

Posudek oponenta bakalářské práce

Student: Rusiňák Petr

Téma: Určování typů entit na základě extrakce informací z Wikipedie (id 20995)

Oponent: Otrusina Lubomír, Ing., UPGM FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Jednalo se o obtížnější zadání. Obtížnost spočívala zejména v práci s velkými daty a manuální klasifikaci dat pro vyhodnocení systému.
- 2. Splnění požadavků zadání** **zadání splněno s vážnými výhradami**
Výhrady mám zejména k bodu 2. Student v práci nepopisuje žádné metody převodu nestrukturovaných textových informací do strukturované podoby. Popis souvisejících projektů je také velmi strohý, v zásadě se zde student omezuje jen na DBpedii. Pokud by se student důkladněji seznámil s používanými technikami, mohl by být realizační výstup kvalitnější.
- 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** **75 b. (C)**
Prezentační úroveň technické zprávy je dobrá, práce je čtivá a je dobře pochopitelná pro čtenáře. Problém je zejména v tom, že zde chybí popis metod převodu nestrukturovaných textových informací do strukturované podoby. Kapitola 4 "Návrh systému" je velmi strohá. Mnohé informace, které by měly být v této kapitole, jsou rozptýleny do teoretické části práce a do kapitoly popisující implementaci. Rovněž kapitola 8 s popisem výsledků by měla být obsáhlejší. V práci se občas vyskytují sporná tvrzení.
- 5. Formální úprava technické zprávy** **80 b. (B)**
Formální úprava technické zprávy je dobrá. Po jazykové stránce práce obsahuje několik málo pravopisných chyb (ty data, ty pravidla), překlepů či kostrbatých konstrukcí (článek patří do entity). Po typografické stránce bych vytkl nečíslované úryvky datových souborů či některé tabulky.
- 6. Práce s literaturou** **65 b. (D)**
V seznamu studijních pramenů chybí publikace popisující převod nestrukturovaných textových informací do strukturované podoby, případně publikace popisující další existující systémy. Některé zdroje by se hodily spíše do poznámky pod čarou. V jenom případě se student odkazuje na interní wiki stránku výzkumné skupiny, která není veřejně přístupná.
- 7. Realizační výstup** **85 b. (B)**
Jelikož se student nedostatečně seznámil s existujícími metodami, vychází návrh systému zejména ze studentovi vlastní zkušenosti. Studentem navržený systém je ale zdařilý a provádí automatickou extrakci typů entit z Wikipedie velmi dobře. Při použití pokročilejších technik by ale pravděpodobně bylo dosaženo ještě lepších výsledků.
- 8. Využitelnost výsledků**
Studentem navržený a implementovaný systém je možné úspěšně použít pro automatickou extrakci typů entit z Wikipedie.
- 9. Otázky k obhajobě**
 - V kapitole 8 "Výsledky" bych očekával podrobnější srovnání Vašeho systému s DBpedíí. Můžete komisi ukázat výsledky tohoto srovnání na několika typech entit. Jaké články správně identifikoval Váš systém oproti DBpedii? Jaké články správně identifikovala DBpedie oproti Vašemu systému? Jaké jsou výhody Vašeho systému oproti DBpedii?
- 10. Souhrnné hodnocení** **75 b. dobře (C)**
Jednalo se o obtížnější zadání, které student splnil s výhradami. Hlavní pochybení je zejména v nedostatečném nastudování existujících přístupů a systémů. Na druhou stranu je prezentační úroveň technické zprávy na velmi dobré úrovni a studentem vytvořený systém dává dobré výsledky. Z těchto důvodů navrhuji hodnocení stupněm C.

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

V Brně dne: 31. května 2018

.....
podpis