

Posudek oponenta diplomové práce

Autor práce: **Bc. Tomáš Hanke**

Název práce: **Tvorba datových zdrojů pro bibliometrická měření**

Student Bc. Tomáš Hanke se v průběhu své diplomové práce seznámil s bibliografickými databázemi *Google Scholar*, *ACM Digital Library* a *SciVerse Scopus*. Z těchto databází doporučil k dalšímu zpracování databázi Scopus, která poskytuje pro přístup k datům dvě vlastní API – Javascript API a RESTful API. V textu práce blíže popsal Javascriptové API a zmínil svůj neúspěšný pokus o získání přístupu k RESTful API. Následně ale student doporučil a získal data z webu Scopus využitím meta-vyhledávače, který byl implementován v jedné z dříve zadaných bakalářských prací. Z webu Scopus byla získána data o publikacích, autorech, afiliacích a státech autorů a jejich citacích, která byla uložena do relační databáze. Získaná data student dále analyzoval využitím programové knihovny JUNG s cílem vytvoření žebříčků publikací, autorů a států. Byla provedena hodnocení dle počtu publikací a dle počtu citací, citační grafy byly vyhodnoceny algoritmy in-degree, PageRank a HITS a také byly určeny dvojice nejvíce spolupracujících autorů a států. Výsledné žebříčky byly prezentovány a částečně diskutovány. Tímto student bez výhrad splnil zadání práce.

Student získal data z odpovědí webu Scopus na dotazy se slovy *internet*, *pagerank* a *telecommunication* a prezentované výsledky jsou tedy vázané pouze na sjednocení těchto tří oblastí. Zde předpokládám, že tuto volbu student provedl po konzultaci se svým vedoucím diplomové práce. Lehce udivující je pro mě skutečnost, že z odpovědí na slovo *pagerank* student získal více publikačních záznamů, než z odpovědí na zbylá dvě slova současně. Prezentovaný počet celkově získaných publikačních záznamů není příliš vysoký (55 tisíc) a v textu práce jsem nenalezl informaci, v jakém časovém rozpětí byly publikace publikovány.

Předkládaný text diplomové práce má velmi dobrou logickou strukturu a dopovídající obsah i rozsah. V textu bylo odkázáno na 23 relevantních zdrojů a text byl téměř bez gramatických chyb či překlepů, proto práci s literaturou a formální úroveň textu hodnotím jako výborné. Kvalitu řešení a dosažených výsledků hodnotím jako velmi dobrou, protože např. diskuse výsledků mohla, dle mého názoru, být obsáhlejší.

Dotazy k práci:


Co představuje „číslo generace“, které ukládáte v záznamech publikací?

Dokázal byste říci, proč bylo využitím slova *pagerank* v dotazu získáno více záznamů o publikacích, než využitím slov *internet* a *telecommunication*?

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

V Plzni 1. 8. 2014

**SOUHLASÍ
S ORIGINÁLEM**


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


Ing. Michal Nykl