

Posudek oponenta bakalářské práce

Student: Balaga Róbert**Téma:** Navrhování klíčových slov v Centrálním portálu knihoven (id 15372)**Oponent:** Otrusina Lubomír, Ing., UPGM FIT VUT**1. Náročnost zadání****průměrně obtížné zadání****2. Splnění požadavků zadání****zadání splněno s drobnými výhradami**

Zadání práce bylo splněno, ale u některých bodů pouze v minimální možné míře nebo pouze částečně. Student se v práci nevěnuje dostatečně předzpracování českého textu (bod 1). Zcela pak chybí informace o začlenění výsledků do projektu CPK (bod 3). Závažnost těchto pochybení je ale spíše malá.

3. Rozsah technické zprávy**je v obvyklém rozmezí**

Rozsah technické zprávy je v obvyklém rozmezí. Všechny části práce jsou informačně bohaté a pro práci nezbytné.

4. Prezentační úroveň předložené práce**55 b. (E)**

Prezentační úroveň práce je podprůměrná. Některé kapitoly jsou hůře pochopitelné pro čtenáře (např. 2.2.4 Keygraph). Grafy nejsou dostatečně popsány a student se na ně v textu neodkazuje. Popis vzorců je občas nepřehledný. V textu se objevují termíny, které student nevysvětluje, anebo je vysvětluje až v dalších kapitolách. V kapitole 2 chybí podkapitola o lingvistických metodách. Popis dosažených výsledků v kapitole 4 je nepřehledný a chybí zde jejich důkladné zhodnocení. Závěr práce je příliš krátký. Práce také neobsahuje žádné informace o projektu CPK, čtenář tak nezíská dostatečný přehled o problému.

5. Formální úprava technické zprávy**60 b. (D)**

Formální úprava technické zprávy má některé nedostatky. Obrázky jsou většinou vloženy doprostřed věty. Některé termíny nejsou přeloženy (dataset, precision, recall, part of speech). V textu jsou časté mezery před čárkami nebo tečkami, jinde zase mezery chybí. Občas chybí odkazy na použité nástroje.

6. Práce s literaturou**55 b. (E)**

Výběr studijních pramenů je dobrý. U některých zdrojů chybí informace o tom, kde byly publikovány. Student ve většině případů uvádí, odkud dané informace čerpal. V některých případech to ale není úplně zřejmé. Výjimečně pak někde chybí citace úplně (např. v kapitole 2.2.2 SingleRank, u obrázku 3.1 nebo obrázku 4.9). Ve zdrojích se také nachází některé položky, které by se hodily spíše do poznámky pod čarou (např. odkaz na github).

7. Realizační výstup**55 b. (E)**

Student implementoval systém, který dokáže navrhovat klíčová slova z českých dokumentů. Realizační výstup patří spíše k jednodušším. Na přiloženém médiu se nachází skript, který implementuje navrženou metodu. Chybí tam ale skripty, které student používá k získání výsledků jiných metod (TF-IDF a RAKE). Není proto jasné, jak tyto výsledky získal. Vlastní kód skriptu je jednoduchý a nepřehledný, neobsahuje žádné funkce ani objekty.

8. Využitelnost výsledků

Jedná se o práci kompilačního charakteru. Využitelnost dosažených výsledků v praxi je diskutabilní.

9. Otázky k obhajobě

- V kapitole 4.2.5 popisujete výsledky Vámi implementované metody. Jsou zde uvedeny hodnoty přesnosti a pokrytí pro různá nastavení parametru alfa. Jak si vysvětlujete to, že se hodnota přesnosti pro různá nastavení parametru alfa téměř nemění (narozdíl od ostatních metod)?
- Pišete také, že pro využití Vašeho systému v projektu CPK, je důležitější jeho pokrytí na úkor přesnosti. Jak jste došel k tomuto závěru?

10. Souhrnné hodnocení**55 b. dostatečně (E)**

Jedná se o průměrně obtížné zadání. Zadání bylo splněno s drobnými výhradami. Prezentační úroveň práce je podprůměrná. Formální úprava technické zprávy má také nedostatky. Práce s literaturou obsahuje mnohé prohřešky, v některých částech není jasné, odkud student čerpal. Realizační výstup práce patří k těm jednodušším. Využití dosažených výsledků v projektu CPK je diskutabilní. Z těchto důvodů navrhuji hodnocení stupněm E.

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

V Brně dne: 31. května 2018

.....
podpis