

Cena SAV

za výsledky v oblasti spolupráce s vysokými školami

Výskum nových prístupov a metód na odhaľovanie a spracovanie znalostí a informácií v prostredí heterogénnych informačných zdrojov patrí v súčasnosti k intenzívne sa rozvíjajúcim odborom aplikovanej informatiky. Súvisí to s narastajúcim množstvom informácií dostupných na webe. Práve neustále sa zväčšujúce množstvo dostupných dát a informácií si vyžaduje vývoj pokročilých metód na prácu s nimi. Medzi najdôležitejšie problémy, ku ktorým výskumníci prispeli, patrí oblasť automatickej anotácie, personalizovanej prezentácie a v nemalej miere aj orientácia na prácu so slovenčinou.

Výskum sa realizoval v rámci riešenia úlohy výskumu a vývoja štátneho programu „Budovanie informačnej spoločnosti“ č. 1025/04. Riešitelia na základe výskumu navrhli a implementovali softvérové nástroje, ktoré sa overovali v rámci dvoch pilotných aplikácií orientovaných na získavanie informácií a znalostí v doméne pracovných ponúk. Vyvinuté softvérové nástroje boli integrované do niekoľkých celkov, ktoré spracovávajú informácie od ich rozpoznania na webe až po prezentovanie užívateľovi v takej forme, ktorá čo najviac zodpovedá jeho profilu. Výskum nemal za úlohu vyvíjať nový vyhľadávač, ale prispieť k zlepšeniu procesu získavania informácií z webu.

Výsledky výskumu sa dajú rozdeliť do troch základných oblastí: vyhľadávanie informácií v prostredí heterogénnych zdrojov; spracovanie informácií takým spôsobom, aby mali čo najviac homogénnu štruktúru, prispôbenie výsledkov vyhľadávania používateľovi.

Získavanie a anotovanie ponúk

Internet obsahuje množstvo informácií, ktoré sú uložené väčšinou vo forme html dokumentov. Na automatizované získavanie dát z internetu sa používajú metódy známe ako webcrawling, prípadne spidering. Bežne používané metódy *wecrawlingu* a *spideringu* riešitelia doplnili o nástroje RIDAR a ERID, ktoré umožňujú vyhľadávanie zamerané na cieľovú doménu. Nástroj ERID používa na vyhodnotenie relevantnosti nájdených stránok metódu jednoduchého neurónu. Ďalším nástrojom, ktorý sa používa na extrakciu informácií a na sledovanie zmien obsahu webových stránok je WRAPPER. Pri jeho používaní sa využíva podobnosť štruktúry stránok v rámci jedného webového sídla. Pomocou nástrojov na semiautomatickú anotáciu (OnTeA) sa v dokumentoch určujú kľúčové slová, ktoré sa použijú pri personalizovanom prezentovaní ponúk používateľom.

Určenie najlepšej ponuky

Reťazec služina vyhľadávanie ponúk podľa používateľských preferencií, napríklad na nájdenie najrelevantnejších objektov pre používateľov s podobným profilom, ktoré sa

Dostal ju široký kolektív pracovníkov organizácií SAV, vysokých škôl a podnikateľského sektora pod vedením prof. Ing. Pavla Návrata, PhD., z FIIT STU (koordinátor projektu), prof. Ing. Márie Bielikovej, PhD., z FIIT STU, doc. Ing. Ladislava Hluchého z ÚI SAV, prof. RNDr. Petra Vojtáša, DrSc., z UPJŠ Košice a Ing. Pavla Bartoša z firmy Softec, s. r. o. za prácu: *Pokročilé metódy pre získavanie, organizovanie a udržiavanie znalostí v prostredí heterogénnych informačných zdrojov.*

reprezentujú pomocou fuzzy množín. Tento prístup sa kombinuje s vyhľadávaním pomocou indexovania a fulltextového vyhľadávania. Nástroje navrhnuté a implementované v rámci tohto reťazca vykonávajú prácu skrytú používateľom. Ide o predspracovanie získaných údajov rôznymi spôsob-



Zástupcovia širokého kolektívu, ktorý získal Cenu SAV za výsledky spolupráce s vysokými školami.

mi, ktoré umožnia neskôr vyhľadávanie informácií z rôznych zdrojov podľa rovnakých preferencií a profilov. Táto skupina nástrojov obsahuje veľké množstvo implementovaných a otestovaných softvérov.

Adaptívna prezentácia ponúk

Táto skupina metód a implementovaných nástrojov je z hľadiska používateľa najviditeľnejšia. Bez dobrej prezentácie výsledkov, ktoré sa urobili nástrojmi v prvých dvoch skupinách, by sa výsledky nedali použiť v praxi. Metódy a nástroje v tejto skupine umožňujú:

- *Modelovanie používateľa* – vytváranie zmysluplnej charakteristiky používateľa pomocou zaznamenávania jeho akcií a následného sémantického vyhodnocovania získaných dát;
 - *Personalizované vyhľadávanie* – vyhľadávanie, pri ktorom sa ako vstupné dáta vkládajú charakteristiky používateľov;
 - *Personalizovaná navigácia* – navigáciu pomocou adaptívnych fazetových sémantických prehľadávačov.
- Pri vytváraní výsledkov riešenia sa použili, v súlade so zmluvou, dostupné softvérové prostriedky a knižnice distribuované pod

licenciou OpenSource. Táto licencia sa zachováva pri ďalšom šírení softvéru.

■ Výsledkom projektu je vytvorený portál pracovných ponúk, ktorý primárne slúži na zadávanie a editovanie pracovných ponúk. Tento portál sa stal integračným nástrojom pre metódy a softvéry vyvinuté v rámci projektu. Navrhnuté metódy predstavujú prínos v procese aktívneho vyhľadávania informácií z rôznych zdrojov na internete. Zvyšujú šance ich nájdenia uchádzačom o zamestnanie podľa ich osobnej charakteristiky a preferencií a v neposlednej rade zvyšujú šance zamestnávateľov nájsť vhodného kandidáta na ponúkanú pozíciu. Výskum sa robil so zreteľom na sprístupnenie vyhľadávacích služieb čo najširším vrstvám obyvateľov pri hľadaní zamestnania. Aplikáciou výsledkov výskumu je možné značne zlepšiť šance na nájdenie vhodnej ponuky v rozsiahlom a heterogénnom prostredí zdrojov ponúkajúcich zamestnanie, od portálov pracovných agentúr, cez firemné portály a web stránky, až po inzeráty ponúkajúce zamestnanie u malých zamestnávateľov. Výsledky výskumu sa konzultovali s odborníkmi z Úradu práce a sociálnych vecí, ktorí sa zúčastňovali priebežných oponentúr počas riešenia projektu.

Aplikované výsledky – softvérové produkty:

Výsledkom projektu je súbor softvérových nástrojov a tri reťazce nástrojov, ktoré sa integrovali do portálu pracovných ponúk v rámci dvoch pilotných aplikácií. Vyvinuté metódy a ich softvérová implementácia sú popísané v dokumentácii k projektu NÁZOU na stránke (<http://nazou.fiit.stuba.sk/home/?page=methods>).

Spolupráca nástrojov v rámci troch reťazcov je v časti Methods, link na *Collaboration of these methods* (<http://nazou.fiit.stuba.sk/home/?page=collaborations>).

V tejto časti sa nachádzajú aj demonštračné videá k všetkým trom vytvoreným reťazcom, ktoré patria k aplikačným výstupom projektu.

Ďalšie informácie o úlohe a jej výsledkoch sa nachádzajú na <http://nazou.fiit.stuba.sk/home/>.

■ Ing. IVANA BUDINSKA
Ústav informatiky SAV