

Posudek oponenta diplomové práce

Student: Bukovčák Jakub, Bc.
Téma: Extrakce informací z webových stránek (id 21836)
Oponent: Rychlý Marek, RNDr., Ph.D., UIFS FIT VUT

- 1. Náročnost zadání** průměrně obtížné zadání
Jedná se o průměrně obtížné zadání, podobné nástroje již existují a lze je použít v řešení (např. v práci použita knihovna Crawler4j).
- 2. Splnění požadavků zadání** zadání splněno
Zadání je splněno ve všech bodech bez výhrad.
- 3. Rozsah technické zprávy** je v obvyklém rozmezí
Od úvodu po závěr obsahuje technická zpráva 59 vysázených stran a její rozsah je tedy v obvyklém rozmezí.
- 4. Prezentací úroveň předložené práce** 75 b. (C)
Technická zpráva má logickou strukturu, která odpovídá postupu vývoje realizačního výstupu. Obsah je srozumitelný, jednotlivé části na sebe dobře navazují. Vytknout lze mnoho obsáhlých diagramů a tabulek analýzy a návrhu systému v kap. 4 s příliš stručným popisem - např. tabulky se strukturovaným popisem případů užití v kap. 4.3 by bylo vhodnější dát spíše do příloh a v textu zprávy vysvětlit a odkazovat. Oceňuji objektový návrh s diagramy tříd v kap. 4.5, avšak opět by bylo vhodnější popsat spíše architekturu a rozhraní a rozsáhlé diagramy doplňkových částí (DAO a Config) umístit do příloh.
- 5. Formální úprava technické zprávy** 90 b. (A)
Zpráva má velmi dobrou formální úpravu bez zjevných nedostatků. Drobnou výtku mám pouze ke skloňování anglických slov (např. "crawlera" na str. 14).
- 6. Práce s literaturou** 85 b. (B)
Seznam literatury obsahuje 26 zdrojů, z nichž se asi 4 odborné publikace zabývají problematikou získávání a extrakce dat z webu a zbytek pojednává o použitých technologiích. Zdroje jsou v práci rozumě použity a dobře citovány.
- 7. Realizační výstup** 85 b. (B)
Realizačním výstupem je Java EE aplikace s webovým rozhraním pro zadávání a provádění úloh stahování webového obsahu a extrakce z něj metodou HLRT. Aplikace je dobře navržena a velmi pečlivě implementována (vhodné technologie, architektura, bezpečnost, persistence, atd.). Jedinou výtku lze mít k návrhu komponent pro stahování (crawler) a extrakci, kde chybí obecná a robustní rozhraní těchto komponent, takže bude problematické jejich pozdější rozšíření (implementace jiných stahovačů a extraktorů) a také škálování (distribuované stahování/extrakce).
- 8. Využitelnost výsledků**
Výsledná aplikace je dobře použitelná, avšak širší uplatnění v praxi by vyžadovalo drobné rozšíření, např. pro dynamický HTML obsah či obsah ze stránek s omezeným přístupem (autentizace).
- 9. Otázky k obhajobě**
 - Popište možný způsob rozšíření aplikace o další komponenty pro stahování a extrakci.
- 10. Souhrnné hodnocení** 85 b. velmi dobře (B)
Výsledkem práce je velmi dobrá technická zpráva a s drobnou výtku vhodně navržené a prakticky použitelné programové řešení. Celkově považuji řešení za mírně nadstandardní a navrhuji hodnotit práci stupněm **velmi dobře (B)**.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 6. června 2019

Rychlý Marek, RNDr., Ph.D.
oponent