

Posudek oponenta bakalářské práce

Student: Gallovič Lubomír
Téma: Automatická extrakce klíčových slov v češtině (id 15374)
Oponent: Otrusina Lubomír, Ing., UPGM FIT VUT

- 1. Náročnost zadání** **průměrně obtížné zadání**
- 2. Splnění požadavků zadání** **zadání splněno s drobnými výhradami**

V rámci bodu 4 měly být diskutovány možnosti zvýšení výkonosti a úspěšnosti vytvořeného systému. Zvýšení úspěšnosti je věnováno jen několik málo vět v závěru práce v kapitole 12.1 Možné vylepšení. Zvýšení výkonosti diskutováno nebylo. Tento bod tedy nebyl zcela splněn. Závažnost tohoto prohrěšku je ale poměrně nízká.
- 3. Rozsah technické zprávy** **přesahuje obvyklé rozmezí**

Rozsah technické zprávy přesahuje obvyklé rozmezí. To je způsobeno zejména rozsahy kapitol 9 a 10, které se věnují analýze použitých přístupů. Tyto kapitoly, které zabírají téměř polovinu celé technické zprávy, obsahují velké množství nepříliš zajímavých grafů a měly být zkráceny.
- 4. Prezentací úroveň předložené práce** **50 b. (E)**

Prezentací úroveň technické zprávy není příliš dobrá. Některé kapitoly věnující se popisu použitých algoritmů mohly být sloučeny. Kapitoly 9 a 10 měly být zkráceny a zpřehledněny. Některé výsledky bylo lepší prezentovat v tabulkách. Návaznost kapitol je vcelku dobrá, ale práce je v některých místech hůře pochopitelná pro čtenáře. Největším problémem jsou nevysvětlené proměnné ve vzorcích nebo nedostatečně vysvětlené a nepřehledné grafy a tabulky. Práce rovněž obsahuje mnoho nepravdivých, nepřesných a nepodložených tvrzení. Některé použité pojmy nejsou v práci vysvětleny. Srozumitelnosti práce by zcela jistě pomohlo, kdyby student v práci uváděl více příkladů.
- 5. Formální úprava technické zprávy** **55 b. (E)**

Formální úprava technické zprávy také není příliš dobrá. Po jazykové stránce práce obsahuje značné množství překlepů, nepřeložených termínů a nesrozumitelných vět. Kapitoly často začínají přímo podkapitolami, bez jakéhokoliv úvodu do kapitoly. V práci chybí odkazy na některé použité technologie či datové sady.
- 6. Práce s literaturou** **65 b. (D)**

Výběr studijních pramenů je poměrně dobrý, ale mám k němu několik výhrad. Na podobné téma bylo na fakultě řešeno již několik bakalářských či diplomových prací, student však necituje žádnou z nich. Spousta referencí by se hodila spíše do poznámky pod čarou. Student uvádí, kdy daný zdroj navštívil online i u zdrojů, kde to nedává příliš smysl, např. u vědeckých článků publikovaných ve sborníku.
- 7. Realizační výstup** **55 b. (E)**

Výsledky práce jsou těžko reprodukovatelné, např. není jasné, odkud student bral seznam stop slov. Taktéž využívaný korpus fakulty informačních technologií VUT, který student použil pro vyhodnocení systému, není v práci dostatečně popsán, a není tak zcela jasné, o jaké publikace se jedná. Úspěšnost použitých algoritmů je na dané datové sadě velmi nízká, např. přesnost nalezení tříslavných termínů je nižší než 5 %. Zdrojové kódy jsou špatně komentované.
- 8. Využitelnost výsledků**

Jedná se o práci kompilačního charakteru. Dosažené výsledky pravděpodobně nejsou v praxi nijak využitelné.
- 9. Otázky k obhajobě**
 - V tabulce 9.1 uvádíte seznam nejčastějších vzorů POS tagů z rejstříků testovacích publikací. Jejich četnosti, ale v práci uvedeny nejsou. Můžete je komisi prezentovat?
 - Čím je podle Vás způsobena tak malá úspěšnost Vámi použitých algoritmů? Můžete komisi znovu popsat nejlepší Vámi dosažené výsledky? Jak jste na tom ve srovnání s konkurenčními systémy?
 - V práci uvádíte, že je na závěr potřeba identifikované termíny převést do správného tvaru. V práci to ale nijak více nerozebíráte. Jak tedy termíny převádíte do správného tvaru a jaká je úspěšnost převodu? Na jakých tvarech provádíte vyhodnocení Vašeho systému?
- 10. Souhrnné hodnocení** **55 b. dostatečně (E)**

Prezentací úroveň technické zprávy je špatná. Práce je v některých místech špatně pochopitelná pro čtenáře. Je zde rovněž několik formálních nedostatků. Kapitoly s vyhodnocením metod jsou příliš dlouhé a nezajímavé. V práci chybí srovnání s podobnými systémy. 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: 1. června 2017

.....
podpis