

# Posudek oponenta diplomové práce

Autor práce: Radek Šmolík

Název práce: Vyhledávání expertů zadaného oboru z bibliografických vědeckých databází

## Obsah práce

Obsahem této diplomové práce (DP) je popis systému pro vyhledávání expertů na základě bibliografických dat a jeho implementace. DP se rovněž zabývá teorií vyhledávání vzorků v textu, indexací, tvorbou dotazů a algoritmy zjišťování významnosti uzlů v grafu.

## Kvalita řešení a dosažených výsledků

Osobně se domnívám, že kvalita řešení a dosažených výsledků je na spodní hranici přijatelnosti. Už jenom design a layout výsledné aplikace je odrazující a i kdybych připustil, že její jednoduchost je záměrná z důvodu přenositelnosti na mobilní telefony (což se ovšem na s. 60 uvádí až jako budoucí práce), zdá se mi přehnaná. Dokladem toho jsou např. obrázky 8.7 a 8.15 na s. 49 a 53, které jsou diplomové práce nedůstojné. Aplikace je zřejmě funkční, ale předložený dokument uživatele nikterak neláká k jejímu vyzkoušení. Teoretická část práce jde velmi málo do hloubky, čehož důkazem je např. ukázka taxonomie na s. 12.

## Formální úroveň

Formální úroveň diplomové práce je podprůměrná. Odsazování prvních řádek odstavců po prázdných řádkách je zbytečné, prázdná půlstrana na s. 19 je nevhodná. Některé číslované kapitoly či podkapitoly jsou neuvěřitelně krátké, např. kap. 3.6 nebo kap. 4. Před úvodem na s. 1 se nachází druhá strana obsahu s číslem 4. V angličtině napsaný abstrakt se ani vzdáleně neblíží správně napsanému souhrnu. Úsekům zdrojového kódu (např. na s. 36, 37 a v přílohách) by určitě pomohla barevně zvýrazněná syntax. První dva dokonce představují algoritmy, které by bylo vhodné pojmenovat a očíslovat. Určitě není vhodné pojmenovat kapitolu „Praktická část“ (kapitola 7) a uživatelská příručka z kapitoly 8 by zcela jistě měla patřit do příloh.

## Práce s literaturou

Práci s literaturou považuji za slabou. Je zde pouze 12 referencí, z toho jen 5 „newebových“. Reference č. 4 ovšem postrádá svazek i číslo časopisu a také stránkování. Někde odkazy na relevantní literaturu vyloženě chybějí, např. na s. 11 při zmínce o lemmatizaci a stemmingu by bylo určitě vhodné odkazovat na Porterův algoritmus nebo na s. 18 není vůbec diskutován faktor tlumení *d*.

## Splnění zadání

Zadání DP bylo splněno, ovšem použitelnost systému personálními agenturami je neprůkazná.

## Doplňující informace k práci

V diplomové práci se mimo řadu překlepů a pravopisných chyb vyskytují i hrubé věcné chyby. Např. není pravda, že „Indegree nám poskytuje pouze pořadí dokumentů“ (s. 20), že zkoumaná bibliografická data pokrývají roky 1995 až 2007 (s. 28) nebo že změna velikosti dat z 539 MB na 115 MB znamená redukci „o 21 %“ (s. 32) apod. Soubor „index.html“ na CD by měl být opravdový index s odkazy.

## Dotazy k práci

Jaká je výhoda užití XML + Lucene oproti importu dat do relační databáze?

Proč má epsilon na straně 37 zrovna tuto hodnotu?

Jak vznikne 37,5 % z pěti? (Prosím vysvětlete tabulky 7.3 a 7.4 na straně 40.)

Proč jste porovnání metod hodnocení autorů neřešil korelací?

Jak jsou řazení autoři na obrázku 8.10 na straně 50?

Navrhuji hodnocení známkou **dobře** a práci doporučuji k obhajobě.

V Plzni dne 2. června 2015

Západočeská univerzita v Plzni  
Fakulta aplikovaných věd  
katedra informatiky a výpočetní techniky

②

  
doc. Ing. Dalibor Fiala, Ph.D.