

Posudek oponenta bakalářské práce

Student: Panov Sergey
Téma: Extrakce informací z Wikipedie (id 18751)
Oponent: Otrusina Lubomír, Ing., UPGM FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
2. **Splnění požadavků zadání** zadání splněno s vážnými výhradami
Zadáním práce bylo navrhnout a implementovat systém pro automatickou extrakci vztahů vybraného typu a pro identifikaci nekonzistentních prvků ve Wikipedii a jiných polostrukturovaných zdrojích. Dále také realizovat nástroj pro automatické opravování textů Wikipedie. Předložený nástroj pro identifikaci nekonzistentních prvků ve Wikipedii a pro automatické opravování textů Wikipedie je však velmi triviální a zdaleka nesplňuje očekávání.
3. **Rozsah technické zprávy** je v obvyklém rozmezí
Rozsah technické zprávy je spíše na dolní hranici obvyklého rozsahu. Většina částí technické zprávy je rozsahově v pořádku. Výhrady však mám ke kapitole 5 - Testování a vyhodnocení výsledků, která je velmi strohá a o implementovaném systému toho příliš nevypráví.
4. **Prezentační úroveň předložené práce** 40 b. (F)
Prezentační úroveň práce je velmi špatná. Student není rodilý mluvčí a i přesto se pustil do psaní práce v českém jazyce. V celé práci by člověk jen stěží hledal gramaticky správně postavenou větu. Už jen pouhé použití spell checkeru by odhalilo velké množství chyb. Většina částí textu tak působí velmi jednoduše, některé pasáže jsou nepochopitelné nebo nepravdivé. Mnohem větší smysl by dávalo psát práci v anglickém jazyce. Návaznosti jednotlivých kapitol jsou dobré a kapitoly většinou obsahují to, co by měly obsahovat.
5. **Formální úprava technické zprávy** 30 b. (F)
Jak bylo zmíněno výše, po jazykové stránce je ve velmi špatném stavu. Po typografické stránce nemám k práci vážnější výhrady.
6. **Práce s literaturou** 50 b. (E)
Ačkoliv seznam literatury obsahuje relevantní publikace, občas není zřejmé, odkud student čerpal. V některých podkapitolách (např. 2.5.2 Techniky založené na NLP) citace zcela chybí. Zdroje [1] a [2] jsou totožné.
7. **Realizační výstup** 40 b. (F)
Realizační výstup práce není příliš zdařilý. Ačkoliv student v nejvyšší možné míře splnil zadání a vytvořil systém pro extrakci informací, systém je velmi úzce zaměřen a není schopen extrahovat téměř žádné použitelné informace. Výhrady mám zejména k identifikaci nekonzistentních prvků ve Wikipedii, což student realizoval tak, že zpracoval celou Wikipedii a našel v ní odkazy na zatím neexistující stránky. Nástroj pro automatické opravování textů Wikipedie pak umí pouze doplňovat chybějící informace o seriálech, přičemž student nevyužívá dat, která vyextrahoval, ale dat z DBpedie.
8. **Využitelnost výsledků**
Jde o práci kompilačního charakteru. V současném stavu práce není použitelná.
9. **Otázky k obhajobě**
-
10. **Souhrnné hodnocení** 40 b. nevyhovující (F)
Student se ve své práci zaměřuje na velmi úzce specifickou extrakci informací. Dá se říci, že zadání splnil pouze v nejmenší možné míře. Vyhodnocení systému toho příliš mnoho nevypráví. Práce obsahuje obrovské množství gramatických chyb, nesrozumitelných pasáží a je psána velmi jednoduchou češtinou. Z těchto důvodů navrhuji hodnocení stupněm F a práci nedoporučuji k obhajobě.

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

V Brně dne: 2. června 2016

.....
podpis

