



Premijerina nagrada i Zlatno Teslino jaje u kategoriji Akademske institucije

Jednostavno i brzo do propisa

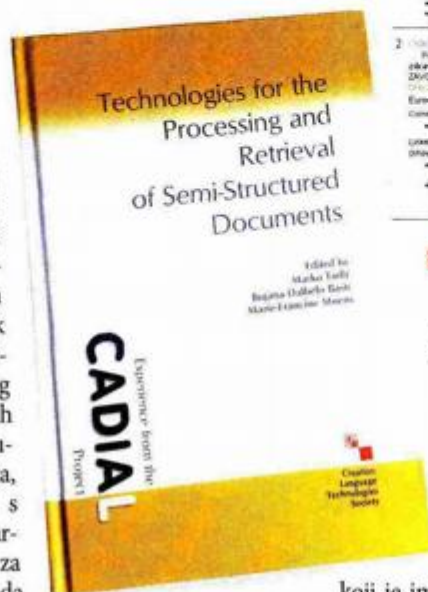
Nakon zapaženog uspjeha na Vidi e-novation natječaju primjena CADIAL-ovih modula već se lagano širi na regionalna tržišta

priređio: Goran Novačić

U poznavanje sa zakonima i propisima postalo je olakšano pojavom Interneta i weba. Još je lakše do njih doći putem dobro organiziranih i moćnih pretraživača takve specijalne namjene. Jedan od takvih jest i sustav CADIAL, dobitnik Premijerine nagrade izdavačke kuće Vidi te Zlatnog Teslinog jajeta u kategoriji Akademske institucije. Kako nam je istaknula voditeljica nagrađenog projekta, prof. dr. Bojana Dalbelo Bašić s Fakulteta elektrotehnike i računarstva Sveučilišta u Zagrebu, ideja za projekt nastala je 2005. godine kada je Hidra (Hrvatska informacijsko-dokumentacijska referalna agencija – www.hidra.hr) izrazila potrebu da se postupak označavanja dokumenata ključnim riječima iz pojmovnika Eurovoc (http://europa.eu/eurovoc/) automatizira. Potreba da se dokumenti na taj način sadržajno označe predstavlja usklađivanje Hrvatske sa standardima Europske unije jer je Eurovoc službeni pojmovnik EU i postoji na svim jezicima Europske unije, ali i na drugim jezicima pa tako i na hrvatskom.

Potpuno automatski sustav

Označavanje dokumenata, ističe Dalbelo Bašić, je izuzetno intelektualno naporan posao, zahtijeva razumijevanje sadržaja dokumenta (propisa, zakona ili sl.) i poznavanje strukture Eurovoca (oko 6000 pojmova organiziranih u 8 razina). To je skup i vremenski zahtjevan posao koji rade visoko obrazova-



The screenshot shows the CADIAL search engine interface. At the top, there's a search bar and navigation links. Below, search results are displayed for the query 'Zlatno Teslino jaje'. The results list two documents with their titles, authors, and publication dates. A sidebar on the right shows a hierarchical tree of Eurovoc categories, with 'Health' and 'Health policy' highlighted.

Baza: 20.000 označenih službenih dokumenata RH slobodno, javno dostupno i pretraživo

ni stručnjaci iz Hidre. Do 2005. godine postojalo je svega stotinjak dokumenata, "ručno" označenih pravnih propisa. Hidra, Fakultet elektrotehnike i računarstva i Filozofski fakultet zajednički su započeli projekt AIDE

koji je imao za cilj napraviti sustav

za poluautomatsko označavanje dokumenata. Projekt je uspješno završen, a njegov nastavak, projekt CADIAL, imao je zadatak napraviti potpuno automatski sustav kojim će se označiti sva službena dokumentacija Republike Hrvatske i zatim napraviti tražilica nad takvom zbirkom označenih dokumenata koja će biti javno dostupna svim građanima RH. Svi su ti ciljevi u potpunosti ostvareni i danas je više od 20 000 označenih službenih dokumenata RH slobodno, javno dostupno i pretraživo, na adresi cadial.hidra.hr. U projektu CADIAL je, osim gore navedenih institucija, sudjelovala i grupa profesorice Marie-Francine Moens, s Katoličkog Sveučilišta u Leuvenu koje je jedno od najboljih europskih sveučilišta. Na pitanje jesu li zadovoljni dosadašnjim stupnjem razvoja projekta CADIAL, Dalbelo Bašić odgovara potvrdno te dodaje da su i ponosni postignutim rezultatima.

"Željela bih napomenuti da je tražilica samo jedan od rezultata međunarodnog projekta CADIAL. Uz tražilicu, napravili smo sustav za automatsko dodjeljivanje ključnih riječi iz Eurovoca, sustav za morfološku normalizaciju hrvatskoga jezika, sustav za automatsku ekstrakciju višeriječnih termina (ili fraza) iz dokumenta te smo izdali knjigu "Technologies for the Processing and Retrieval of Semi-Structured Documents" i publicirali cijeli niz vrlo vrijednih znanstvenih radova. K tome smo ostvarili izvrsnu međuinstitucijsku i međunarodnu suradnju. Kao voditeljica, posebno sam ponosna na talentirane mlade ljude s FER-a koji su s puno entuzijazma i kreativnosti radili na ovom projektu

(Jure Mijić, Jan Šnajder, Frane Šarić, Artur Šilić, Saša Petrović, Davor Delač), a i zahvalna sam kolegicama iz Hidre, Nedi Erceg i Maji Cvitaš te njihovim suradnicima koji su stvarali bazu pravnih dokumenata. Također sam zahvalna koordinatrici prof. Marie-Francine Moens i kolegi prof. Marku Tadiću. Popis svih rezultata, opis knjige i popis radova dostupan je na adresi www.cadial.org. Što se tiče same tražilice, ona je funkcionalna i javno dostupna.

Tražilica služi svim građanima Hrvatske, ali i šire, budući da je, zahvaljujući višezječnosti pojmovnika Eurovoc, moguće pretraživati i na engleskom jeziku. Međutim, kako ističe Dalbelo Bašić, na tražilici se svakodnevno radi: uklanjaju se neke manji nedostaci, optimiziraju neki parametri tražilice i poboljšava njezin rad na temelju sugestija korisnika. Tako će uskoro web-sučelje tražilice dobiti novi dizajn i još neke nove opcije.

U svrhu javnog doprinosa

Zanimalo nas je dakako i koliko je do sada otprilike uloženo u razvoj same tražilice. Profesorica Dalbelo Bašić ističe kako su projekt CADIAL financirali Vlada Flandrije (Belgija) i Ministarstvo znanosti obrazovanja i športa. Vrijednost cijelog projekta jest 160 000,00 eura, od čega je MZOŠ sudjelovao s 20 000,00 eura. Inače, kako dodaje, projekt CADIAL je neprofitabilan i njegov je cilj bio napraviti doprinos koji će biti javan, koji će doprinijeti približavanju Hrvatske standardima EU u dokumentarističkoj domeni.

Međutim, važno je napomenuti da je FER, kroz rad na ovom projektu, stekao ekspertno znanje i iskustvo te je izgradio osnovne module koji mogu biti dijelovi bilo kojeg drugog komercijalnog sustava. Tako je na primjer sustav za morfološku normalizaciju hrvatskoga jezika, koji nam omogućava da za zadani upit pronađemo dokumente u kojima se nalazi bilo koji oblik riječi iz upita (bez obzira na rod, broj i padež), primjenjiv i u drugim sustavima za obradu hrvatskoga jezika.

The screenshot shows the CADIAL website interface. It features the CADIAL logo, a search bar, and a list of search results. At the bottom, there's a call to action: 'www.cadial.org Na web stranicama Cadiala pronaći sve potrebne informacije o tražilici'.