

Posudek oponenta diplomové práce

Název DP: **Shlukování textových dokumentů a jejich částí**
Diplomant: **Radoslav Zápotocký**

Obsah práce:

Předmětem diplomové práce (DP) byl návrh a implementace systému pro shlukování dokumentů a jejich částí (tj. celých dokumentů, kapitol, odstavců, vět), a vizualizaci. V první kapitole autor uvádí do problematiky fulltextového vyhledávání a shlukování, ve druhé pak analyzuje funkcionalitu implementovaného systému. Třetí kapitola se věnuje architektuře systému, čtvrtá příkladům použití systému a pátá práci uzavírá. Šestou kapitolu tvoří uživatelská příručka implementovaného systému a sedmou programátorská dokumentace.

Hodnocení:

Práce je poměrně dost odfláknutá. Sice se jedná o implementační diplomovou práci, ovšem rozsah implementace tomu neodpovídá. Text práce je velmi krátký, hlavní text (kap. 1-5) má pouhých 23 stran, přílohy 10 (uživ. příručka a programátorská dokumentace). Navíc některé části hlavního textu by měly být součástí programátorské příručky (celá kap. 3), či uživatelské příručky (kap 4), čímž by počet stran hlavního textu klesl na pouhých 11 stran. Zcela chybí rešerše a bibliografie čítá 12 odkazů, z toho pouze 5 na publikace (zbytek webové odkazy).

Samotný program také nepůsobí dojmem, že by na něm autor strávil alespoň rok práce, spíš působí jako polotovár. Zadání práce nevyžadovalo implementaci nových technik shlukování či modelů pro fulltextové vyhledávání, pouze implementaci léty prověřeného vektorového modelu a shlukovacích metod. Program nabízí základní vyhledávání a vizualizaci. Zajímavým prvkem je použití shlukování jako nástroje pro tvorbu abstraktů z dlouhých textů, nicméně tato funkcionalita není dále rozváděna, je ponechána uživateli. Celkově práci chybí nějaká přidaná hodnota, a to jak textu, tak programu. V textu by například mohl být detailněji popsán model abstraktování a jeho ověření sadou solidních experimentů. Program by mohl uživateli nabídnout nějakou komplexní funkcionalitu, ze které by šlo vyvozovat nějaké závěry. V současné podobě program nabízí pouze výpis základních pojmů souvisejících s vektorovým modelem DIS, případně primitivní vizualizaci (k čemu je užitečná obří Dot matrix???)

Podrobnější hodnocení, připomínky a otázky:

Pokud si uživatel program nainstaluje ve Windows 7 standardním způsobem do složky C:\Program Files\... a zkusí založit nový projekt, program spadne, neboť chce zapisovat data tam, kam je instalován – což se mu pochopitelně díky absenci práv zápisu nepovede. Program se tedy musí instalovat nestandardně do prostoru uživatelských dat (např. na plochu).

Angličtině je většinou slušná, nicméně na mnoha místech jsou gramatické chyby. Celkově je vidět, že práce byla psána ve spěchu.

Závěr:

Práce teoreticky mohla splnit vágní zadání a proto ji s výhradou doporučuji k obhajobě.

V Praze dne 17. května 2011

