Názov výsledku: Vyhľadávanie entít a kategorizácia dopytov vyhľadávania
Anglický názov výsledku: Entity Search and Web Queries Categorization
Číslo projektu a agentúra: Spolupráca s podnikateľskou sférou
Hlavní odberatelia: Magnetic Media Online, Inc., USA
Autori: Ladislav Hluchý, Michal Laclavík, Marek Ciglan, Martin Šeleng, Štefan Dlugolinský
Celková cena zákazky: 28 261€.

Dielo predstavuje súbor metód a nástrojov na spracovanie anglickej Wikipédie ako znalostnej bázy pre podporu vyhľadávania entít. Anglická Wikipédia obsahuje 40GB textu a 6 miliónov wiki stránok, ktoré reprezentujú abstraktné ale aj reálne koncepty alebo entity. Tieto dáta boli spracované na základe techník z oblasti vyhľadávania informácií (information retrieval) ako aj spracovania komplexných sietí a grafových algoritmov. Ďalej bol vytvorený prototyp na kategorizáciu dopytov vyhľadávania. Firma Magnetic sa zameriava na Internetovú reklamu, kde je potrebné kategorizovať množstvo dopytov vyhľadávania do kategórií za účelom modelovania záujmu používateľov. Teda napríklad ak používateľ vyhľadáva dopyt „Sochi 2014“ je potrebné zaradiť používateľa do kategórie „Sport/Olympics“ a „Travel/Travel Destinations”. Firma Magnetic musí spracovať 5000 dopytov za sekundu. Na kategorizáciu dopytov bola použitá Wikipédia, kde sa najprv vyhľadajú entity, ktoré reprezentujú daný dopyt a následne sa tieto entity zaradia do kategórie. Metóda kategorizácie dopytov bola overená na dopytoch firmy Magnetic ako aj štandardnej dátovej sade pre túto úlohu zo súťaže KDD 2005, kde boli dosiahnuté výsledky porovnateľné z najlepšími dostupnými riešeniami, ktoré však používajú internetový vyhľadávač. Bolo teda vytvorené inovatívne riešenie, ktoré dosahuje výborné výsledky, používa Wikipédiu namiesto korpusu celého internetu a dokáže kategorizovať niekoľko tisíc dopytov za sekundu. Vytvorené riešenia sú aplikovateľné aj na iné úlohy vyhľadávania entít, napríklad pri vyhľadávaní v podnikoch (enterprise search).

Hlavné scientometrické výstupy:

[1] LACLAVÍK, M. - CIGLAN, M. Towards entity search: Research roadmap. In 8th Workshop on Intelligent and Knowledge Oriented Technologies : WIKT 2013 proceedings. Eds. Babič, F., Paralič, J. - Košice : Centre for Information Technologies, Technical University in Košice, 2013, p. 161-166. ISBN 978-80-8143-128-9.