

Hol fáj? – A jelentésrepresentáció nehézségei egy kórlapkitöltő rendszerben

Gröbler Tamás és Szóts Miklós

Alkalmazott Logikai Laboratórium
1022 Budapest, Hankóczy J. u. 7.
{grobler, szots}@all.hu

Kivonat: Szövegek jelentésének formális reprezentálására szolgáló módszeren dolgozunk. A szövegek morfológiai és szintaktikai elemzése után a jelentésrepresentációt ontológiában állítjuk elő. A szemantikai elemzés feladata a világmodell tartalmazó ontológia példányosítása a szöveg alapján, vagyis a szöveg jelentésének megfelelő individuumok és a közöttük fennálló relációk létrehozása. A szemantikai elemző tudásbázisát a kutatás keretében épülő ontológia adja, amelyet az OWL ontológia-leíró nyelven fogalmazunk meg. A cikkben áttekintjük a rendszer felépítését és működését, és közben körbejárjuk azokat a pontokat, amelyek a jelentésrepresentáció szempontjából különleges bánásmódot igényelnek. Az egyes problémák kezelésére javaslatot teszünk. A konkrét feladat, amely keretében a kísérletet végezzük a kórházi panaszfelvételek során lejegyzett magyar nyelvű szövegek egységes kórlap-representációvá alakítása.

1 Bevezetés

Az a kutatás, amelyről beszámolunk⁵⁷, nem nyelvészeti ihletből született, hanem az ontológiák tanulmányozásából. A kutatás célja, hogy egy szöveg jelentését formálisan reprezentáljuk. Ennek több praktikus felhasználási területe lehet, a jelenlegi projekt orvosi szabad szövegeket formális kórlapstruktúrába interpretál, de az igazi célkitűzés szemantikus kereső fejlesztése.

Noha, a kutatás nem nyelvészeti ihletből született, a kísérlet során egyre szorosabb nyelvészeti kapcsolatokra döbbsentünk rá. A legfontosabb: egy diskurzusrepresentációt építünk, az egyetlen újdonság az, hogy nem csak a szövegből nyelvészeti módszerekkel kinyerhető információra alapozunk, hanem az ontológiában strukturálisan tárolt háttértudásra is támaszkodunk.

Az ontológiát magát sokan mint nyelvi elemek (terminusok) rendszerét értik. Azonban valójában az ontológia fogalmak rendszere, amely fogalmakat természetesen nyelvi kifejezésekkel nevezünk meg. A nyelvi elemek és az ontológia fogalmait már sokan megkülönböztetik, pl. a DOLCE csúcsontológiájához már kidolgozták a

⁵⁷ A kutatást a GVOP-3.1.1-2004-05-0363/3.0 sz. és az NKFP-2/042/04 (MEO) pályázat támogatja.

WordNet lexikonnal való kapcsolatot [3]. Többen használják a szóelőforduláson alapuló visszakeresés javítására ([2], [5], [8]) vagy szakmai szövegek generálására.

Mi az ontológiát **világmodellnek** fogjuk fel, és a szöveg jelentését ebben reprezentáljuk. Ehhez természetesen szükséges a nyelvi tudás tárháza (a továbbiakban lexikonnak nevezzük), valamint ennek elemeinek az ontológiába való leképzése.

Pragmatikus okok miatt feltételezzük, hogy a szöveg egy jól meghatározott témakörre korlátozódik („universe of discourse”), és csak leíró jellegű szövegeket kezelünk. Elvileg ezek a korlátok átléphetők, de a megfelelő ontológia elkészítése rendkívüli erőfeszítést igényelne.

Az ontológia mint „világmodell” tartalmazza azon fogalmakat⁵⁸, amelyek előfordulásaira a reprezentálandó szövegek referálnak, legalábbis azokat, amelyek a szöveg megértése szempontjából fontosak – ezt tárgyaljuk a 2. fejezetben.

A nyelvi tudástár elemei mindazon elemi nyelvi objektumok (szavak, ragok, idiómák stb.), amelyek jelentése az ontológiában valamilyen módon reprezentálva van – erről a 3. fejezetben szólnunk.

A továbbiakban ismertetjük azt a rendszert, amelyen dolgozunk, illetve bemutatjuk, hogyan működik a már megvalósított rendszer orvosi szövegeken.

Az utolsó fejezetben röviden összefoglaljuk a továbbfejlesztés terveit.

2 Ontológia – a világról szóló tudás

Ahhoz, hogy az ontológiában egy tudásterület szövegeinek jelentését reprezentálhassuk, az ontológiának képesnek kell lennie leírni azokat a szituációkat, amelyek a várt szövegekben tipikusan előfordulnak. Ehhez elsősorban egy megfelelően kidolgozott **csúcsontológiára** van szükség, amely megszabja mi az, ami kifejezhető, és mi az, ami nem. A következőkben néhány fontos kérdést veszünk sorra. Ezeket a kérdéseket általános szinten vetjük fel, de felhívjuk az olvasó figyelmét arra, hogy a szakontológiák kategóriái radikálisan eltérhetnek az általános célú ontológiákéitól. Példa: egy orvosi ontológiában a PÁCIENS és az ELLÁTÓ SZEMÉLYZET TAGJA fogalmak közös nemét, az EMBER fogalmat teljesen felesleges felvenni.

2.1 Eseményszerűségek reprezentálása

Az eseményszerűség (occurrence, perdurant: az események, folyamatok stb. gyűjtőfogalma) a természetes nyelvfeldolgozás szempontjából kulcsfontosságú: ezek azok a fogalmak, amelyekre (általában) igével referálunk. A csúcsontológiában az ESEMÉNYSZERŰSÉG és az ENDURANT (térben anyagi és absztrakt létező) kategóriák közt értelmezzük a *résztevője* relációt, amely fajtái lesznek a szereprelációk, és kijelölik milyen fogalom előfordulása lehet pl. egy esemény *aktora*, *tárgya* stb. – lásd [7].

⁵⁸ A MEO projektben kidolgozott ontológiamodell szellemében mind az osztályokat, mind a relációkat fogalmaknak értjük.

2.2 Tulajdonságok

Hogyan reprezentáljuk azt, hogy az „ég kék”, vagyis hogy „az ég színe kék”? Vagy azt, hogy „a páciens vérnyomása 2006. nov. 8-án ülő helyzetben, bal karon mérve 220/178 Hgmm”? Míg az első példa egyszerűnek látszik – bár ott is van egy rejtett időfüggés –, a másodikban már nyilvánvaló a probléma: egy fogalomhoz (PÁCIENS) egy tulajdonság (VÉRNYOMÁS) egy értékét (*vérnyomása*) kell rendelni, amely különböző paraméterektől (TESTHELYZET, MÉRÉSI HELY, IDŐ) függ. Nyilvánvaló, hogy különböző tulajdonságoknál különböző paraméterekkel kell számolnunk. Azt a megoldást választjuk, hogy a „tulajdonsága” relációt reifikáljuk, és azt a *hordozója, értéke*, valamint a paramétereknek megfelelő relációk (példánkban: *ideje, mérési helye, testhelyzetben*) kapcsolják a megfelelő objektumok, értékek osztályaihoz.

2.3 Idő

Az időkezelés minden tárgykör jellemzője, viszonylag egyszerű: megkülönböztetünk időintervallumot és időpontot, valamint egy lineáris skálát rendelünk az időpontokhoz. Az IDŐINTERVALLUM és IDŐPONT fogalmak közt értelmezve van a *kezdőpontja* és a *végpontja* relációk. Azonban az egyszerűséget elrontja a granularitási probléma: az egyes időfogalmak mint NAP, ÓRA stb. hol időintervallumot, hol időpontot jelentenek⁵⁹. Ennek feloldására több út van, mi azt választottuk, hogy minden ilyen időfogalom az IDŐPONT fajtái, amelyek előfordulásai egy eseményszerűséghez rendelt IDŐINTERVALLUM előfordulás kezdőpontja ill. végpontja lehetnek. A szövegből kiindulva a szemantikus elemző határozza meg, hogy milyen fogalom előfordulásai szerepelnek mint időintervallum-határok.

2.4 Hely

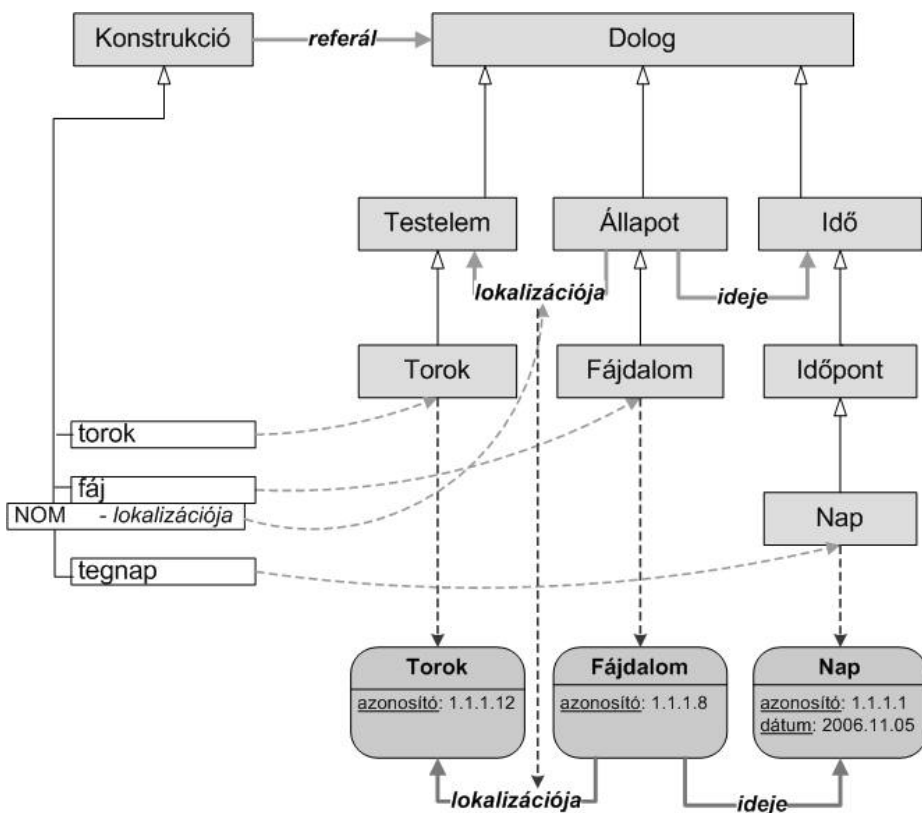
A hely kezelése lényegében különbözik az időtől: nincs, illetve általában nem használatos egy jól meghatározott koordináta-rendszere. Egyéb fogalmak szolgálnak helymeghatározásra. Az orvosi témakörben két, egymástól teljesen különálló helymeghatározással találkozunk: egyrészt az emberi szervezet részeihez kapcsolódnak orvosi fogalmak (pl. „daganat a májon”), másrészt az egyes ellátó egységek mint az ellátás helyei. Két különböző relációval modellezzük ezt a két összefüggést.

3 A nyelvi tudás és az ontológia kapcsolata

A MEO modellnek [6] megfelelően az ontológiának két rétege van: a fogalmi réteg (a tulajdonképpeni ontológia), és a nyelvi réteg, amelyben a fogalmakat megnevező nyelvi elemek vannak. A nyelvi réteg egyelőre nem több, mint egy lexikon, amelyben az egyes lexémák (idiómák, ragok stb.) alatt ezek egyértelműsített változatai szerepelnek. Ezeknek a *referál* függvény adja meg jelentésüket (1. ábra). A lexémák egy csoportja és az igei vonzatkeretekben szereplő morfológiai jegyek közvetlenül is

⁵⁹ Az egy külön probléma, hogy időtartam-egységet is jelentenek.

utalnak egy-egy fogalomra ill. relációra. Elemzés közben a lexémák által referált fogalmak egy-egy előfordulását (példányát) hozzuk létre, amelyeket a fogalmak közötti megfelelő relációval kötünk össze. Eközben a példányok adat típusú tulajdonságokat is felvehetnek.



1. ábra A „Tegnap fájt a toroka.” mondatnak megfelelő ontológiai részlet és jelentésrepresentáció. Az üres nyilak a generikus relációknak felelnek meg. A lexémáknak megfelelő fogalmakat a *referál* relációra kimondott megszorítások kötik a világmodell megfelelő fogalmaihoz. Az alsó sorban a fogalmak egy-egy előfordulása adja a jelentésrepresentációt.

A jelentésrepresentáció generálásában rejlő igazi probléma természetesen nem az, hogy egyes szavak jelentését megadjuk, hanem az, hogy reprezentálhatók legyenek azok a szintaktikai viszonyok, amelyek az egyes szavakat, morfémákat értelmes mondattá szervezik, azaz a régens-bővítmény viszonyokat. Ezek reprezentálása két módon történhet, de mindkét módon a többértelműség problémájával kell megküzdenünk.

- 1. vonzatkeret reprezentálása** A vonzatviszony olyan reláció, amely szó-előfordulások közt értelmezett – ennek az ontológiában a szereprelációk felelnek meg. Ez a megfeleltetés nem univerzális: például a nominatívusz vonzat – bár általában az *ágens*e vagy *elszenvedője* szereprelációknak felel meg – jóformán bármely szereprelációnak megfelelhet, az orvosi szövegekben például sokszor a *lokazáció*-nak. Ezért egyértelműsített lexikonelemként külön ad-

juk meg a vonzatkeretet: a vonzatoknak relációkat feletetünk meg, amelyek kijelölik a lexikonelem megfelelő szereprelációját. Például, a „fáj” szó nominatívusz vonzata a FÁJDALOM fogalom *lokalizáció* szereprelációjának felel meg. Ha egy szóhoz több különböző vonzatkeret is járulhat, ezeket külön-külön egyértelműsített változatként kezeljük.

2. **szabad határozók** A legtöbb rag, amely a magyarban a szabad határozókat jelzi, általában megfeleltethető az ontológiában szereplő relációknak: pl. a -bAn ragnak egyrészt az *ideje*, másrészt a *helye* reláció egy-egy fajtája. Ez az egyik olyan terület, amelyen az OWL által nyújtott definíciós lehetőségek nem elegendőek ahhoz, hogy az ontológiában leríjünk egy nyelvi elem pontos értelmezését.

Természetesen vannak olyan nyelvi elemek, amelyeknek az ontológiában nem feltethető meg referátum. Ilyenek a névelők, névmások, tagadósók stb. Ezeket az alaktani és szintaktikai elemzéshez csatolódva külön szabályokkal kell kezelnünk (l. 5. fejezet).

4 A kórlapkitöltő rendszer

A kifejlesztett rendszer feladata a kórházi panaszfelvételek során lejegyzett magyar nyelvű szövegek egységes kórlap-reprezentációvá alakítása. A kórlapnak tartalmaznia kell a beteg adatait, a rá vonatkozó panaszfelvételek körülményeit, és az egyes panaszfelvételek során leírt tüneteket, panaszokat, azok tulajdonságait, fellépésük és megszűnésük idejét, ill. az esetleg fontos egyéb körülményeket.

Egy szöveg jelentésének meghatározása a morfológiai és szintaktikai törvényszerűségek figyelembe vétele nélkül elképzelhetetlen. Ugyanakkor a nyelvi (elsősorban szintaktikai) elemzés is csak akkor lehet elegendően pontos, ha kiegészül szemantikai információval. Ebből a szempontból ideális helyzet az lenne, ha a morfológiai, szintaktikai és szemantikai elemzés egyszerre lenne elvégezhető. Erre a magyar nyelv esetében is van példa [1]. Egy ilyen elemző megvalósítása hosszabb távú terveinkben szerepel, addig azonban a morfológiai és szintaktikai elemzést előfeldolgozásnak tekintjük.

A szöveg nyelvi előfeldolgozását a MorphoLogic kft. morfológiai és szintaktikai elemzője [4] végzi. A szintaktikailag elemzett szöveget XML formátumú elemzési fájl formájában kapjuk meg. Az elemzés fejközpontú: minden nagyobb egység, csoport, frázis (lehetőség szerint) hivatkozik arra a terminális elemre, amelyik a fejének tekinthető. A terminálisok releváns morfológiai jegyei is fel vannak tüntetve, ezek nagymértékben segítik az egyes szövegelemek közötti szemantikai relációk felderítését.

A projekt keretében rendelkezésre álló orvosi szövegek esetében különös jelentősége van a szövegnormalizálásnak. Kezeleni kell a szakterületre jellemző idegen (jellemzően latin) szavakat, a rövidítéseket ill. ezek legkülönfélébb változatait, a számokat és a szövegek sietős lejegyzése miatt feltűnően gyakori hibákat, elgépeléseket. A jelenleg működő rendszer egyelőre normalizált bemenettel dolgozik, a szövegnormalizálást a következő fázisban szintén a MorphoLogic-kal közösen valósítjuk meg.

Szoftverrendszerünk tehát tartalmazza a nyelvi előfeldolgozót, a jelentésreprezentációt létrehozó szemantikai elemzőt, és a reprezentációból a kórlap egyes mezőinek megfelelő információt kiolvasó kórlapkitöltőt. A nyelvi előfeldolgozó saját szótári és

nyelvtani adatbázissal rendelkezik, a szemantikai elemző adatbázisa pedig a fent tárgyalt ontológia. Az ontológiát a Protégé ontológiaszerkesztővel szerkesztjük, és a Java nyelven írott programkódunk a Jena fejlesztői programcsomag segítségével használja.

5 Eddigi eredményeink

A jelentésrepresentáció ontológiában való megvalósítása számos problémát vet fel. Ezek egy részét az ontológia megfelelő kialakításával és használatával kezelni tudjuk, más esetekben viszont ki kell lépni az ontológia megszabta keretek közül, és külön kell kezelnünk a problémát. Reményeink szerint a korábban említett lexikalista nyelvtenba integrálódva ezek az esetek is a többivel együtt lesznek kezelhetők.

5.1 Egyértelműsítés, anafora-feloldás

A szabad szövegek elemzésekor rendre felmerülő egyik jelentős kihívás a rendszerben több szinten megjelenő többértelműség. A nyelvi elemzés eredménye akár a lexikális, morfológiai, akár a szintaktikai szinten is többértelmű lehet (l. a címben szereplő „Hol fáj?” kérdést). A lexémáknak az ontológiában több konstrukció is megfelelhet, amelyek különböző fogalmakra referálnak.. A fogalmak közötti kapcsolatrendszerben is többféle reláció állhat fenn két fogalom között.

Egyfajta többértelműségnek tekinthetjük az anaforák jelenlétét is, hiszen általában több, korábban előforduló fogalom közül kell kiválasztanunk, hogy melyikre utalnak. Ilyenkor a már feldolgozott fogalmak közül az anafora által meghatározott tulajdonságúakat keressük ki.

Az egyértelműsítés alapja minden esetben az, hogy melyik reprezentáció a teljesebb abban az értelemben, hogy a szövegből több információt tud kinyerni. Ennek mérésére a fogalmak közti relációkat aszerint értékeljük, hogy az általuk összekötött előfordulások szempontjából mennyire specifikusak. Azt a reprezentációt választjuk ki, amelyik minél több és minél specifikusabb relációt tartalmaz. Az egyértelműsítést praktikus okokból mindig a lehető legalacsonyabb szinten végezzük el.

5.2 Ismeretlen szavak

Bármekkora ontológiát is építünk, a feldolgozandó szöveg szükségképpen fog tartalmazni ismeretlen szavakat. Megtehetnénk, hogy ezeket kihagyjuk a jelentésrepresentációból, de előfordulhat, hogy éppen egy ismeretlen szó forrasztja össze a jelentés egyébként széteső részeit, hiszen a mondat szintaktikai szerkezete egyértelműen kijelölheti a szerepét. Éppen ezért az ismeretlen szavak reprezentálására a VALAMI csúcscategória egy előfordulását hozzuk létre, illetve az ismeretlen kapcsolatok jelölésére bevezettük a *dummyReláció* nevű relációt.

Az egynél nagyobb számosságú objektumok kezelése is kihívást jelent az ontológiaszerkesztők számára. Ebben a projektben az a megoldás tűnt a leginkább kezelhetőnek, hogy minden megszámlálható dolog egyben halmaz is lehet, amelynek elemei ugyanolyan típusú objektumok, mint önmaga. Egy előfordulás azáltal válik halmazzá, hogy vagy a számosságát vagy az elemeit meghatározzuk. A számmal kifejezhető SZÁMOSSÁG mellett megkülönböztetjük a KVALITATÍV_SZÁMOSSÁGOT (pl. sok, kevés) is (l. 3. ábra).

5.4 Ontológián kívüli módszerek

A természetes nyelvi szöveg számos olyan elemet tartalmaz, amelyek nem a világmodell fogalmaira referálnak, hanem a többi fogalom jelentését vagy egymáshoz való viszonyát módosítják. Ezért létrehoztuk a nem tartalmas szavak szótárát, amely egyszerű szintaxissal megírt utasításokat ad a szemantikai elemzőnek az egyes szavak kezelésére vonatkozóan.

A legegyszerűbb esetben az adott szóval egyáltalán nem foglalkozunk (pl. kötőszók, „csak” stb.). Ilyenkor vagy a szintaktikai elemzés hordozza a megfelelő információt (pl. mellérendelés a kötőszók esetében), vagy a szóban rejlő információt nem kívánjuk reprezentálni (pl. szubjektív minősítés a „csak” esetében).

A tagadószókat és fosztóképzőket már a nyelvi előfeldolgozás során megjelöljük, és ezt figyelembe véve kell a létrehozott példányok igazságértékét megállapítani. Szintén még az előfeldolgozás során a mellérendelő szerkezeteknél, felsorolásoknál a közös bővítményeket a felsorolás minden elemére kiterjesztjük.

Vannak speciális jelentéssel bíró kifejezések, amelyek az adott mondat komplex kezelését kívánják meg. Például az „egyéb panasza nincs” kifejezés azt jelenti, hogy a panaszok halmaza (amelynek korábbi panaszok az elemei) lezárt, abba újabb elemet nem lehet felvenni.

A 4. ábra egy konkrét példába sűrítve mutat meg néhányat a fenti jelenségek közül.

6 Hogyan tovább?

A jelenleg elkészült kórlapkitöltő rendszer számos kérdésre adott választ, de az itt bemutatott vízióknak szempontjából egyelőre csak egy demó, amelynek számos korlátja van. További feladatok az alábbi területeken adódnak.

A rendszer jelenlegi architektúrája hosszabb távon nem tartható: a szintaktikai és a szemantikai elemzés szétválasztása a módszer szellemével ütközik. Az elsődleges kutatási feladat olyan nyelvtan kidolgozása, amely akár a szemantikus elemzéssel párhuzamosan, egymást segítve képes szöveget elemezni. A szintaktikai és a szemantikai elemzés párhuzamosítása hatékonyabbá tenné a többértelműségek feloldását. Mindenképp valamilyen lexikális nyelvtanra gondolunk, így azokat a szabályokat, amelyeket most procedurálisan építünk a szemantikai elemzésbe, leíró szabályokként használhatnánk.


```

"Tegnap óta lázas, és fáj a torka, ma sokat köhögött. Egyéb panasza
nem volt."
Panaszfelvétel ("") [id=Test_20061106]
:eredményezi-->
  Panasz ("panasza") [id=Test_20061106_1.2.1.3]
  :ideje-->
    Nap ("ma") [id=Test_20061106_1.1.1.15]
    : dátuma--> 2006.11.06.
  :zárt-e--> true
  :eleme-->
    Köhögés_Állapot ("köhögött") [id=Test_20061106_1.1.1.17]
    Fájdalom ("fáj") [id=Test_20061106_1.1.1.8]
    Láz ("lázás") [id=Test_20061106_1.1.1.4]
  Köhögés_Állapot ("köhögött") [id=Test_20061106_1.1.1.17]
  :aktora-->
    Páciens ("") [id=Test]
  :kvalitatív_száma-->
    Sok ("sokat") [id=Test_20061106_1.1.1.16]
  :ideje-->
    Nap ("ma") [id=Test_20061106_1.1.1.15]
    : dátuma--> 2006.11.06.
  Fájdalom ("fáj") [id=Test_20061106_1.1.1.8]
  :ideje-->
    Időintervallum_tól ("óta") [id=Test_20061106_1.1.1.3]
    :kezdőpontja-->
      Nap ("Tegnap") [id=Test_20061106_1.1.1.1]
      : dátuma--> 2006.11.05.
  :lokalizációja-->
    Torok ("torka") [id=Test_20061106_1.1.1.12]
  Láz ("lázás") [id=Test_20061106_1.1.1.4]
  :ideje-->
    Nap ("Tegnap") [id=Test_20061106_1.1.1.1]
    :kezdi-->
      Időintervallum_tól ("óta") [id=Test_20061106_1.1.1.3]
      : dátuma--> 2006.11.05.
:elszenvedője-->
  Páciens ("") [id=Test]
:ideje-->
  Nap ("") [id=Test_20061106_time]
  : dátuma--> 2006.11.06.

```

4. ábra Egy példaszöveg reprezentációja a program által adott karakteres kimeneten.

A jelenleg használt Protégé ontológiaszerkesztő adottságai sok mesterséges megoldást kényszerítettek ránk, amelyek kissé elbonyolították az ontológia szerkezetét. Majd ha a MEO projekt során fejlesztett ontológiaszerkesztő [6] használatra kész lesz, sokkal természetesebb ontológiát használhatunk.

Valódi alkalmazás rendkívüli méretű ontológiát igényel – egy ilyen megkonstruálása külön több éves projekt. Olyan megoldáson gondolkodunk, amely – legalábbis néhány alkalmazás esetén – nem kíván teljes ontológiát; illetve az alkalmazás használata során épülne az ontológia. A legfontosabb ilyen alkalmazás olyan keresőprogram kifejlesztése lenne, amely a kereső-kifejezés jelentésének megfelelő találatokat adna.

Bár a bemutatott rendszer is bizonyítja módszerünk alkalmazhatóságát, számos kérdés vehető fel bonyolultabb szövegek elemzésénél. A módszer továbbfejlesztése is előttünk áll.

Bibliográfia

1. Alberti G., Balogh K., Kleiber J., Viszket A.: A totális lexikalizmus elve és a GASG nyelvtan-modell. In: Maleczki M. (szerk.): A mai magyar nyelv leírásának újabb módszerei V. SZTE, Szeged (2002) 193–218.
2. Kiryakov, A., Popov, B., Terziev, I., Manov, D., Ognyanoff, D.: Semantic annotation, indexing, and retrieval. *Journal of Web Semantics* 2 (2005)
3. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: *Ontology Library*. WonderWeb Deliverable D18. URL: www.loa-crn.it/Publications.html
4. Merényi Cs., Tihanyi L.: A MetaMorpho fordítóprogram projekt 2006-ban. In: MSZNY 2006, Szeged (2006)
5. Nagypál, G.: Improving Information Retrieval Effectiveness by Using Domain Knowledge Stored in Ontologies. *OTM Workshops 2005*, LNCS 3762, Springer-Verlag (2005) 780-789.
6. Szakadát I., Szóts M., Gyepesi Gy.: MEO - Ontology Infrastructure. In: Magyar G., Knapp G., Wojtkowski, W., Wojtkowski, G., Zupancic, J., Wrycza, S. (eds.): *Advances in Information Systems Development: New Methods and Practice for the Networked Society*, Proceedings Information Systems Development, Springer, in press
7. Szóts M., Lévay Á.: Szerepfogalmak az ontológiában – az OntoClean metodológia továbbfejlesztése. In: MSZNY 2005, Szeged (2005)
8. Tomassen, S.L.; Gulla, J.A.; Strasunskas, D.: Document Space Adapted Ontology: Application in Query Enrichment. *Proceedings of 11th International Conference on Applications of Natural Language to Information Systems (NLDB'2006)*, Klagenfurt, Austria, LNCS 3999, Springer-Verlag (2006) 46-57.