszen magát az OntoClean módszertant is továbbfejleszthetőknek érezzük, de az ontológiák alaprelációira (a generikus és partitív relációkra) is további vizsgálatokat, módszertani fejlesztéseket látunk megvalósíthatónak, illetve megvalósítandónak.

Az OntoClean módszertan továbbfejlesztése

Megállapíthatjuk, hogy az OntoClean az egyetlen olyan metodológia, amely formális eszközökkel közelíti meg az ontológiák alapvető problémáit. Azonban alapos elemzése számos problémát

vetett fel, és előnyei újabb lehetőségeket kínálnak fel. Ezért célul tűzzük ki továbbfejlesztését, melyet az alábbi irányokban tartunk elképzelhetőnek:

- módszereket és eszközöket kidolgozni a metatulajdonságok megállapításának segítésére,
- a fogalmak osztályozásának finomítása és pontosítása (attribution, keverék és phased sortal fogalomtípusok esetén),
- meg kell vizsgálni, hogyan alkalmazzuk a metodológia fogalmait a relációfogalmakra,
- ki kell dolgozni, hogy fogalomtípusokként hogyan lehet szabványosított relációk segítségével a fogalmakat alaposabban jellemezni (a szerep esetére ez megkezdődött

Bővebben: [OntoClean], [Szóts & Lévay 2005]

A partitív reláció típusainak és tranzitivitási feltételeinek kidolgozása

A *partitív reláció* tipizálásával sokat foglalkoztak a különböző szaktudományok területein. Ezek elsődleges feldolgozását elvégeztük. További munkára van szükség egy olyan szempontrendszer kidolgozásához, amely módszertani segítséget ad a partitív reláció típusbasorolásához. A tipizálási munka ilyen támogatása azért különösen fontos, mert a partitív reláció esetében nem mindig érvényes a tranzitivitás tulajdonsága. Márpedig ez a tulajdonság az öröklődést kihasználó és másfajta következtetések alapja. Ezért minden ontológia számára alapvető, hogy meg tudjuk mondani, mikor teljesül a tranzitivitás és mikor nem? Ehhez pedig a partitív reláció megfelelő típusainak kialakítására és használatára van szükség.

Bővebben: [Szakadát 2005]

A generikus reláció elméletének kialakítása a mereológiai elméletek mintájára

A projekt során feldolgoztuk, hogyan lehet axiomatizálni a *partitív reláció* fogalmát a mereológiai elméletek egymásra épülő rendszerében. Elkezdtük a generikus reláció formális leírását is, de ezzel a kísérlettel csak az első lépésig jutotunk, amit érdemes lenne folytatni. A generikus reláció alapját ugyanúgy egy rendezési reláció jelenti, mint ahogy ez igaz a partitív reláció esetében is. A partitív relációt formalizáló mereológia elmélete egyre több axiómát vesz fel, amivel egyre szűkíti az eredeti rendezési reláció jelentését, vagyis egyre pontosabban írja le a partitív reláció jelentését. Hasonlót kellene (lehetne) elvégezni a generikus reláció esetében is, hiszen az alapvető rendezési relációhoz ugyanez szükséges további axiómákat felvennünk, hogy a generikus relációt képesek legyünk elkülöníteni egyrészt más rendezési fogalmaktól, másrészt a generikus relációhoz hasonló jelentésű, de bizonyos pontokon mégis eltérő tartalmú (pl. hiponíma, is-a) relációktól.

Bővebben: [Szakadát 2005]

Jelentésreprezentáció

A projekt a jelentésreprezentáció területén tovább lépett a korábban lefektett elvekhez képest azzal, hogy elfogadta Jerry Hobbs megoldását, aki a mindennapi tudás formalizálását a davidsoni megközelítés radikális kiterjesztésével éri el.

Hobbs reprezentációs nyelvének sajátosságai

Hobbs egy nagyon fontos ponton általánosítja Davidsont. Egy „komprehenziós séma” segítségével posztulálja, hogy bármely n -argumentumú predikátumhoz tartozik egy $n+1$ -argumentumú predikátum, amelynek van egy kitüntetett eseményargumentuma (ezt alább több példán be fogjuk mutatni). Ennek az általánosításnak azonban közvetlen következménye, hogy az eseményváltozó már nem csak a davidsoni értelemben vett eseményekre, hanem tetszőleges *tényállásra* vonatkozhat. Más szóval Hobbs tetszőleges (ténylegesen fennálló vagy pusztán lehetséges) tények fölött kvantifikál és azt, hogy egy tény *valójában* is fennáll, egy sajátos predikátummal, a **Rexist(e)** („really exists”) predikátummal fejezi ki. Egy **p(a)** mondat igazsága

így ekvivalens módon kifejezhető lesz a $\exists e(p'(e,a)\wedge\text{Rexist}(e))$ kijelentéssel, ahol a vesszős predikátum első argumentuma egy tényekre utaló változó. Általános esetben tehát:

$$\forall x_1, x_2, \dots, x_n (p(x_1, x_2, \dots, x_n) \equiv \exists e(\text{Rexist}(e)\wedge p'(e, x_1, x_2, \dots, x_n)))$$

Hobbs magyarázata szerint a felsővessző mint operátor egy olyan predikátumot állít elő a kiinduló predikátumból, amelynek első argumentuma olyan körülményeket (*conditions*) tud denotálni, amelyek mellett a $p(x_1, x_2, \dots, x_n)$ kijelentés igaz. Egzisztenciálisan kvantifikálva ezt a változót olyan formulát kapunk, amely azt állítja, hogy *van* olyan körülmény, amelyek között a $p(x_1, x_2, \dots, x_n)$ kijelentés igaz; ezt azután kiegészítve a **Rexist** predikátummal a kapott formula azt állítja, hogy a $p(x_1, x_2, \dots, x_n)$ kijelentés *ténylegesen* (aktuálisan) is igaz.

Az analitikus filozófiában az utóbbi években egyre többször kerül terítékre az ún. *truth-makers* elmélete (ezt bevett terminológia híján esetleg „az igazságalapok elméletnek” nevezhetjük). Eszerint az elmélet szerint az igazság fogalmával kapcsolatba hozható dolgok két osztályba sorolhatók: igazsághordozók (*truth-bearers*): ezek azok a dolgok, amik egyáltalán igazak vagy hamisak lehetnek (alapvetően a mondatok és a kijelentések), illetve az igazságalapok (*truth-makers*): ezek azok a dolgok, amelyek igazzá vagy hamissá teszik az igazsághordozókat. Az elfogadott definíció szerint egy igazsághordozó igazságalapja az, aminek létezése kikényszeríti („necessitates”) az igazsághordozó igazzá válását. Davidson fent említett elmélete az igazságalap-elmélet egy sajátos alkalmazásának tekinthető, ahol a kijelentések (Davidson cikkében a cselekvésleíró mondatok) mint igazsághordozók igazságalapjául intencionális cselekvési események szolgálnak. Hobbs megközelítése ezzel szemben az igazságalap-elmélet korlátlan elfogadására épül. Illusztráljuk most ezt a megközelítést Hobbs egy példájával. A

Maybe the boy wanted to build a boat.

mondat „hagyományos” (Montagoviánus) reprezentációja Hobbs szerint így nézne ki:

$$\exists x(\text{boy}(x)\wedge\text{POSS}(\text{PAST}(\text{WANT}(x,\lambda z(\exists y(\text{boat}(y)\wedge\text{Quick}(\text{build}(z,y))))))))$$

míg a reifikáció alkalmazásán keresztül a fentiek a következőkre egyszerűsödnek:

$$\exists e_0, e_1, e_2, e_3, e_4, x, y (\text{Rexist}(e_0) \wedge \text{Past}'(e_0, e_1) \wedge \text{possible}'(e_1, e_2) \wedge \text{want}'(e_2, x, e_3) \wedge \text{quick}'(e_3, e_4) \wedge \text{build}(e_4, x, y) \wedge \text{boy}(x)\wedge \text{boat}(y))$$

Az egyszerűsödés az operátortípusok szintaktikai „kilapulásában” nyilvánul meg.

Thematisz szerepek

A thematisz szerep (vagy théta-szerep) fogalma a nyelvtudományból származik, ahol abból a megfigyelésből kiindulva, hogy az egyes igék argumentumaikra tipikus, ám rendkívül általános feltételeket rónak ki, a nyelvészek körülhatároltak néhány ilyen típust. A thematisz szerep (Θ -szerep) fogalma azonban igen általánosan is definiálható. Voltaképpen tetszőleges predikátum (nem csak ige) esetében is van értelme. Általános esetben egy Θ thematisz szerep azonos argumentumszámú predikátumok valamelyik argumentumhelyét jellemzi. Így a thematisz szerepek két tényezőtől függenek: n -argumentumú predikátumok (pl. igék) egy P halmazának megválasztásától (n rögzített), illetve attól, hogy pontosan *melyik* argumentumhelyet vizsgáljuk. A legáltalánosabb szinten tehát nem beszélhetünk thematisz szerepről *per se*, csak egy *adott predikátumhalmazban* a predikátumok i . argumentumhelyét jellemző thematisz szerepről.

Általánosan azt mondhatjuk, hogy adott P n -argumentumú igékből (n -argumentumú predikátumokból) álló ige-halmaz (predikátum-halmaz) elemeinek i . argumentumhelyét jellemző thematisz szerep nem más, mint azon tulajdonságok (1 -argumentumú predikátumok) összessége, amelyeket a halmazban található igék (predikátumok) az i . argumentumukra vonatkozóan mind megkövetelnek:

$$\Theta(x_i, P) =_{\text{def}} \bigcap_{p \in P} \{ \{ \Phi \mid \forall x_1 \forall x_2 \dots \forall x_n (p(x_1, x_2, \dots, x_i, \dots, x_n) \rightarrow \Phi(x_i)) \} \}$$

Ebből a szokásos théta-szerep fogalomhoz úgy jutunk, hogy a $\Theta(x_i, P)$ halmaz elemei közül csak a legspecifikusabbakat tartjuk meg (a „legkisebbeket” a tulajdonságok (predikátumok) implikációs (generikus) rendezése szerint).

A nyelvészek pragmatikus alapon csoportosították az igéket, miközben párhuzamosan kialakították a csoportokhoz tartozó théta-szerepeket is. Kialakították pl. az ún. *ágens-igék* osztályát, amiben olyan igéket szerepeltetnek, amelyek (felszínen alanyként megjelenő) argumentumára *nagyjából* a következő érvényes:

{aktív-az-eseményben, felelős-kezdeményezője-az-eseménynek}

Ha tehát egy ige teljesíti azt a feltételt, hogy bármely vele képzett igaz mondatban az alanyi argumentumára igazak e halmaz elemei, akkor a szóbanforgó ige eleme az ágens-igék osztályának.

A nyelvészek által használt tematikus szerepek főbb típusai a következők.

- ágens (AG):
tudatos, a cselekvésben aktív, felelős érte, és ő kezdeményezi (esetleg folyamatosan kontrollálja) azt (*Brutus a Brutus megölte Cézárt* mondatban)
- páciens (PAT):
egy ágens által kifejtett hatás, változás elszenvedője; az esemény során tulajdonságai az ige jelentéséből megjósolható módon megváltoznak (*Cézár a Brutus megölte Cézárt* mondatban)
- eszköz (INSTR):
a cselekvés közvetlen eszköze; jelenléte feltételezi egy ágens és egy páciens jelenlétét; az ágens az eszköz közvetlen manipulálásával gyakorol hatást a cselekvés páciensére (*a tőr a Brutus egy tőrrel ledöfte Cézárt* mondatban)
- experiens (EXP):
tudatos szereplő; az ige az ő tudatállapotáról illetve annak valamilyen megváltozásáról tesz állítást (*János fél a pókoktól, Mari megjedte a kutyától*)
- cél (GOAL):
Ágens által végzett tevékenység szándékolt végállapota ill. végpontja (János sokat tanul, hogy sikeresen le tudjon vizsgálni, Mari kihordta a bútorokat az udvarra)
- téma:
aminek a helyét vagy tulajdonságait adjuk meg; ez a szemantikailag legkevésbé tartalmaz tematikus reláció (*A malac az ólban van, A könyv borítója piros*)

A tematikus szerep fogalma a jelen vállalkozás számára két szempontból is lényeges. Egyrészt azért, mert a többargumentumú predikátumkonstansok argumentumai között nemigen lehet máshogy különbséget tenni (a sorrend nyilván konvencionális, ám a konvenció alapjául többnyire éppen a tematikus szerepek szolgálnak). Másrészt azért, mert a reifikációval együttesen használva lehetővé teszik a többargumentumú relációk kétargumentumúra redukálását. Terence Parsons ugyanis módszeresen továbbfejleszti Davidson eredeti elméletét egy olyan irányban, amelyben az eseményben résztvevő szereplőket egy-egy tematikus szerep köti magához az eseményhez. Vegyük példaként a következő mondatot:

Brutus egy tőrrel ledöfte Cézárt.

A mondat parsonsi ábrázolása így néz ki:

$$\exists e (\text{ledöfés}(e) \wedge \text{AG}(\text{brutus}, e) \wedge \text{PAT}(\text{cézár}, e) \wedge \text{INSTR}(\text{tőr}, e))$$

Mikroelméletek

A *mikroelmélet* fogalma a CYC ontológiában széleskörűen használatos fogalom. A mikroelméletek olyan tudásterületek (egy csoportba tartozó állítások), amelyek osztoznak bizonyos általánosnak tekinthető axiómákban. A mikroelmélet fogalma a jelentéslírások esetében is alapvető, mert ezen keresztül ragadható meg az azonos „jelentésmezőkhöz” tartozó kifejezésekben lévő közös

tartalom. Az egy szemantikai mezőhöz tartozó mikroelméletet jellemző axiómák meghatározása azért fontos, mert segítségükkel a mezőhöz tartozó egyes szemémák jelentésének leírása „parametrizálhatóvá válik”, így a konkrét jelentések leírása már „csak” az egyes paraméterek értékének meghatározásán múlik. A mikroelméletek részletes kidolgozása és a köztük lévő kapcsolatok megszervezése halaszthatatlan feladat, hiszen a példaanyag felgyűléseivel párhuzamosan az egyes állítások egymáshoz való viszonyait egyre inkább átlátható formára kell hozni. Ennek eszköze az absztrakció, amelynek segítségével az egyes jelentésrepresentációkban ismétlődő minták időről időre kigyűjthetők, és a rendszer magasabb pontjaihoz rendelhetők, ahonnan öröklés útján „jutnak el” a konkrét jelentésleírásokhoz.

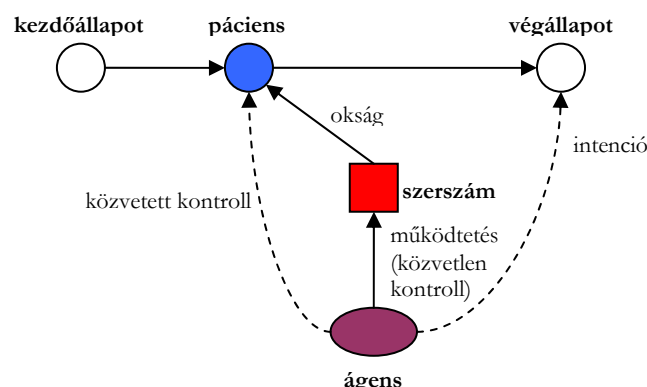
Bővebben: [jelrepyelv]

Prototipikus példák

A fentebb kifejtett elméleti megfontolásokra, szempontokra támaszkodva a projekt második szakaszában olyan fogalmak formalizálását végeztük el, amelyek „önmagukon túlmutató jelentőséggel” bírtak azért, hogy tipikusnak mondható problémákat jelentettek az axiomatizálási munka során. Ezek a definíciós megoldás jelentősen felgyorsíthatják a további fogalmak formális leírását célzó munkát. Egy önálló tanulmányban bemutatjuk az alábbi fogalmak formalizálását:

- háziállatok, házityúk, házinyúl, házisertés, házimacska
- természeti jelenségek, csapadék
- intézmények, intézmények épülete, kórház
- foglalkozások, bűvész
- szerszám, munkaeszköz

A formalizálási feladat, illetve a felmerülő problémák érzékeltetésére kiemeljük az eszköz példáját. Ebben az esetben egyértelműen látszik, hogy nem pusztán a szerszám, illetve (munka-)eszköz fogalmát jellemezzük, hanem azt a környező tudásdarabot is, amit ez a fogalom implikál. Valaki csak akkor rendelkezik egy adott eszköz fogalmával, ha tudja, milyen körülmények között, azaz mikor lenne szüksége a szóban forgó eszközre. Mivel ez minden eszközre igaz, az eszköz fogalmához hozzátartozik ez a feltétel is. A fogalmat két szinten írhatjuk le. Az egyik a sematikus összetevő, ami minden eszközre igaz. A másik a speciális (és tipikus) rész, ami eszközről eszközre változhat. A sematikus rész bizonyos slot-jait töltik ki a konkrét eszközökhöz rendelt specifikus információk. A sematikus összetevőhöz az alábbi ábrát jellemezhetjük:



Természetes nyelvi értelmezése: 'az Ágens az eszköz használatával valamilyen átmenetet akar megvalósítani (problémahelyzet)'. A specifikus összetevő vizsgálatokor mindig két alapkérdésre kell tudnunk válaszolni:

- tipikusan mire jó?
- tipikusan hogyan használjuk?

A bemutatott értelmezési keretben már minden eszköz fogalomra adható megfelelő formális leírás, melyek közül néhányat a hivatkozott tanulmányban mutatunk be.

Bővebben: [jelreppeldak]

Az ontológiaszerkesztés gyakorlata

A MEO projektben kiemelt hangsúllyal kezeljük az ontológiainfrastruktúra kérdéskörét, mert minden ontológiát használó projektben arra kell felkészülni, hogy az ontológiák felépítéséhez, módosításához, szűkítéséhez, bővítéséhez szükség van arra, hogy az adott tudásterülethez értő – informatikailag esetleg/gyakran laikus – szakemberek munkáját támogatni lehessen valamilyen tartalomnedzsment alkalmazás igénybevételével.

A projekt első szakaszában a közel 100 ontológiaszerkesztő közül kellett választanunk. A vizsgálatok után a *Protégé* mellett döntöttünk, és ma is ez tűnik a legmegfelelőbb választásnak. Bizonyos funkciók esetében azonban maradtak fenntartásaink, ami miatt a rendszer teljes körű alkalmazhatósága még nem véglegesen eldöntött. A „hagyományos“ ontológiaszerkesztők közül kilógott a Relex, ezért nem is lehetett jól összehasonlítani a többi eszközzel. A Relex egy nagy múlttal rendelkező, hazai teauruszszerkesztő program, amely a rendszer elemei közti külső kapcsolatokat felhasználóbarát módon engedi kezelni. Megvizsgáltuk, hogy a *Protégé* és a Relex milyen funkcionalitásokban erősek, illetve miben vannak hiányosságai. Az összehasonlító végén arra jutottunk, hogy továbbra is megtartjuk a *Protégé* „elsődlegességét”, de elképzelhető, hogy az ontológiainfrastruktúra felépítésével kapcsolatos követelményspecifikáció kidolgozása után szükség lehet arra, hogy olyan ontológiaszerkesztőt is kifejlesszünk, amely beépíti magába a Relex jó tulajdonságait, hálózati üzemmódban képes működni és adatbázisból tud közvetlenül dolgozni.

A *Protégé* a gyakorlatban

A második projektszakaszban már úgy vizsgáltuk a *Protégé* programot, hogy mennyiben és hogyan alkalmas az ontológiaszerkesztési feladatok elvégzésére, és elkezdtünk egy olyan anyag összeállítását is, amely egyfajta használati útmutatóként is kezelhető.

Ebben az anyagban összefoglaljuk, hogyan lehet felépíteni egy új ontológiát, különös tekintettel arra, hogyan lehet az olyan alapvető elemeket létrehozni, mint az osztályhierarchia, az attribútumok (slotok) vagy a példányok. Bemutatjuk azt is, milyen a *Protégé* 3.1.1 többfelhasználós módja. A *Protégé* program ugyanis támogatja ontológiák több személy általi, párhuzamos elérését, mely során a felhasználók ugyanazt az ontológia-adatbázist szerkesztik. A hálózati kommunikációt, valamint a szinkronizációs problémákat a szerkesztő program megoldja. Fontos képessége az alkalmazásnak az ontológiák konzisztenciájának ellenőrzése: a *Facet Constraints Tab* plug-in segítségével megkereshetjük, hogy az ontológiában alkalmazott megszorítások valahol sérülnek-e. Az emberi feldolgozás számára olykor komoly segítséget nyújthat a *Protégé* grafikus megjelenítési képessége (a *TGVizTab* plug-in segítségével), melyet az útmutatónkban szintén bemutatunk. A *Leíró logikát* alkalmazó ontológiák könnyen szerkeszthetők a *Protégé* OWL Plug-in segítségével. Ez különösen fontos a MEO-projekt számára, ezért az anyagban alaposabban körbejárjuk a *Logikai kifejezések* és *Osztályleírások* szerkesztésével kapcsolatos feladatokat, illetve megvizsgáljuk a *Következtetés* és a *Méretezhetőség* problémáit.

A *Protégé* lehetőséget ad a segítségével készült, leíró logikát alkalmazó ontológiák ellenőrzésére is úgy, hogy külső, leíró logikai következtető gépeket tud felhasználni. A tesztelési munkák során elvégeztünk néhány kísérletet pár ismertebb leíró logikai következtető motor és a *Protégé* két változatának együttműködésére vonatkozóan. A felmérés eredményét az alábbi táblában összegezzük:

Név	Protégé 3.0		Protégé 3.1.1	
	Kapcsolódási felület	Működött?	Kapcsolódási felület	Működött?
Pellet	szoftver/HTTP	nem	szoftver/HTTP	nem
Fact	HTTP DIG	nem	HTTP DIG	nem
Fact++	HTTP DIG	nem	HTTP DIG	igen
Racer	HTTP DIG	igen	HTTP DIG	igen

a *Protégé* külső, leíró logikai következtető gépei

A teszteléshez a `pizza.owl` közismert példa ontológiát használtuk fel. A Pellet rendszernek csak a legújabb béta verziója támogatja a HTTP DIG felületet, amely azonban még csak béta verzióban volt meg és nem működött egyik *Protégé* verzióval sem. Ugyanez volt a helyzet a Fact következtető géppel is. A Fact++ azonban már működött a *Protégé* 3.1.1-gyel. A letöltött Racer következtető gép pedig mind a két *Protégé* verzióval jól működött. Egyre nagyobb méretű ontológiák esetén azonban a Fact++ is elkezdett akadozni, és használhatatlanná vált.

Bővebben: [Protege]

Követelmény- és funkcionális specifikáció

Belekezdünk egy olyan dokumentum kidolgozásába, amely az ontológiaépítés érvényességi feltételeivel, ellenőrzési lehetőségeivel kapcsolatos kérdésekkel foglalkozik. Ez az dokumentum az ontológiaszerkesztési tevékenységre vonatkozó követelmény- és funkcionális specifikációként is értelmezhető.

Bővebben: [kovspec]

Ontológiaépítés

A projektterv szerint csak a 3.2. részfeladatról kellene beszámolnunk ebben a jelentésben, de röviden jelzést adunk más eseményekről is.

MEO-csúcsontológia

A projekt első szakaszában összeállítottuk a MEO csúcsontológiájába tartozó elemek gyűjteményét. A második projektszakaszban elkezdtük az elemek közti struktúra kialakítását. Elkészült az anyag, a tárgy és nagyrészt a tevékenység fogalmi kategóriák generikus hierarchiája. Ezen belül minden fogalmi egység tezaszuscikke is megszerkesztődött. Ezen kívül számos, a generikus hierarchiába még nem beillesztett tulajdonság és állapot kategóriába tartozó fogalom tezaszuscikke kialakult, de jelenleg még nincs teljesen lezárva. A már az ontológiai hierarchiába szerkesztett fogalmak tezaszuscikkei is bővílni fognak még azáltal, hogy a még beépületlen állapot- és tulajdonság kategóriába tartozó fogalmak beépülnek a saját kategóriájuk generikus hierarchiájába.

Bővebben: [MEO-gener], [MEO-top]

Ontológiamodell

Az 2005 elején javasolt első változatához képest az ontológia adatbázisának modelljét több ponton is érdemi módon bővíteni, pontosítani kellett. A modell első változatában szereplő 'gloss', 'sense' és 'concept' alapentitások megmaradtak, bár két helyen is nevet cseréltünk. Egyfelől a 'gloss' entitást nevet – a ráakodó jelentésrétegek miatt – megváltoztattuk, s a modell legelemibb nyelvi egységét 'graph'-ként kezeljük, másfelől a jelentés (sense) szintjén a 'sense' táblát átkereszteltük 'lexeme'-nek. Az igaz változást azonban mégis egy új szint beiktatása hozta. A morfoszintaktikai jegyek felvétele céljából fel kellett vennünk egy teljesen új entitást, amit 'morph'-nak nevezünk el. A 'morph' egységhez tudjuk hozzá rendelni a jegyeket (többek között az egység szófaji besorolását), és így ezen a szinten jellemezni tudjuk az elemek morfoszintaktikai viselkedését. Abból a célból, hogy a 'morph'-ok segítségével a 'lexeme' tábla elemeit felépíthessük, vagyis tartalommal/jelentéssel látjuk el a szóalakokat/formákat, két fontos modellezési kiegészítést kellett tennünk. Egyfelől meg kellett oldanunk a morfológiai szavak kezelésének lehetőségét, másfelől a 'morph'-okra támaszkodva biztosítanunk kellett a kifejezések felépítésének lehetőségét is. A 'sense' szintre tehát (aminek alapegységét 'lexeme'-nek nevezzük, mint ahogy lehetne 'semene'-nek is) vagy szó vagy kifejezés mehet be.

A modell fontos eleme még az, hogy a szavakhoz, terminusokhoz rendelendő „metaadatok” tartalmát, értékét szintén a rendszerből kell kivenni (ha már egyszer benne van, benne lehet minden szó és kifejezés, illetve a rendszerhez lehet csatolni a nyelvészeti gold-ontológia elemeit is), amit megfelelő nézetek előállításával lehet biztosítani (pl. a szófaj-, nyelv- vagy relációlisták).

Bővebben: [ontomodell], [gold]

MEO-szakontológia

A 3.2. részfeladatban a távközlési teaurusz felújítása és transzformálása volt a feladat. Elkészült a távközlési teaurusz felső és középső szintű fogalmainak generikus hierarchiája, azaz a teljes „anyagok”, „eszközök (alkatrészek és berendezések)”, „folyamatok és műveletek” és „jellemzők” kategóriába tartozó fogalmak struktúrája. Megtörtént a legspeciálisabb kifejezések szelektálása, rendezése, egy részük már kapcsolódik a középszintű fogalmak kifejezéseikhez. A szükséges transzformálási munka azáltal valósult meg, hogy a távközlési teaurusz technikai formátumát is úgy alakítottuk ki, hogy azt a Relex teauruszszerkesztővel lehet kezelni.

A projektervekben nem szerepelt, de a második projektszakaszban adatbázisba rendeztünk egy szabad forrású, Gold néven futó, nyelvészeti ontológiát. Ennek elemeit bármikor be lehet emelni a MEO-csúcsontológiába exoszteletális elemekként.

Bővebben: [telcotez], [telcotezhier], [gold]

SUMO-csúcsontológia

A 4.1. munkaszakaszban a szabadon felhasználható csúcsontológiák esetleges hasznosítását kellett előkészítenünk. Két nagy ontológiára (a WordNet és a SUMO-ra) hivatkoznak világszerte és ezeket egyre több célra kezdik el használni. Bár a projekt tervezési fázisában felmerült, hogy próbáljuk meg a WordNet valamilyen hasznosítását, de ezt elvetettük. A SUMO ontológiát viszont – legalább részeiben – alkalmazhatónak véljük. Éppen ezért az idő problémakörét kiválasztva megvizsgáltuk, hogy ezen a területen milyen a SUMO és MEO elemeinek egymáshoz való viszonya – az illeszkedés szempontjából.

A SUMO mintegy 4 000 állításból áll, köztük 800 szabályt (rule, implication, „ha ... akkor ...” típusú implikációt tartalmazó speciális állítás) és kb. 1 000 fogalmat definiál. A SUMO egységei a fogalmak (concepts), melyeket a terminusok (terms) jelölnek. A fogalmak négyféle alaptípusba (metafogalomba) sorolhatók: **Class** (osztály), **Relation** (reláció), **Individual** (individuum), **Attribute** (attribútum / tulajdonság), melyek között négyféle reláció (metareláció) vehető fel: **Instance** (példánya), **subClass** (részosztálya), **subRelation** (reláció-altípus), **subAttribute** („Attribútum-altípusa).

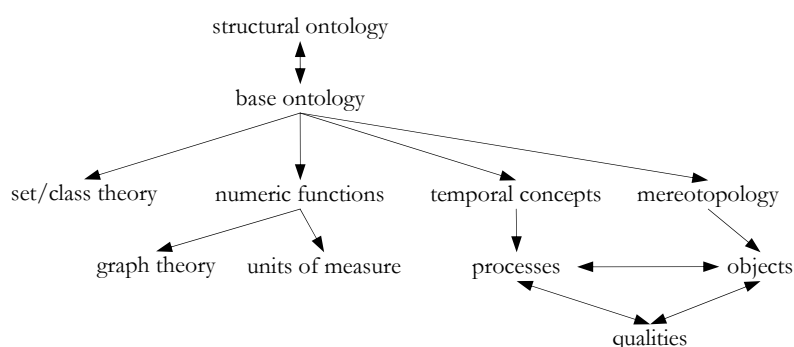
A SUMO moduláris ontológia, vagyis több, egymásra épülő részontológiából (subontologies) áll.

A 11 részontológia az alábbi előzmények felhasználásával született:

- **STRUCTURAL ONTOLOGY**: hasznos szintaktikai rövidítések definíciói (alapvető strukturális relációk, logikai, aritmetikai műveletek stb.) 38 fogalom, 220 állítás.
- **BASE ONTOLOGY**: a teljes ontológia legfelső szintje, az alapvető fogalmi kategóriák. Források: Sowa (1995), Russell & Norvig, ITBM-CNR felsőszintű ontológia. 223 fogalom, 972 állítás.
- **SET/CLASS THEORY**: halmazelméleti és osztályokkal kapcsolatos predikátumok és függvények. Forrás: kif-sets ontology (Ontolingua server). 13 fogalom, 64 állítás.
- **GRAPH THEORY**: gráfelméleti fogalmak, műveletek. 27 fogalom, 105 állítás.
- **NUMERIC FUNCTIONS**: matematikai műveletek, függvények. 35 fogalom, 210 állítás.
- **TEMPORAL CONCEPTS**: idővel kapcsolatos fogalmak. Források: James Allen temporális axióma-rendszere (Allen 1984), Simple-Time ontology (Ontolingua server). 60 fogalom, 356 állítás.

- ❑ **MEREOTOPOLOGY:** mereológiával, a rész-egész viszonyokkal kapcsolatos fogalmak. Forrás: Barry Smith és Nicola Guarino, a lyukak elméletéről Casati és Varzi. 24 fogalom, 162 állítás.
- ❑ **UNITS OF MEASURE:** mértékegységek. Források: Quantities ontology (ITBM-CNR), "Standard Units" és "Standard Dimensions" ontológiák (Ontolingua server), átdolgozta Helena Sofia Pinto (Instituto Superior Tecnico in Portugal), Taylor (1995) és az Encyclopaedia Britannica alapján. 102 fogalom, 391 állítás.
- ❑ **PROCESSES:** folyamat-fogalmak tipológiája, Beth Levin „Verb Classes And Alternations” c. munkája nyomán. 163 fogalom, 552 állítás.
- ❑ **OBJECTS:** a 3-dimenziós térben létezésük bármely pillanatában megtalálható általános fizikai dolgok, földrajzi helyek, élőlények stb. 151 fogalom, 499 állítás.
- ❑ **QUALITIES:** szerep-relációk, attribútum-osztályok (pl. halmazállapot, biológiai nem stb.) 112 fogalom, 346 állítás.

A SUMO 11 ontológiai moduljának egymásra épülését az alábbi ábra szemlélteti:



Az egyes fogalmak több részontológiába is beletartozhatnak. A SUMO részontológiái azért lehetnek fontosak a MEO-projekt számára, mert egyrészt segítséget adhatnak a csúcsontológia-építés – most következő – konszolidálási szakaszában, másrészt a SUMO és MEO további részterületeinek összerendelésével egyre nagyobb felületen teremthetünk kapcsolódási pontokat más ontológiák irányába (úgy tűnik, esély van arra, hogy a SUMO ilyen vonatkoztatási, referencia ontológia legyen), harmadrészt a jelentésreprezentációs munka során számunkra szükséges mikroelméletekhez is hasznosíthatjuk a SUMO részontológiáit.

Bővebben: [SUMO-ido]

SUMO/MEO időontológia

A különböző ontológiák közti integrációs lehetőségek megvalósíthatóságát, az integrációhoz szükséges feladatokat a SUMO és a MEO idővel kapcsolatos részontológiái között „próbáltuk ki”. Bár az összehasonlításnak, összerendelésnek egyelőre komoly „műfaji” nehézségei voltak, mivel a SUMO idővel kapcsolatos fogalmai formalizált axiómákban, logikai kijelentésekben álltak rendelkezésre, míg a MEO-ban az idő-fogalmaknak csak a külső kapcsolatai készültek el eddig (melyek főként a két hierarchikus reláció mentén épültek fel). Mindenesetre elkészült az idő-fogalmak magyar fordítása és a két ontológia elemei közti összerendelés. Az idő-fogalmak fordításainak egy részére az alábbi javaslat született:

SUMO angol terminus	MEO magyar terminus
TimeMeasure (class)	IdőMérték
TimeDuration (class)	IdőHossz
DayDuration (class)	NapHossz
TimePosition (class)	IdőPozíció
TimeInterval (class)	IdőIntervallum
Day (class)	Nap
TimePoint (class)	IdőPont

TemporalRelation (relation)	IdőReláció
time (relation)	idő
date (relation)	dátum
before (relation)	előtt
during (relation)	közben
DayFn (relation)	NapFv
duration (relation)	időtartam

Az összerendelési munkához tisztázni kellett, milyen módon, milyen metarelációkkal fejezzük ki a két rendszer elemei közti különböző jellegű kapcsolatokat. Más, hasonló célú projektek gyakorlatát és terminusait átvéve az alábbi összefüggéseket rögzítettük az összehasonlítás során:

- pontos megfelelés („=” szimbólum):
a két fogalom ekvivalens, ugyanazt jelenti; egy-egyhez megfeleltetés
- nem pontos megfeleltetés („~” szimbólum):
a két fogalom nem pontosan azonos, az egyik a másik legközelebbi párja a saját rendszerében (a „közelség” konkrétan nem részletezett); a megfeleltetés több-többhöz is lehet.
- alárendelő ekvivalencia („<” szimbólum):
az egyik rendszerben csak specifikusabb fogalom van
- fölrendelő ekvivalencia („>” szimbólum):
az egyik rendszerben csak általánosabb fogalom van
- nincs megfeleltetés („!” szimbólum):
az egyik oldal egy fogalmával megfeleltetés nem lehetséges a másik oldalon (pl. még közelítő fogalom sincs).

A fenti metarelációk segítségével a SUMO és a MEO idő-fogalmai között az alábbi kapcsolatokat tudtuk megállapítani:

MEO-fogalmak	SUMO-fogalmak	típus	megjegyzés
viszonyidő, időnévmások	TemporalRelation	<	nincs teljes lefedés (SUMO)
naptári idő, óraidő	TimeDuration	<	
időpont	TimePoint	~	oszthatatlan?
időtartam	TimeInterval	~	
dátum	date	~	eltérő pontosság
időkezdet	BeginFn	=	
nap, stb.	DayDuration, stb.	=	
-	Day, stb.	!	hiányzó MEO-I fogalmak
múlt	PastFn	=	
jövő	FutureFn	=	
korábban	earlier	=	
közben	during, holdsDuring	~	több szintaktikai változat

Bővebben: [SUMO-ido], [idotez]

A MEO igei vonzatkeret-tárának felépítése

A projekt 4.3. szakaszában el kellett kezdenünk az igei vonzatkeret-tár felépítését. Újra és újra fel kell hívnunk a figyelmet arra a fontos különbségre, hogy az igei vonzatkeret-tár nyelvi-lexikai, tehát szavak közti, nem pedig ontológiai, tehát fogalmak közti információt rögzít. Emiatt az igei vonzatkeretek „csak” útmutatást, ráutalást nyújthatnak az ontológiai építkezés számára, közvetlenül nem vehetjük át a szó-szinten megállapított kapcsolatokat a fogalmi szintre. Ennek tudatában persze nyugodtan mondhatjuk, hogy a fogalmak közti kapcsolatok formalizálása során igen komoly segítséget jelenthet a vonzatkeret-információ.

A 4.3. munkaszakaszban a MEO-csúcsontológiába tartozó igék vonzatkereteit töltöttük fel adatokkal. 430 ige esetében 1780 kapcsolatrendszerrel, argumentumszerkezetet rögzítettük. A munka eredménye egyelőre egyszerű táblázatos formában áll rendelkezésre. Nem volt fontos a belső struktúra kialakítása, mivel a vonzatkeret-információt nem direkt módon kellett használnunk. Ebben a fázisban csak az eseményszerű fogalmak formalizálásához adott olykor segítséget ez a tár. Az igékhez kötődő vonzatkeret-adatok nagyvolumenű hasznosítása a következő munkaszakaszban várható, amikor már rendelkezésre állnak az eseményszerű (általában igékkel reprezentált) fogalmak közti kapcsolatok (relációfogalmak) is, amikor a konzisztenciaellenőrzés egyik eszközeként alkalmazhatjuk az ilyen típusú információt.

Bővebben: [MEO-verb]

Projektkommunikáció

A projekt kommunikációja a 2. munkaszakaszban is két csatornán keresztül zajlott. Egyrészt a munka során létrejött dokumentumokat, részeredményeket folyamatosan közzé tettük az ontológia portálon (<http://ontologia.hu>), másrészt a projekt résztvevői egy belső levelezési listát használtak az adminisztratív teendők megbeszélésére. Olykor előfordult, hogy a MEO levelezési listán is szakmai vita, beszélgetés alakult ki, ilyenkor törekedtünk arra, hogy a lényegi információk kikerüljenek valamilyen formában a portál felületére is.

A projekt eredményeként két publikáció jelenik meg a szegedi III. Magyar Számítógépes Nyelvészeti Konferencia kötetében.

Hivatkozások, projektdokumentumok

Résztanulmányok

[CC-spec]

Scriptum Rt., *Távközlési közönségzolgálat gépi támogatásának lehetőségei – Specifikáció*

http://ontologia.hu/document/proj_doc/cc_szakontologia/1_3_Tavkozlesi_kozonsegszolgalat_gepi_tamogatasanak_lehetosegei.pdf

[CC-tb]

Scriptum Rt., *Távközlési közönségzolgálati tudásbázis konszolidálása*

http://ontologia.hu/document/proj_doc/cc_szakontologia/1_4_Tavkozlesi_kozonsegszolgalati_tudasbazis_konszolidalasa.pdf

[jelrepyelv]

Gyarmathy Zsófia, Héja Enikő, Szeredi Dániel, Szóts Miklós, Varasdi Károly, *Jelentésreprezentációs nyelv*

<http://ontologia.hu/Members/varasdi/jelentesrep.pdf/download>

[jelreppeldak]

Gyarmathy Zsófia, Héja Enikő, Szeredi Dániel, Szóts Miklós, Varasdi Károly, *Prototipikus példák a jelentésreprezentációra*

<http://ontologia.hu/Members/varasdi/melleklet.pdf/download>

[kovrend]

Scriptum Rt., *Következtető rendszerek részletes vizsgálata*

http://ontologia.hu/Members/Bakos_Gabor/tanulmanyok/kovetkeztetorendszerek/kovrend.pdf

[kovrendfunkt]

Scriptum Rt., *Következtető rendszerek összetétele - funkcionális összehasonlítás*

http://ontologia.hu/Members/Bakos_Gabor/tanulmanyok/kovetkeztetorendszerek/kovrendfunkt.pdf

[kovrendsajat]

Scriptum Rt., *Következtető rendszerek összetétele - funkcionális összetétel - saját tesztheink*

http://ontologia.hu/Members/Bakos_Gabor/tanulmanyok/kovetkeztetorendszerek/kovrendsajat.pdf

[kovrendtptp]

Scriptum Rt., *Következtető rendszerek összetétele - funkcionális összetétel - TPTP-s tesztheink*

http://ontologia.hu/Members/Bakos_Gabor/tanulmanyok/kovetkeztetorendszerek/kovrendtptp.pdf

[kovspec]

Szakadát István, *Ontológiai konzisztenciaviizsgálat (Követelmény- és funkcionális specifikáció)*

http://ontologia.hu/document/proj_doc/consistency

[OntoClean]

Szóts Miklós, *Az OntoClean metodológia ismertetése, problémái és továbbfejlesztési lehetőségei*

<http://ontologia.hu/Members/szots/OntoClean.pdf>

[Protege]

SZTE, *Protégé használhatósági tanulmány*

http://ontologia.hu/document/proj_doc/ontologiai_szerkesztok/MEO_SZTE_20051201.doc

[SUMO-ido]

Miháltz Márton, *Idővel kapcsolatos kifejezések a SUMO ontológiában és összehasonlításuk a MEO időontológiájával*

http://ontologia.hu/Members/mihaltz/Tanulmányok/MM_idoter2.doc.zip

Részontológiák

[idotez]

Ungváry Rudolf, *Időtezaurusz*

<http://ontologia.hu/forum/Members/ungvary/tezauruszok/Idotezaurusz.mdb>

[gold]

Gold – nyelvészeti ontológia (MS Access)

<http://ontologia.hu/document/KOS/Egyeb/gold/goldmdb>

[MEO-gener]

Ungváry Rudolf, *Generikus hierarchia, 2005.10.25.*

http://ontologia.hu/document/KOS/KOS_MEO_folder/KOS-MEO-02/Generikus_hierarchia_2005_10_25.rtf

[MEO-top]

Ungváry Rudolf, *A csúcsontológia lexikai egységeinek csoportosított állománya 2005. 10. 25.*

http://ontologia.hu/document/KOS/KOS_MEO_folder/KOS-MEO-02/Lexikai_egysegek_csoportositasa_2005_10_30.rtf

[MEO-verb]

NYTI, *MEO-igék vonzatkeret-tára*

http://ontologia.hu/document/KOS/KOS_MEO_folder/KOS-MEO-02/vonzatkeret2005

[ontomodell]

Az ontológia adatbázis-modelljének és felhasználói felületének prototípusa (MS Access)

http://ontologia.hu/document/KOS/KOS_MEO_folder/meomodell

[telcotez]

Ungváry Rudolf, *Távközlési tezaurusz 2005. 11. 30. állománya*

http://ontologia.hu/document/KOS/thesaurus/tavkozlesi/Tavkozlesi_tezaurusz_2205_11_25.zip

[telcotezHier]

Ungváry Rudolf, *Távközlési hierarchia 2005. 11. 30. állománya*

http://ontologia.hu/document/KOS/thesaurus/tavkozlesi/Hierarchikus_resz_2005_11_25.rtf

A projekt keretében megszületett publikációk

[Szóts & Lévay 2005]

Szóts Miklós, Lévay Ákos, Szerepfogalmak az ontológiákban - az OntoClean metodológia továbbfejlesztése,

in: Alexin Z., Csendes D., *III. Magyar Számítógépes Nyelvészeti Konferencia, 2005*

<http://ontologia.hu/forum/Members/szots/szerepEloadaas.pdf>

[Szakadát 2005]

Szakadát István, Réteges struktúra, alaprelációk, in: Alexin Z., Csendes D., *III. Magy. Szám. Nyelv. Konf., 2005*

http://mokk.bme.hu/archive/sznyk3_reteg_sy