

KONTROLLÁLT TERMÉSZETES NYELVŰ LEKÉRDEZÉS WEBES ADATBÁZISOKHOZ

*Mészáros Tamás, meszaros@mit.bme.hu**

*Dobrowiecki Tadeusz, dobrowiecki@mit.bme.hu**

*Kiss Margit, kiss.margit@webit.hu***

**BME Méréstechnika és Információs Rendszerek Tanszék*

*** ELTE BTK Nyelvtudományi Doktori Iskola*

Az előadás keretében egy olyan módszert mutatunk be, amely lehetőséget teremt webes adatbázisok és dokumentumtárak természetes nyelvű lekérdezésére. A kidolgozott eljárás alapján megvalósított kísérleti rendszer egy webes adatbázis magyar nyelvű lekérdezésére szolgál. A módszer alapja egy ún. kontrollált természetes nyelvű lekérdező felület, amely egy web űrlapon feltett kérdéseket SQL kifejezéseké fordít. Az elkészült implementáció ezt mesterséges intelligencia, XML és web technológiák segítségével valósítja meg.

Az elektronikusan tárolt információk hatékony és egyszerű elérése egy komplex informatikai feladat, amely a web rendszereiben tárolt és elérhető információk dinamikus bővülésével a felhasználók egyre szélesebb köre számára mindennapi tevékenységgé vált. Napjainkban legszélesebb körben alapvetően statisztikai és katalogizáló módszereket alkalmaznak e feladat megoldására, melyek azonban a tárolt információ mélyebb megértése nélkül nem képesek a feladat kielégítően jó megoldására. Ezért egyre nagyobb hangsúly kerül mind a tárolás mélyebb szemantikai szintjeinek kidolgozására (W3C Szemantikus Web), mind az elérési módszerek közelítésére az emberi szint felé (új lekérdező nyelvek, természetes nyelvfeldolgozás). Ezek az újfajta módszerek azonban nagy részben még kidolgozatlanok, sok esetben a megvalósításukhoz szükséges elméleti eredmények is hiányoznak.

A BME Méréstechnika és Információs Rendszerek tanszékén az elmúlt években több hazai és nemzetközi projektben is vizsgáltuk az információ-elérés sajátosságait, megoldási javaslatokat tettünk ezen rendszerek minőségének javítására. Az elmúlt évben a természetes nyelvek gépi elemzésének olyan alkalmazásait vizsgáltuk, melyek lehetővé tehetik az átlagos felhasználó számára egyszerűbben használható, természetes nyelvű lekérdező felületek megvalósítását.

A bemutatásra kerülő rendszer lényegi eleme egy ún. kontrollált természetes nyelvű lekérdező rendszer. Ez az általános természetes nyelvű lekérdező rendszereket annyiban egyszerűsíti – és teszi így a feladatukat megoldhatóvá –, hogy a lekérdezés során korlátozza a természetes nyelv használatát. Egy ilyen kontrollált nyelv a számítógép által könnyebben feldolgozható, egyértelmű jelentés rendelhető a lekérdezéshez, és így elkészíthető annak gépi fordítása (alkalmazásunkban SQL kifejezéssé). Ez a módszer ugyanakkor a felhasználó számára lehetővé teszi természetes nyelvű kérdések megfogalmazását, a hagyományos űrlapos lekérdezéseknél közérthetőbb és egyszerűbb módon.

Az előadás során a teljes rendszer technikai bemutatása mellett kitérünk a kontrollált nyelvi elemzés sajátosságaira, valamint a korlátozott természetes nyelv számítógépes generálásának lehetőségeire is. Az elkészült implementáció az ELTE doktori programjában elkészített magyar főnévi vonzattár webes adatbázisának bonyolult lekérdező felületét cseréli le egy természetes nyelvű felületre. A kísérleti rendszer kipróbálható a <http://www.webit.hu/fonevlista/?l=askme> című weblapon.