

Posudek oponenta diplomové práce

Student: Kocman Tomáš, Bc.
Téma: Automatizovaná analýza a archivace dat z webu (id 21459)
Oponent: Veselý Vladimír, Ing., Ph.D., UIFS FIT VUT

- 1. Náročnost zadání** **průměrně obtížné zadání**
Zadání navazuje na probíhající vědecký výzkum v rámci projektu TARZAN (dr. Matoušek). Obtížnost není definována zadáním jako takovým (automatizované procházení webu a následná rekonstrukce/archivace dat), ale hloubkou jeho zpracování (mírou realizovaných detailů). Za tímto účelem student pojmul poměrně komplexní problematiku rekonstrukce webového obsahu se všemi jeho úskalími (SPA stránky, relativní/absolutní odkazy, exekovaný JavaScript, prohledávání do šířky/hloubky).
- 2. Splnění požadavků zadání** **zadání splněno**
Všechny body zadání splněny a nad jejich rámec byla kromě plnohodnotného řešení realizována i odlehčená monolitická verze nástroje pro potřeby Policie ČR.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
Práce je psána v husté LaTeXové šabloně a má 61 stránek (se všemi "pomocnými provozy" dohromady pak 64 stránek). Ve výsledku je tedy v obvyklém rozmezí.
- 4. Prezentací úroveň předložené práce** **95 b. (A)**
Práce má logickou strukturu, která odráží jednotlivé body zadání. Práce se hezky čte a (i v problematice neznalého) čtenáře dobře provádí tématem.
- 5. Formální úprava technické zprávy** **90 b. (A)**
Práce je psána v češtině bez zásadnějších prohřešků vůči spisovnosti. Co se týče typografie, anotování obrázků a diagramů, tak se zdá být bez problémů.
- 6. Práce s literaturou** **85 b. (B)**
Student v práci cituje z relevantních zdrojů. Mnohdy se jedná i o konferenční a žurnálové příspěvky, jejichž pochopení studentem je v práci patrné. Kromě standardních citací obsahuje práce i celou řadu souvisejících poznámek pod čarou. Nicméně ve výsledné bibliografii se vyskytují nekonzistence ve formátu (např. [6] vs. [10] či [12] vs. [15]) pramenů pravděpodobně vlivem špatných oddělovačů ve jménech autorů.
- 7. Realizační výstup** **96 b. (A)**
Realizační výstup tvoří komentované zdrojové kódy v Pythonu k několika modulům (lemmit, scrapit, winit) plus recepty na sestavení a orchestraci Docker kontejnerů. Na tomto místě bych rád vyzdvihнул obě realizované implementace (jak plnohodnotnou linuxovou, tak monolitickou pro Windows), které uživatelům umožňují anotovaný sběr selektivních dat z webových stránek.
- 8. Využitelnost výsledků**
Výstupy práce jsou zajímavé pro koncové uživatele projektu TARZAN. Odlehčená verze nástroje byla pak s úspěchem prezentována na konferencích LEA-der 2019 a ISS World Europe 2019 různým složkám policie z celého světa.
- 9. Otázky k obhajobě**
 - Diskutujte možnosti implementovaných nástrojů v kontextu 1-2 let s ohledem na rozvoj SPA aplikací.
- 10. Souhrnné hodnocení** **91 b. výborně (A)**
Diplomovou práci hodnotím jako výbornou (tedy stupněm A). Práce má kvalitní a prakticky použitelné realizační výstupy, které byly na více místech prezentovány/publikovány s pozitivním ohlasem. Z textové i implementační části je patrné, že se student tématu zhostil inženýrským způsobem.

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

V Brně dne: 7. června 2019

Veselý Vladimír, Ing., Ph.D.
oponent