

Természetes nyelvi interfész adatbázisok lekérdezéséhez

Vajda Péter

Nyelvtudományi Intézet, Korpusznyelvészeti Osztály
vajda@corpus.nytud.hu

Kulcsszavak: adatbázis, kérdés, keresés, ontológia, Szemantikus Web, RDF

Az Interneten található folyamatosan bővülő információmennyiségből a releváns adatok megtalálása egyre nehezebb feladat. Ahhoz, hogy a keresett információhoz gyorsan hozzájussunk, szükséges – részben a természetes nyelv többértelműségei miatt –, hogy ne csak karaktereket, hanem a gép számára jelentést hordozó adatokat dolgozzunk fel. Így a felhasználó keresésének sem csupán kulcsszavakból kell állnia, hanem számára egyszerűbb és kérdését pontosabban kifejező mondatokat is megfogalmazhat. A *Szemantikus Web* kezdeményezés alapján mindehhez szükség van egy *ontológiára*, ami tulajdonképpen a gép számára megfogalmazott „világismeret”.

A mai keresők másik hiányossága, hogy nem férnek hozzá a nagyrészt publikus adatbázisokból álló ún. *Deep Web*hez. Az ilyen adatbázisokban szereplő adatok igen hasznosnak bizonyulhatnak olyan egyszerű kérdések megválaszolásában, amelyekre adható válasz megtalálása kulcsszavas kereséssel szinte lehetetlen. Pl.: „Hány Oscar-díjat nyert Woody Allen”, vagy „Melyik a 90-es évek legrágább filmje?”.

A Budapesti Műszaki Egyetemen és a Nyelvtudományi Intézetben végzett munka során a két problémát együtt kezelő rendszer megtervezésére, ill. megvalósítására tesztek kísérletet. Eddig végzett munkám során egy adatbázishoz kapcsolódó természetes nyelvű interfész architektúráját terveztem meg. Ennek feladata egy magyar nyelvű kérdés lefordítása egy vagy több olyan SQL queryvé, amely(ek)et a rendelkezésre álló adatbázis(ok) feldolgozhat(nak). A megvalósítandó komponens feladata a „szemantikai illesztés”, amely a feltett kérdés szintaktikailag elemzett alakjából, a kérdés interpretációját adja meg egy RDF-en alapuló logikai nyelven (ebből és a konkrét adatbázis sémájának ismeretéből fog végül előállni az SQL query). A szemantikai illesztés felhasznál egy ontológiát, ami a témakör fogalmi hierarchiájából és szavakat fogalmakra leképező szemantikai szabályokból áll. A fogalmi hierarchia elemei entitások, valamint közöttük fennálló relációk, a szemantikai szabályok pedig az egyes szavakhoz rendelik hozzá a megfelelő típusú entitást (fogalom) vagy relációt (predikátum). Az ontológia felépítéséhez szükség volt a kiválasztott témakörben előforduló lehetséges kérdések összegyűjtésére és nyelvtani vizsgálatára is. A szemantikai illesztés során a kérdésben lévő szavakhoz a rendszer megadja a hozzájuk tartozó fogalmakat és predikátumokat úgy, hogy a végeredmény a kérdésnek megfelelő logikai kifejezés legyen. Előadásomban a szemantikai illesztés folyamatát szeretném néhány példán keresztül bemutatni az erre kialakított program segítségével.