TÖBBNYELVŰ INFORMÁCIÓ-KERESŐ RENDSZEREK

Douglas W. Oard
a Maryland-i Állami Egyetem tanára
http://www.glue.umd.edu/~oard/

A beszámolót készítette:
Moldován István

A maryland-i egyetem tanárát Komlódi Anita, washingtonban dolgozó magyar könyvtáros kollégánk tanácsára és segítségével hívtuk meg egy előadás megtartására. Douglas Oard főbb kutatási területe az információ keresési tevékenysében az új technológiák alkalmazása. Jelen előadásában ennek egyik fontos részéről a többnyelvű információkeresésről számolt be hallgatóságának.

Kiindulásként ismertette az Internet várható fejlődési tendenciáit a következő 5 évben. Becsléseik szerint 2005-re az Internet felhasználók millió között kisebségbe kerülnek (32 %) az angol anyanyelvűek, míg jelentős többséget fognak alkotni a kínai, spanyol, japán és más nyelvű felhasználók. Egy másik tudományos vizsgálaton alapuló becslés alapján Európában az angol nyelvű oldalak számát 2005-re utoléri a nem-angol nyelvű Web-oldalak száma. Következésképpen egyre nagyobb igény várható a többnyelvű keresőrendszerek iránt (Cross-Language Search). Erre szüksége lesz mind a többnyelven beszélőknek, akik inkább egy nyelven keresnek jobban, mind a csak egy nyelvet beszélőknek, hogy a számukra idegen nyelveken is releváns információkat találjanak.

A többnyelvű információkeresés egy összetett, többlépcsős folyamat, amely nemcsak a keresőkérdés megfogalmazásából, más nyelvekre való átültetéséből, a kapott dokumentumok értékeléséből, szelektálásából áll, hanem azok megszerzéséből, lefordításából. Az előadásban ismertetett projekt azonban ezekre az utóbbi lépésekre, a dokumentum idegennyelvű tartalmának megismerésére már nem tér ki, csak a keresésre.

A többnyelvű információkeresés alapja megfelelő szótárak felhasználása. Ezek segítéségvel határozza meg a kereső azokat a szinonimákat, amelyek a saját nyelvén meghatározott keresőfogalmakat legjobban leképezik. A bemutatott keresőrendszer nemcsak szó-megfeleltetéseket, de környezetben lévő kulcsszavakat (KWIC index) is képes megadni a megfelelő fogalmak kiválasztása érdekében.

Egy 8 országból álló projekt keretén belül (The Cross-Language Evaluation Forum - CLEF) két éve indult meg a fejlesztés egy többnyelvű információkereső alkalmazás létrehozására. A fejlesztést eddig 8 európai nyelvre valamint a kínaira és a japánra alkalmazták.

A projekt keretében előre kiválogatott dokumentumokkal és felhasználókkal teszteltették a

megadott kérdésekre kapott válaszok relevanciáját. Az Európai Unió támogatta az alkalmazás tesztelését, értékelését, amelyet a csatlakozott intézmények végeztettek el kísérleti személyekkel.

Az előadás végén az előadó külön felhívta a hallgatóságot, segítsen partnereket keresni a magyar nyelvű teszteléshez is. A projektről bővebb információ az alábbi címeken található:

http://clef.iei.pi.cnr.it
http://terral.lsi.uned.es/iCLEF

A rendezvény alatt a tolmácsolásban Szalóki Gabriella, az OSZK munkatársa segített, szakértő munkáját ezúton is köszönjük.

Teljes angol nyelvű PowerPoint bemutató: clef-eng.ppt
Rövidített magyar nyelvű PowerPoint bemutató: clef-hun.ppt