

Posudek oponenta bakalářské práce

Student: Valaštin Samuel
Téma: Získávání znalostí z webových logů (id 24994)
Oponent: Rychlý Marek, RNDr., Ph.D., UIFS FIT VUT

- 1. Náročnost zadání** **průměrně obtížné zadání**
Bakalářská práce pojednává o předzpracování a dolování dat z logů webových serverů. Jedná se o průměrně obtížné zadání, které využívá dobře známé a zavedené postupy a technologie.
- 2. Splnění požadavků zadání** **zadání splněno s drobnými výhradami**
Zadání je splněno s jednou drobnou výhradou k bodu 2, kdy student vhodné datové sady sice vyhledal a použil (jsou přítomny v souborech odevzdané práce a používány vytvořenou aplikací), avšak nepopsal je v textu technické zprávy a nezdůvodnil tak ani jejich vhodnost pro tuto práci.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
Technická zpráva bakalářské práce obsahuje od úvodu po závěr 47 vysázených stran a je tedy svým rozsahem v obvyklém rozmezí. Rozsah jednotlivých kapitol je přiměřený a svým obsahem jsou kapitoly většinou informačně bohaté a pro práci nezbytné. Poněkud stručný je popis návrhu aplikace v kap. 5, kde postrádám popis architektury (komponenty, rozhraní, datové toky).
- 4. Prezentací úroveň předložené práce** **75 b. (C)**
Struktura technické zprávy je logická a odpovídá průběhu procesu získávání znalostí od předzpracování, vlastního dolování dat, až po vizualizaci výsledků v aplikaci. Experimenty by bylo vhodné uvést do samostatné a lépe strukturované kapitoly s výstižnějším názvem, než má současná podkap. 5.6 "Scenáře pro získávání znalostí z webových logů". Oceňuji dobré obrázky, které vhodně ilustrují text technické zprávy a demonstrují v něm popisované postupy. Škoda, že zmiňované obrázky nejsou většinou v doprovodném textu odkazovány.
- 5. Formální úprava technické zprávy** **85 b. (B)**
Z hlediska formální úpravy je technická zpráva na velmi dobré úrovni bez vážnějších nedostatků. Zpráva je psána ve slovenském jazyce.
- 6. Práce s literaturou** **75 b. (C)**
Seznam použité literatury obsahuje 28 zdrojů, z nichž podstatná většina jsou odborné publikace (knihy a články). Zdroje vhodně pokrývají téma bakalářské práce a jsou správně v textu citovány. V textu lze snadno rozlišit převzaté části od vlastních úvah autora a také většina obrázků je původních. Drobnou výtku mám k obr. 2.1 "Diagram modelu CRISP-DM" na str. 4, který je pravděpodobně autorovým překladem vizuálně totožného obrázku tak, jak je dostupný v mnoha zdrojích, a citace u něj chybí.
- 7. Realizační výstup** **75 b. (C)**
Realizačním výstupem je webová aplikace pro nahrání, předzpracování, dolování a vizualizaci dat z logů webových serverů. Aplikace je funkční a provádí požadované úlohy dolování dat za pomoci Python knihovny Mlxtend. Poněkud překvapivé je použití úložiště souborů Google Drive pro data aplikace, namísto běžnějšího přístupu, který by využil databázi (lokální nebo jako službu cloud computing). Zdrojový kód aplikace je vhodně členěn a dostatečně komentován. Vytknout lze slabou architekturu řešení, zejména úzkou provázanost dolování s vizualizací (soubor mine.py), která není v souladu s doporučeným principem oddělení vrstev pohledu a modelu/řadiče.
- 8. Využitelnost výsledků**
Samostatná aplikace je použitelná pro analýzu logů webových serverů. Pro praktičtější použití by bylo vhodnější aplikaci lépe navrhnout (architektura) a integrovat s dalšími nástroji běžně používanými pro sbírání logů a vizualizaci (Loki, Grafana).
- 9. Otázky k obhajobě**
 - Vysvětlíte důvod použití úložiště Google Drive v aplikaci, zejména ve srovnání s možným použitím databáze.
 - Na obrázku objasněte architekturu aplikace (komponenty, rozhraní, datové toky) a způsob, jak by se dala integrovat s jinými nástroji.
- 10. Souhrnné hodnocení** **75 b. dobře (C)**
Výsledkem práce je dobrá technická zpráva a funkční aplikace. Přestože existuje drobná výhrada ke splnění zadání v případě chybějící analýzy dostupných datových sad v technické zprávě, student jejich vhodnost prokázal v aplikaci. Navrhují hodnotit práci stupně **dobře (C)**.

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

V Brně dne: 1. června 2022

Rychlý Marek, RNDr., Ph.D.
oponent