

Využití vizualizačních technologií pro vyhledávání dokumentů

Petr Jenček

Posudek vedoucího diplomové práce

Zadání práce bylo motivováno problémem vyhledávání informací ve velkém množství textových dokumentů (ať jde o webové stránky, zprávy elektronické pošty, nebo např. možné plagiáty školních prací), které jsou v současné době ve valné většině zobrazeny jako seznam odkazů na relevantní dokumenty. Vzhledem k tomu, že člověk (uživatel) je schopen lépe vstřebat grafickou informaci než textovou, je možné uživateli vyhledané dokumenty zobrazit graficky podle toho, jak jsou si tyto dokumenty podobné. Cílem práce bylo zkoumat modely a metody vizualizace a metrik na vyjádření preferencí, implementovat systém a hlavně provést experimenty na několika datových sadách.

V práci byl proveden

- průzkum současných metod vizualizace a analýza jejich použitelnosti pro zobrazení dokumentů vyhledaných dle uživatelem zadaného řetězce;
- připravena testovací data a navrženy experimenty;
- analýza možné implementace a prototypová implementace s použitím SocioMap[®], dále už jen SocioMap;
- provedeny experimenty.

Formální úroveň není obzvlášť vysoká, software funguje za předpokladu, že máme klíč do SocioMap.

Lze ale konstatovat, že výsledky experimentů jsou zajímavé a publikovatelné.

K obhajobě předkládám uchazeči otázku, proč dotazy nad RCV nezpracoval také pomocí SocioMap, protože zde měl k dispozici i lidské hodnocení výsledků.

Nelze souhlasit s tvrzením str. 9¹³⁻¹⁵ „Majority of key words ...“ protože ty mohou mít veliký idf.

Editační vzdálenost 3.3.7.1 a 3.3.1.1 je velice malá.

Na závěr konstatuji, že práce splňuje požadavky na diplomovou práci a doporučuji ji k obhajobě.

V Praze 18.5.2009

Prof. RNDr. Peter Vojtáš, DrSc.
vedoucí DP