Názov výsledku: Connectik - Sémantické služby pre sociálnu sieť
Anglický názov výsledku: Connectik - Semantic services for social network
Autori: Martin Šeleng, Štefan Dlugolinský
Typ a číslo projektu: Zakázka na externý výskum
Koordinátor projektu: Ladislav Hluchý
Hlavní odberatelia: SCS GROUP SA
Získané finančné prostriedky: 40 000 EUR

Anotácia výsledku: V projekte Connectik bolo našou úlohou dodať sémantické služby pre rovnomennú sociálnu sieť ľudí a firiem podľa požiadavky „automaticky odporúčať používateľom iných používateľov na základe obsahu ich profilov a aktivity v sociálnej sieti, t.j. obsahu zverejnených príspevkov, udalostí z kalendára a komentárov v diskusiách“.
Naše riešenie spočívalo v sémantickom modelovaní jednotlivých používateľov. Sémantický model bol vektor váh kategórií z preddefinovanej taxonómie. Podľa toho, akým témam sa používateľ vo svojich príspevkoch venoval, tak sa aj modeloval jeho vektor kategórii. Pomocou modelov a kosínusovej podobnosti sme vedeli následne porovnávať jednotlivé vektory medzi sebou a tým aj odporúčať podobných používateľov, či skupiny používateľov. Pri skupinách sme v modelovaní postupovali podobne ako u individuálnych používateľov s tým rozdielom, že na vytváranie modelu skupiny sme nepoužívali modely príspevkov, ale už len modely používateľov danej skupiny. Kategorizáciu textu podľa taxonómie mal na starosti kategorizačný systém, ktorý sme vytvorili spracovaním Wikipédie, Freebase databázy a DBpédie. Kategorizačný systém bežal nad systémom Hadoop (paralelne a distribuovane). Ďalším krokom bolo namapovanie taxonómie na takto obohatené stránky. To sme dosiahli tak, že sme jednotlivým kategóriám taxonómie manuálne priradili niekoľko príkladov stránok (seed pages). Tieto príklady slúžili na nájdenie ďalších príbuzných stránok. Stránky namapované na taxonómiu sme nakoniec zaindexovali a index využívali ako znalostnú bázu pri kategorizácii textu.

Hlavné scientometrické výstupy:
1. LACLAVÍK, Michal - CIGLAN, Marek - STEINGOLD, Sam - ŠELENG, Martin - DORMAN, Alex - DLUGOLINSKÝ, Štefan. Search query categorization at scale. In World Wide Web 2015 Companion : TargetAd2015 - Workshop on Ad Targeting at Scale. - Florence, Italy : ACM, 2015, p. 1281-1286. ISBN 978-1-4503-3473-0. Typ: AFC