

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Web Media Content Aggregator
Jméno autora:	Bc. Petr Cipra
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Vedoucí práce:	Mgr. Miroslav Blaško, Ph.D.
Pracoviště vedoucího práce:	KBSS

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
---------------	-------------------------

Zadání považuji za průměrně náročné. Složitost implementace vychází zejména z použití méně rozšířených technologií Sémantického webu. Byla použita grafová databáze Ontotext GraphDB a framework JOPA pro mapování ontologií na objekty v jazyce Java.

Splnění zadání	splněno
-----------------------	----------------

Zadání považuji za splněno. Naopak analýzu považuji za rozsáhlejší a implementaci za výrazně rozsáhlejší oproti typické práci. Autor např. implementoval nad rámec optimalizace stahování dat na základě odhadu pravděpodobnosti aktualizace dat, obcházení blokování geolokace a deduplikaci dat.

Zvolený postup řešení	vynikající
------------------------------	-------------------

Student zvolil správný postup a metody řešení.

Odborná úroveň	B - velmi dobře
-----------------------	------------------------

Po odborné stránce je práce kvalitní. V práci oceňuji preciznost jakou je vypracována většina kapitol. Oceňuji hlavně detailní zpracování požadavků a analýzu řešení.

Za největší nedostatek považuji málo detailní zhodnocení splnění požadavků. Našel jsem jenom vágnejší zhodnocení v závěru. Bylo by vhodné explicitně adresovat požadavky, které splněné nebyly a ty které byly splněny částečně.

Za menší nedostatky považuji:

- Práce zahrnuje podrobné aktivní diagramy, avšak vhodné by bylo zahrnout jeden agregující diagram případů užití, ideálně i s označením částečného/plného splnění těchto případů v sekci implementace.
- Sekce 4.4.2 kde autor hledá vhodný software pro extrakci webu jsou jenom Java-based aplikace, co není odůvodněno.
- Považuji za matoucí, že sekce Graphical User Interface (str. 83) obsahuje podsekce „Application Extensibility“ nebo „Extraction Process“.
- U sekce 7.1 Uživatelské testování bych ocenil informaci uživatelích, kteří aplikaci testovali

Formální a jazyková úroveň, rozsah práce	A - výborně
---	--------------------

Rozsah práce je 115 stran bez příloh a obsahuje přiměřený počet zvolených obrázků, včetně přibližně 10 UML diagramů. Text je srozumitelný, čtivý a obsahuje jen několik drobných překlepů, které nenarušují plynulost čtení.

Výběr zdrojů, korektnost citací

C - dobře

Bibliografie obsahuje 42 záznamu, z toho jenom 2 recenzované články, které jsou korektně citovány. Zbytek jsou online zdroje, kde není uveden datum citace a u pár z nich nejsou uvedeny autoři nebo datum publikace i když jsou tyto informace dohledatelné (např. autoři i datum chybí u 2, 3, 5, 24). Navíc celkově považují referencování online zdrojů za nesystematické. Autor vůbec nevyužívá footnote, který by byl vhodnější pro méně podstatné zdroje (např. odkaz na konkrétní program). V jiných částech textu je url přímo v textu (např. strana 98 „This solution stems from the compose-wireguard-squid ...“). Reference v sekci 6.4 „... NFKD - Normalization Form KD (Compatibility Decomposition).?“ se zdá být nefunkční.

Další komentáře a hodnocení

Implementační část je pečlivě zpracována a rozsáhlejší, obsahuje kolem 35 tisíc řádků zdrojového kódu. Zdrojový kód je srozumitelný a dobře zdokumentován. Aplikace je dockerizovaná.

U implementace vyhledávání bych doporučil použití Full-text search (FTS) indexu podporovaného přímo GraphDB, které by výrazně zrychlilo vyhledávání a umožnilo např. i fuzzy vyhledávání. (Realizuje se pomocí konfiguraci databáze a drobnou úpravou SPARQL dotazů)

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Se studentem se mi pracovalo velmi dobře. Oceňuji jeho aktivní přístup a efektivní komunikaci. Celkově považuji jeho práci za velmi kvalitní; precizně implementoval jak teoretickou, tak praktickou část, přičemž obě části jsou rozsáhlejší. Rád bych práci ohodnotil nejlepším stupněm, ale bohužel se student v závěrečné fázi diplomové práce zaměřil především na implementaci, což vedlo k některým závažným, i když snadno odstranitelným, chybám – špatně zvoleným a místy chybně citovaným zdrojům a nejasnému vyhodnocení splnění požadavků aplikace.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm B - velmi dobře.

Otázky k obhajobě:

1. Objasněte jak funguje algoritmus pro extrakci obsahu popsany v sekci Extraction process na stránce 89.
2. Detailněji popište, které z požadavků ze sekce 5.1 byly splněny, případně částečně splněny.

Datum: 13.6.2024

Podpis: