

MEO ontológiamodell

Szakadát István¹, Szóts Miklós², Gyepesi György², Varasdi Károly⁴,
Ungváry Rudolf³, Simonyi András^{1,2}, Gyarmathy Zsófia⁴,
Szaszko Sándor⁵, Szeredi Dániel^{1,4}

¹ BME SZKT MOKK, Budapest Stoczek utca 2.
i@syi.hu, dani@szeredi.hu, andras.simonyi@gmail.com

² ALL, Budapest Hankóczy u. 7.
{szots,gyepesi}@all.hu

³ OSZK, Budapest Budavári Palota,
rudi@hungary.com

⁴ MTA NYTI, Budapest Benczur u. 33.
varasdi@nytud.hu, gyzsof@budling.nytud.hu

⁵ BME TMIT, Budapest Magyar tudósok krt. 2.
szaszko@tmit.bme.hu

Kivonat Cikkünkben először röviden értékeljük a MEO ontológia-infrastruktúra építő projekt eredményeit, majd bemutatjuk azt a keretrendszer, ontológia modellt, amely – saját ontológiai elkötelezettségeink szerint – mindenféle más ontológiaépítő munka alapjául, keretétül szolgálhat. Ennek során foglalkozunk a tárgy- és metaszintű fogalmak, illetve a nyelvi és fogalmi réteg elkülönítésével. ⁶

Kulcsszavak: ontológia, fogalom, reláció, osztály, attribútum, metafogalom, modell, fogalmi modell, nyelvi modell

1. Bevezetés

A Magyar Egységes Ontológia (MEO) projekt 2004 decemberében indult és 2006 végén zárul le. A projekt célja között szerepelt egy csúcsontológia és egy szakontológia felépítése, egy szakterületi ontológia támogatásával működő alkalmazás kifejlesztése, illetve az ontológiák építésére, menedzselésére alkalmas ontológia-infrastruktúra felépítése.

A projekt lezárultával egy demoalkalmazás segítségével szemléltetni tudjuk, hogy miként lehet hasznosítani az ontológiai tudást a közönségszolgálati tevékenység támogatására a távközlés területén, de a projekt legfontosabb célja mégsem az volt, hogy az ontológiai tudás gyakorlati hasznosításának lehetőségét bizonyítsa. Sokkal inkább arra törekedtük, hogy az ontológiaépítés elméleti és gyakorlati teendőit felfedezzük, leírjuk és elérhetővé tegyük más (későbbi) projektek számára. A MEO-projekt eredményeivel kapcsolatban ki kell emelnünk,

⁶ A MEO-projekt a KPI NKFP 2/042/04. sz. támogatásával jött létre.

hogy – a nyílt forráskódú kezdeményezések mintájára, azok általánosításaként létrejött Creative Commons mozgalom szellemében – a projekt eredményeinek döntő hányada szabadon hozzáférhető. Úgy gondoljuk, hogy a számítógépek számára biztosítandó szemantikai, ontológiai tudásbázist – anyanyelvünkhöz hasonlóan – csak hosszútávon és közös, egymáshoz igazodó erőfeszítések eredményeként tudjuk csak felépíteni és fenntartani. A MEO ehhez a közös munkához próbálta meg letenni az alapokat.

Az ontológiai tudást a konkrét alkalmazásokban szakontológiák segítségével lehet igazán hasznosítani. A MEO-projekt mégis – története során mindvégig – a csúcsontológia felépítésére koncentrált. Bár a csúcsontológiáknak is lehet gyakorlati haszna akkor, amikor egymástól elkülönülten fejlesztett informatikai rendszereket kell kooperációra készíteni, ám figyelemkoncentrációnk és fókuszunk a csúcsontológia irányába inkább annak a szándéknak tulajdonítható, hogy „ezen a terepen” lehetett a legjobban megtanulni, megérteni és megoldani, kezelni remélni az ontológiaépítés problémáit.

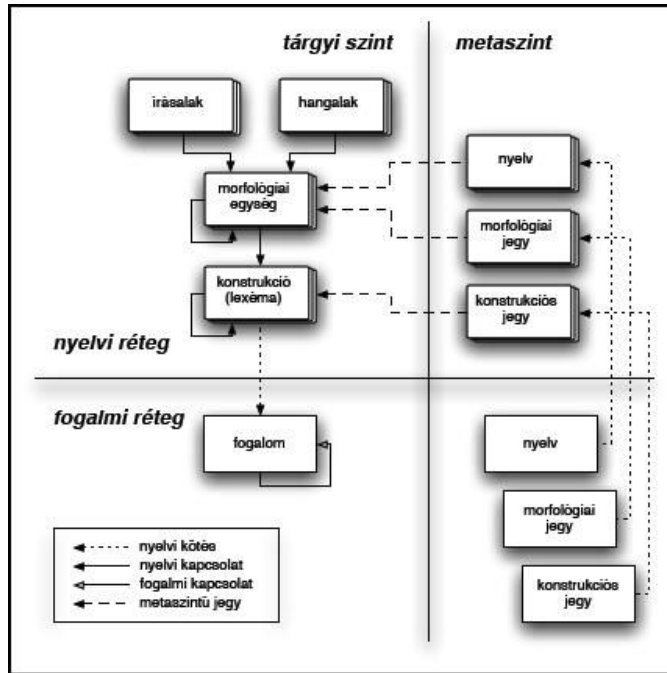
Az ontológia a világ dolgairól szóló tudásunkat tárolja nyelvfüggetlen formában. A MEO projekt kiemelt célja volt az ontológiák tényleges nyelvfüggetlenségének biztosítása, ezért az ontológia modellezése során elkülönítettük a nyelvi és fogalmi rétegeket egymástól. A következőkben röviden bemutatjuk a két réteg szintjeit, elemeit, kapcsolatait.

2. A nyelvi réteg

Mégha az ontológiát nyelvfüggetlennek gondoljuk is, beszélni róla (a rendszer egészéről, a részeiről, az elemeiről) csak valamilyen nyelv segítségével tudunk. Azért van mindenképpen szükség nyelvi kötések létrehozására, hogy az ontológiát használó emberek is beszélni tudjanak róla. A MEO-projekt számára olyan modellt terveztünk, amelyben egyrészt egyértelműen el lehet különíteni a fogalmi és nyelvi rétegeket egymástól (persze mindezt úgy, hogy eközben a fogalmakhoz hozzákötjük az adott nyelven a fogalmakra mutató kifejezéseket, lexikai egységeket), másrészt tetszőleges (számú és típusú) nyelvet lehet a fogalmi réteghez kapcsolni. A nyelvi réteg entitásait mutatja az 1. ábra.

A nyelvi rétegben három – egymásra épülő, de egymástól jól elkülöníthető – entitást definiáltunk. A „legalsó szinten” van a *szóalak* (ami lehet írás- vagy hangalak). A szóalak még nem igazán nyelvfüggő entitás, amire azért van szükség, mert a korpuszokból számolt gyakorisági értéket csak ehhez tudjuk hozzárendelni. A szóalakhhoz valamilyen morfológiai jegyhalmazt rendelve jutunk a *morfológiai egység* fogalmához. Ez a nyelvi entitás már egyértelműen nyelvfüggő, és a hozzá kapcsolt morfológiai jegyek révén lehet megérteni a szóalakok nyelvi megnyilatkozáson belüli viselkedését. Meg kell jegyezzük azonban, hogy a nyelvi réteg első két entitása, a szóalak és a morfológiai egység még nem igazán használható szemantikai, ontológiai célokra.⁷ A nyelvi réteg harmadik entitása, a *konstrukció* az, ami kapcsolatot teremt a nyelvi és fogalmi réteg (vagyis a morfológiai és

⁷ A projekt „melléktermékeként” ugyan belekezdünk egy nyelvi ontológia felépítésébe, de ez nem a MEO közvetlen céljainak eléréséhez volt szükséges.



1. ábra. A MEO modell nyelvi rétege

szintaktikai viselkedés, illetve a jelentés) között. Ha a jelentést az ontológiai réteg hordozza, akkor a konstrukció az, ami a nyelvi megnyilatkozásokat összeköti a szemantikával. A konstrukció gyakran lexikai egység (lexéma), de az ontológiában leírhatjuk más nyelvi egységek (pl. a toldalékok) jelentését is, és az ilyen típusú konstrukciókon keresztül kapcsolatot teremthetünk más típusú morfológiai egységekhez is. A konstrukció természetesen a morfológiai egység szintjéről örökli annak nyelvi kapcsolatát, tehát nyelvfüggő elem.

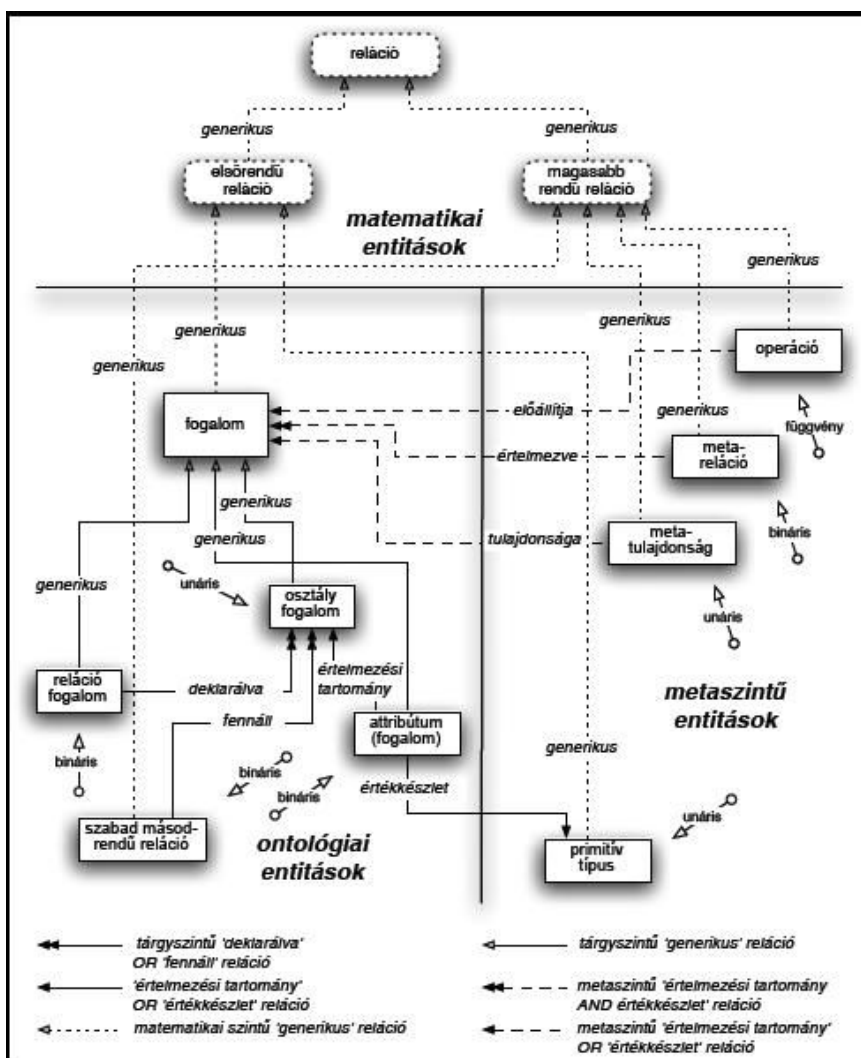
Mivel a konstrukció a nyelvi réteg eleme, ami a kapcsolt fogalmon keresztül kap jelentést, tetszőleges számú nyelven kifejezett konstrukciót lehet ugyanahhoz a fogalomhoz rendelni, ami biztosítja azt, hogy az ontológia elemeihez akárhány nyelven lehessen nyelvi megnyilatkozásokat fűzni. Az ontológia- és szótárépítési munkát ezáltal egymástól elválasztva, de egymáshoz mégis illeszkedve lehet végezni. Ennek során csak arra a szabályra kell figyelni, hogy minden új fogalomhoz biztosítani kell egy alapértelmezett nyelvi konstrukciót, így, ha valamely nyelven nincs egy ontológiai egységnek megfelelő nyelvi konstrukció, akkor még megjeleníthető az ontológiai rétegben rögzített információ. Ilyen esetekben a „nyelvi lyukakat” az „idegen nyelven” megjelenő konstrukciók felbukkanása jelzi.

Bár széles körben hangoztatják, hogy az ontológiákba foglalt tudás nyelvfüggetlen, nem nagyon tudunk olyan ontológiaépítő projektről, illetve olyan ontológiaszerkesztő alkalmazásról, amely hatékonyan támogatná a nyelvfüggetlen építkezést. Ezért döntöttünk a projekt félidejében úgy, hogy saját szerkesztő fejlesztésébe kezdünk

(MEOdit), mert biztosítani szeretnénk volna, hogy a MEO csúcsonológia elemeihez valóban több nyelven lehessen nyelvi szótárt illeszteni (jelenleg négy nyelven, magyarul, angolul, latinul és lengyelül áll rendelkezésre a MEO csúcsgfogalmak durván 2700 elemű készletének nyelvi kötése).

3. A fogalmi réteg

A MEO ontológiamodell másik fontos része a fogalmi réteg, mely az ontológiák építéséhez használható, metaszinten rögzített, a tárgyszinten nem változtatható metafogalmakat és a tárgyszinten az ontológiaépítők által szabadon építhető fogalmakat tartalmazza. A fogalmi réteg vázlatos ábrája a következő:



2. ábra. A MEO modell fogalmi rétege

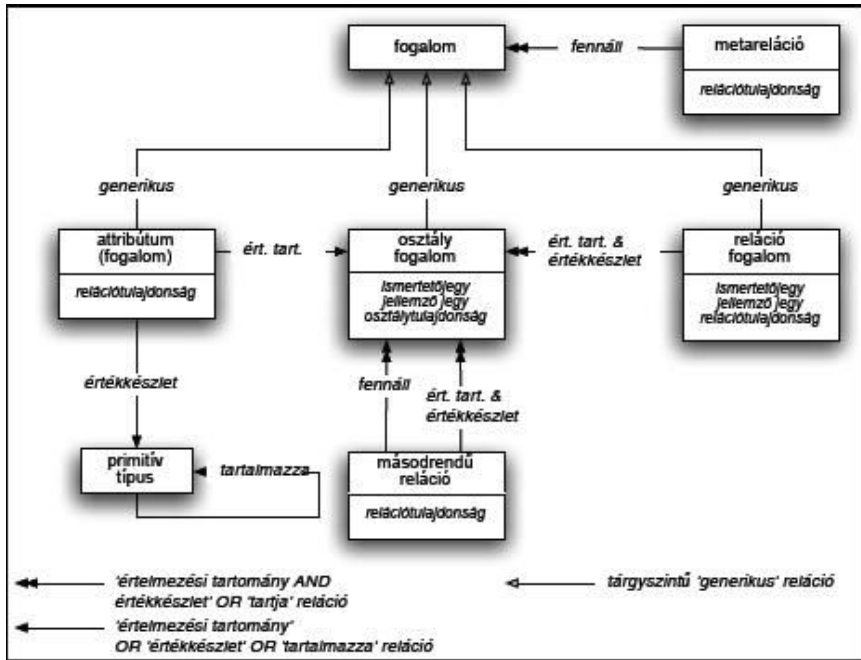
A fogalmi réteg felépítésekor arra törekedtünk, hogy egyértelműen el tudjuk választani a tárgy- és a metasztintű (illetve meta-metasztintű) entitásokat egymástól. A tárgyszint (és persze az egész ontológia) központi entitása a *fogalom*, amelynek három típusát, az *osztály*-, a *reláció*- és az *attribútumfogalmakat* különítettük el (szemben azzal a szokásos megoldással, ami csak az osztályfogalmakat veszi figyelembe). A relációfogalmon elsőrendű relációkat értünk, amelyeket az ontológia építői az osztályfogalmakon deklarálnak. Emellett mód van még másodrendű relációk felvételére is. Az attribútumfogalom is relációnak minősíthető, a különbség csak annyi közöttük, hogy az attribútum értékészlete a metasztinten rögzített primitív adattípusokból állhat. A fogalom típusait (tehát az osztályt, a relációt és az attribútumot) generikus reláció köti a fogalomhoz. A *generikus alárendeltje* reláció (amire használják az is-a, a szubszumpció, az inklúzió, a részosztály és még sok minden más terminust) már a metasztinthez tartozik, és másodrendű reláció, hiszen fogalmakon (vagy másként: predikátumokon) érvényesített univerzális kvantor segítségével lehet definiálni.

A metasztinthez kötöttük az új fogalmak képzésére alkalmazható *műveleteket* (az egyszerű relációsorzással állíthatjuk elő az apja fogalmából a nagyapja fogalmát), illetve ehhez a szinthez rendeltük a fogalmak *metatulajdonságait* (mint például az OntoClean módszertan rigid vagy esszenciális tulajdonságait vagy a relációk algebrai tulajdonságait, mint a szimmetricitás vagy tranzitivitás). A metasztint elemeinek nagyon fontos vonása az, hogy azokat egy adott ontológia építői nem vehetik fel (és nem is törtölhetik) szabadon, azok a „rendszerbe vannak égetve”. A MEO építkezése során ez a megoldás minősíthető a legerősebb ontológiai elkötelezettségnek. A tárgyszintre ebből az ontológiai elkötelezettségből csak a fogalom alá sorolt legfőbb típusok létezésének elismerése került be. Ettől eltekintve szabadon építhető a tárgyszint fogalmi rendszere.

A metasztinten rögzített metafogalmak meghatározását sem végezhetjük el másként, mint hogy valamilyen relációs definíciót adunk rájuk (a művelet homogén, balról egyértelmű, jobbról totális, kétargumentumú reláció, míg a *tulajdonsága* metafogalom olyan másodrendű reláció, amely egy tetszőleges fogalmat és egy tulajdonságot kapcsol össze a tárgyszinten). Ehhez természetesen újra szükség van a reláció fogalmára, vagyis már meta-meta szinten kell a definíciókat megadni. Erre szolgál a modell harmadik szintje, ahol már „matematikai értelemben” használjuk a relációfogalmakat.

A fogalmi réteg a tárgyszint fogalmaira fókuszál, ami az ontológiák főnről lefelé történő (top-down) építését teszi lehetővé. A fogalmi rendszerek kibontásakor azonban nemcsak a fogalmakat szokták megadni, definiálni. Olykor megpróbálják meghatározni a fogalmak (pontosabban az általuk „leírt”, „hivatkozott” létező entitások) tulajdonságait, jegyeit is. Elképzelhető tehát olyan ontológiaépítkezés, amely az alulról fölfelé történő (bottom up) építkezési módszertanra támaszkodik. A MEO-ban kezdettől fogva arra törekedtünk, hogy a modellünkben meg tudjuk ragadni ezt a kettősséget. Ezért egyrészt a projekt egyik munkacsoportjában a tulajdonságok egymásra épülő rendszerének egy lehetséges modelljét próbáltuk meg kidolgozni, másrészt a top-down modellen belül is megteremtettük annak le-

hetőségét, hogy a fogalmakhoz rendelhető ismertetőjegyek rendszerét ki lehessen formálni. Ezt mutatja be a 3. ábra.



3. ábra. A fogalmakhoz rendelhető jegyek rendszere

Az ábrán látható, hogy a fogalmakat leíró tulajdonságok, jegyek változnak a fogalom típusától függően. A MEO-modell alapján megvalósuló ontológiaépítkezési munkák során a jegyeket bármiféle kényszer nélkül lehet csak a fogalmakhoz kapcsolni, tehát – egyelőre – nem lehet a fogalmak közötti relációk és a fogalmakat jellemző jegyek közti összefüggéseket nyomon követni (és kihasználni). A vízió szintjén természetesen megfogalmaztuk azt az Arisztotelészre visszavezethető elképzelést, miszerint csak úgy lehet megadni egy új fogalmat, ha meghatározzuk a fogalmat egyértelműen jellemző, „azonosító” tulajdonságot (differentia specifikát), de a tényleges gyakorlatban mindezt – egyelőre – nem tudtuk implementálni.

Az ontológiaépítő munka egyik fontos összetevője az életciklus-menedzsment. Az ontológiák esetében mindezt azt jelenti, hogy kezelni kell tudnunk az ontológiák részeit, részrendszerait (vagy másként, az ontológiát gráfként értelmezve, a részgráfokat), az ontológia összetevőire (elemeire, részeire, részrendszerreire) vonatkozó egyéni és csoportos jogosítványokat. Az ontológia szerkesztéséhez, a fogalmak definiálásához persze egyfajta konszenzusmenedzsmentre is szükség van, mivel arra kell felkészülni, hogy az ontológiák létrehozását, fenntartását több ember kooperációjától remélhetjük csak, így a közösen épített rendsze-

rek konzisztenciájához biztosítani kell valahogy a vitás kérdések rendezésének lehetőségét. Túl sok választási lehetőségünk persze nincs. Ezt a helyzetet a közösségi döntések egyik típusának minősítve csak az lehet a nyitott kérdés, hogy a közösségi döntéshozatali eljáráshoz milyen konkrét megoldásokat, algoritmusokat választunk magunknak a társadalmi választások elmélete által kínált repertoárból.

Említettük, hogy a tárgyszintre kevés ontológiai elkötelezettséget „vittünk be”. Az már a projekt indulása előtt is nyilvánvaló volt számunkra, amit aztán a projekt tényleges beindulásával a saját tapasztalataink is megerősítettek, hogy a tárgyszinten, főleg a csúc fogalmak esetében alternatív fogalomértelmezésekre kell felkészülni, vagyis nincs esély és remény a teljeskörű konszenzus kialakítására. Minél közelebb (vagyis a hierarchiában minél lejjebb) kerülünk azonban a hétköznapi szinten gyakran használt fogalmakhoz, annál valószínűbb, hogy konszenzust lehet találni ezek meghatározásában, illetve a fogalmak ontológiai struktúrában elfoglalt helyeinek megállapításában. Ezt a konszenzusmenedzsment egyik különös feladatának minősíthetjük. Természetesen a MEO csúcsontológiában határozott ontológiai elkötelezettség érhető tetten, hiszen nekünk is rögzíteni kellett valahogyan a csúc fogalmainkat, de azt reméljük, hogy az ontológiaszerkesztés elveinek, fogalomkészletének tisztázásával és publikálásával lehetővé tesszük bárki számára, hogy saját ontológiát tudjon előállítani céljai eléréséhez.

Hivatkozások

1. Corcho, O., Fernández-López, M., Gómez-Pérez, A., Methodologies, tools and languages for building ontologies. Where is their meeting point? In: *Data & Knowledge Engineering*, Vol. 46, 2003, pp.41-64.
2. Green, R., Bean, C.A., Myaeng, S.H., *The Semantics of Relationships: An Interdisciplinary Perspective*, Dordrecht: Kluwer, 2001.
3. Guarino, N., Welty, C., A Formal Ontology of Properties. In: *Proceedings of 12th Int. Conf. on Knowledge Engineering and Knowledge Management Lecture Notes of Computer Science*, Springer Verlag, 2000.
4. Guarino, N., Welty, C., Supporting ontological analysis of taxonomic relationships. In: *Data & Knowledge Engineering* Vol. 39, 2001, pp 51-74.
5. Staab, S., Studer, R. (eds.), *Handbook of Ontologies*, Springer Verlag, 2004.
6. Szakadát I., Szóts, M., Gyepesi, Gy., MEO - Ontology Infrastructure. In: Gabor Magyar, Gabor Knapp, Wita Wojtkowski, Gregory Wojtkowski, Joze Zupancic, Stanislaw Wrycza (eds.) *Advances in Information Systems Development: New Methods and Practice for the Networked Society, Proceedings Information Systems Development*, Springer (megjelenés alatt).
7. <http://ontologia.hu/meo>