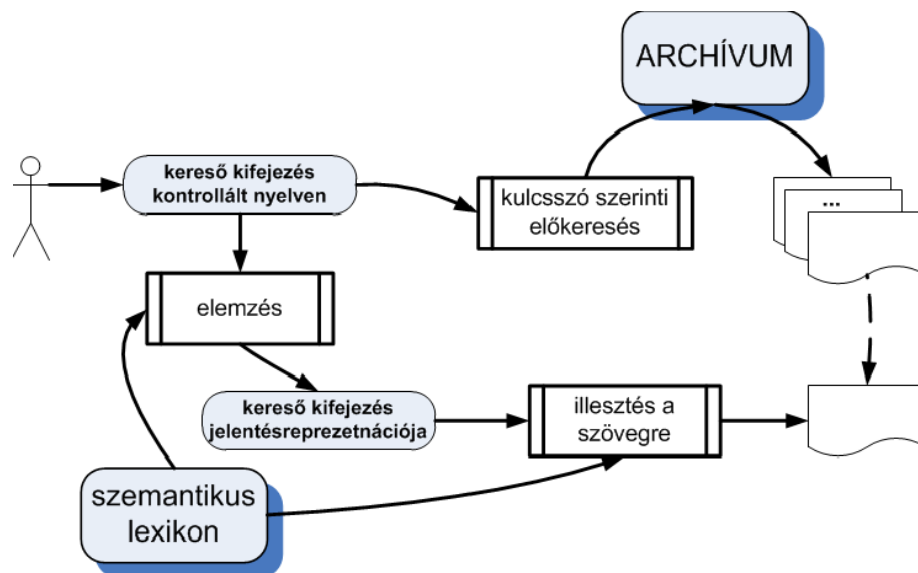


MASZEKER: szemantikus keresőprogram

Hussami Péter¹

¹Alkalmazott Logikai Laboratórium
1022 Budapest, Hankóczy j. u. 7
hussami@all.hu

Az Alkalmazott Logikai Laboratórium és a Szegedi Tudományegyetem Informatikai Tanszékcsoportjával, valamint Könyvtár- és Humán Informatíciótudományi Tanszékével közösen fejlesztette a MaSzeKer magyar szemantikus keresőt (projektazonosító: TECH_08_A2/2-2008-0092). A projekt eredményeképpen olyan szoftvermegoldás született meg, amely különböző forrásokat szemantikus reprezentációs alakra konvertál, és ezekben a reprezentációkban keres. Ennek előnye, hogy a rendszer így morfológiailag eltérő, de szemantikailag megegyező tartalmakat is össze tud kapcsolni. A rendszer áttekintő architektúrája az 1. ábrán látható.



1. ábra: A MASZEKER rendszer áttekintő architektúrája

Az ábrának megfelelően a releváns dokumentumok keresése a következő lépésekből áll:

1. a felhasználó egy kontrollált nyelven adja meg a kereső kifejezést,
2. szintaktikus és szemantikus elemzés előállítja kereső kifejezés jelentésreprezentációját,
3. szavak szerinti keresés előszűri az archívumot,
4. azokra a szöveg szegmensekre, amelyekben a szavak szerinti keresés találatai vannak, illeszti a kereső kifejezés jelentésreprezentációját.

Az MSzNy VIII konferencián tartott előadáson ([1]) ismertették a fenti elemek megvalósítására vonatkozó elméleti alapelveket, elsősorban egy módszert ajánlva a folyamat szemantikus reprezentáció fölé építésére. Az MSzNy VIII-IX konferenciákon tartott bemutatókon ([2, 3]) a rendszer első változatát mutattuk be, amely főnévi csoportos, illetve teljes (mondat) keresőkifejezéseken működött. Ezek a korábbi verziók szabadalmi igénypontokra lettek tervezve – a rendszer ennek megfelelően ezekre lett optimalizálva.

A jelen demó egy viszonylag eltérő nyelvezetű dokumentumhalmazon is működik, nevezetesen angol nyelvű klinikai ajánlásokon. Ezek, a szabadalmi igénypontokkal szemben lényegesen több segédigét / deontikus operátort tartalmaznak, és szabadabb szórendet használnak. Az ehhez tartozó szemantikus reprezentációt shallow parsing elv alapján dolgozzuk fel ([4]).

A demo használata: a felhasználó a kontrollált nyelven adhat meg keresőkifejezést. A keresőkifejezés több mondatból ill. főnévi kifejezésből állhat, egy kontrollált angol nyelven megfogalmazva. A megszorítások az egyértelműséget biztosítják, a tipikusan nehezen egyértelműsíthető fordulatokat akartuk kizárni. Legfontosabb korlátozások (a teljes definíció [5]-ban hozzáférhető):

- csak kijelentő mód, jelenidejű mondat írható,
- tiltott a mellérendelő mellékmondat (viszont a mondatok AND, OR kapcsolóval kapcsolhatóak, zárójelezhetőek),
- tiltott az alárendelő mellékmondat bármiféle lerövidítése (pl. igeneves utómódosítók),
- az alárendelő mellékmondatnak a „which” vonatkozó névmással kell kezdődnie, és ennek a közvetlenül megelőző főnévi csoport fejére kell vonatkoznia,
- tiltottak az igeneves előmódosítók,
- felsorolás, koordináció csak főnévi csoportok közt megengedett, ezeket a felhasználónak jelölnie kell.

A felhasználói interfész segíti a kontrollált nyelv szabályainak betartását, és a morfoszintaktikai elemzés eredménye alapján a rendszer ellenőrzi a szabályok betartását. Mivel teljes szabályrendszer nem ellenőrizhető, a generált jelentésreprezentációt grafikusan bemutatattik – ha szükséges, a felhasználó módosíthatja a keresőkifejezést. Ez a megjelenítés segíti egyértelműsíteni az egyes szavak jelentésének megállapítását is, ha több frame/synset van egy csomópontozat mellett, a felhasználó választhatja a megfelelőt.

A rendszer a keresőkifejezéshez illő frázisokat keres az igénypontok szövegében, és az eredményt a grafikus interfészen megmutatja, kiemelve azokat a szavakat, amelyekből álló frázist a keresőkifejezés egy szegmenséhez hasonlóknak talált. Míg a keresőkifejezés feldolgozásánál maximálisan törekszünk a pontos jelentésreprezentációra, a keresés fázisában az aktuális szövegrészlet vizsgálatánál csak azt ellenőrizzük, hogy jelentheti-e a keresőkifejezés valamely frázisát.

Hivatkozások

1. Szóts M., Csirik J., Gergely T., Karvalics L.: MASZEKER: projekt szemantikus kereső technológia kidolgozására. In: Tanács A., Vincze V. (szerk.): MSzNy 2010 – VII Magyar Számítógépes Nyelvészeti Konferencia, Szegedi Tudomány Egyetem, Szeged (2010) 159–167
2. Hussami P.: MASZEKER: szemantikus kereső program. In: Tanács A., Vincze V. (szerk.) MszNy 2011 – VIII. Magyar Számítógépes Nyelvészeti Konferencia, Szegedi Tudományegyetem, Szeged (2011) 321–322
3. Hussami P.: MASZEKER: szemantikus kereső program. In: Tanács A., Vincze V. (szerk.) MszNy 2013 – IX. Magyar Számítógépes Nyelvészeti Konferencia, Szegedi Tudományegyetem, Szeged (2013) 302–304
4. Gyarmathy Zs., Simonyi A., Szóts M.: Felszíni szintaktikai elemzés és a jóindulatú interpretáció elve információ-visszakeresésben. In: MszNy X – X. Magyar Számítógépes Nyelvészeti Konferencia, megjelenés előtt
5. A kontrollált nyelv definíciója <http://www.maszeker.hu/?page=download>