

Posudek oponenta bakalářské práce

Student: Zrnčík Henrich
Téma: Identifikace uživatelů kryptoměn (id 21558)
Oponent: Matoušek Petr, Ing., Ph.D., M.A., UIFS FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Řešení BP vyžadovalo dopodrobna nastudovat problematiku kryptoměn a navrhnout a implementovat nástroj pro automatizovaný sběr metadat z webových stránek obsahující informace o kryptoměn.
- 2. Splnění požadavků zadání** **zadání splněno pouze částečně**
Student navrhl nástroje pro sběr dat o kryptoměnách z webových stránek, nástroj ale neumožňuje automatizované a periodické sbírání dat včetně zpracování, což byl jeden z cílů projektu. Také chybí otestování řešení a validace výsledků.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
Rozsah zprávy je v obvyklém rozmezí.
- 4. Prezentací úroveň předložené práce** **55 b. (E)**
Většinu technické dokumentace tvoří popis kryptoměn a dostupných frameworků pro stahování dat z webu (str. 1-30), vlastní implementaci je vyhrazeno pouze 10-15 stránek. V kapitole 4 (design) student popisuje zdroje dat, což je vhodnější dát do předchozí kapitoly. Kapitoly 4 (Design) a 5 (Implementace) by bylo vhodné spojit. V textu postrádám popis vlastní implementace a postupu, co jednotlivé části (spidery, plánovač, analyzátor textu) dělají, jaké mají vstupy a výstupy, zda se vyžaduje autentizaci při přístupu na webové stránky, jak to nástroj řeší, apod. U formátu použitých regulárních výrazů v příloze C bych uvítal vysvětlení syntaxe, tj. které znaky tvoří terminály, co jsou meta znaky popisu regulárního výrazu a co znamenají. V kapitole o testování by měl autor uvést, co testuje a proč, pak by měl představit postup testování a výsledky. Toto v práci zcela chybí.
- 5. Formální úprava technické zprávy** **50 b. (E)**
Práce je psána v angličtině. Myslím, že je užitečné, aby si studenti vyzkoušeli psát text v cizím jazyku, na druhou stranu to přináší riziko jazykových chyb. První část BP je psána celkem srozumitelně bez větších chyb. V druhé části práce jsou jazykové chyby častější, text méně srozumitelný (např. 3.1.2, 5.2) občas vypadne sloveso, chybí velké písmeno na začátku věty a tečka na konci věty. Doporučuji konzistenci v použití velkých písmen. V práci se objevují kombinace Smart Contract/Smart contract i smart contract, Proof of Work/proof of work i Proof of work, HTTP i http a další. Při odkazech na jiné části práce používejte výraz Section, nikoliv název a číslo kapitoly, tj. Section 2.3.2, nikoliv Bitcoin 2.3.2. Na str. 28 a 29 se v textu objevují značky Latexu, což by nemělo být.
- 6. Práce s literaturou** **55 b. (E)**
Výběr literatury je rozsáhlý. Seznam obsahuje zejména internetové zdroje, což je u daného tématu obvyklé. V seznamu literatury se kombinují anglické a české názvy (online navštíveno, online visited), některé položky mají špatný formát jmen autorů ([9],[12],[21],[25]). U referencí z elektronických časopisů doporučuji uvést název časopisu, nikoliv jen URL (např. u [20], [31], [39]).
- 7. Realizační výstup** **65 b. (D)**
Realizační výstup tvoří skripty, které přistupují k vybraným zdrojům (diskuzní skupiny, portály, youtube) a vyhledávají na nich řetězce týkající se kryptoměn a identifikaci autorů. Pro každý zdroj je potřeba ručně vytvořit tzv. spider, který by pak měl prohledávat daný zdroj automatizovaně. Data jsou po načtení analyzována a vybrané položky uloženy do databáze SQL. Odevzdaný kód na DVD je nečitelný, student ho na vyžádání oponenta dodal a ukázal použití.
- 8. Využitelnost výsledků**
Výsledky práce demonstrují způsob identifikace uživatelů kryptoměn ze zadaných zdrojů. Vzhledem k nárokům při zpracování knihovnou Splash se paměť rychle vyčerpá a aplikace spadne, následně se musí restartovat (viz část 5.1.1), což omezuje použití v praxi. Nástroj neumí opakovaně navštěvovat stránky, což měla být jedna z nejdůležitějších funkcionalit.
- 9. Otázky k obhajobě**
 - Jak jste testoval váš nástroj a s jakými výsledky? Jak jste měřil výkonnost systému?
 - Řešíte zpracování citlivých informací ze stránek, které mohou obsahovat osobní údaje?
 - Rozšířený abstrakt práce je psán v češtině. Kdo je autorem toho textu?
- 10. Souhrnné hodnocení** **60 b. uspokojivě (D)**
I když práce je rozsáhlá a obsahuje zajímavé části, ve svém výsledku je nedotažená. Chybí implementace

automatizované a periodické stahování dat a otestování výsledku. Celkově hodnotím práci stupněm D.

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

V Brně dne: 22. května 2019

Matoušek Petr, Ing., Ph.D., M.A.
oponent