

Posudek oponenta bakalářské práce

Student: Švub Daniel
Téma: Určování typů a atributů entit napříč jazyky (id 21926)
Oponent: Otrusina Lubomír, Ing., UPGM FIT VUT

- 1. Náročnost zadání** **průměrně obtížné zadání**
- 2. Splnění požadavků zadání** **zadání splněno pouze částečně**

Zadání bylo splněno pouze částečně. Bod č. 1 byl dle mého názoru splněn nedostatečně. Student sice zmiňuje několik metod použitelných pro určování typů entit a extrakci atributů, ale neuvádí žádné dostupné nástroje. Bod č. 2 splněn nebyl, práce neobsahuje přehled jazykově závislých technik využívaných při extrakci a nástrojů pro různé úrovně analýzy jazyka v češtině a slovenštině. Poslední výtka směřuje k bodu č. 4, v práci totiž není žádné vyhodnocení výsledků systému na reprezentativním vzorku dat. Díky pochybení v bodech č. 1 a 2 je implementovaný systém velmi jednoduchý. Kvůli chybějícímu vyhodnocení nevíme, zda systém funguje.
- 3. Rozsah technické zprávy** **splňuje pouze minimální požadavky**

Technická zpráva splňuje pouze minimální požadavky na rozsah. Některé pasáže kapitoly 3 by ale bylo vhodné vynechat nebo je přesunout do příloh, protože slouží spíše jako manuál. Na druhé straně by bylo vhodné rozšířit teoretickou část práce a popis vyhodnocení systému, které je velmi nepřehledné a nedostatečné.
- 4. Prezentací úroveň předložené práce** **50 b. (E)**

Prezentací úroveň technické zprávy je velmi špatná. Práce obsahuje mnoho nezajímavých implementačních detailů (často hraničících se slovním přepisem zdrojových kódů). Naopak některé důležité implementační detaily v práci uvedeny nejsou. Text je špatně strukturován, obsahuje dlouhé souvislé pasáže a práce obecně není příliš čtivá. Některé pojmy zůstávají v práci nevysvětleny a mnohdy není jasné, proč jsou v práci některé informace uvedeny. Práce také obsahuje mnohá nepravdivá nebo nepodložená tvrzení.
- 5. Formální úprava technické zprávy** **65 b. (D)**

Formální úprava technické zprávy je podprůměrná. Jazyk práce občas připomíná spíše vypravování, než technickou zprávu. Práce obsahuje několik málo překlepů a stylistických chyb. Ukázky formátů výstupních souborů a některé obrázky jsou nepřehledné.
- 6. Práce s literaturou** **55 b. (E)**

Výběr studijních pramenů není příliš dobrý. Většinu zdrojů tvoří weby zmiňovaných technologií, případně bakalářské nebo diplomové práce na dané téma. Chybí zde seriózní zdroje pokrývající problematiku extrakce informací z Wikipedie. Na jednotlivé studijní prameny je v práci většinou vhodně odkazováno, výjimečně ale není jasné, odkud student čerpal (např. odkud pochází definice báze znalostí na straně 9). V seznamu použité literatury nalezneme i interní wiki stránky výzkumné skupiny KNOT@FIT. Student v práci neuvádí odkazy na některé zmiňované technologie či nástroje.
- 7. Realizační výstup** **45 b. (F)**

Implementovaný systém splňuje zadání jen v minimální možné míře. To je do značné míry způsobeno tím, že student nezískal dostatečný přehled o metodách zpracování přirozeného jazyka, které se v oblasti extrakce informací z Wikipedie používají. Extrakci definičních slov implementoval pomocí ručně vytvořených pravidel. Obdobně identifikace typů entit je založena na ručně vytvořených seznamech klíčových slov. Úspěšnost implementovaného systému není v práci nijak zjišťována, je zde pouze srovnání s jiným obdobným systémem vytvořeným ve výzkumné skupině KNOT@FIT. Srovnání s dalšími systémy, jako je např. DBpedia, zcela chybí. Zdrojové kódy realizačního výstupu jsou nedostatečně okomentovány.
- 8. Využitelnost výsledků**

Jedná se o práci kompilačního charakteru. Některé výsledky práce mohou být pravděpodobně využity v rámci výzkumné skupiny KNOT@FIT.
- 9. Otázky k obhajobě**
 - Na straně 14 píšete: "Na české Wikipedii program dosahuje úspěšnosti 98 %, na slovenské verzi 96 %." Můžete komisi vysvětlit, jak jste k těmto číslům dospěl? V práci to není dostatečně vysvětleno.
 - Na straně 29 píšete: "Systém převádí wikitext na prostý text s vysokou úspěšností." Čím můžete toto tvrzení podložit? V práci to není uvedeno.
- 10. Souhrnné hodnocení** **49 b. nevyhovující (F)**

Zadání práce bylo splněno pouze částečně. Prezentací úroveň technické zprávy je velmi špatná, mnohé

Vysoké učení technické v Brně

Fakulta informačních technologií

informace v ní chybí. V seznamu použitých zdrojů chybí seriózní zdroje pokrývající problematiku extrakce informací z Wikipedie. Implementovaný systém je velmi jednoduchý a není v práci nijak rozumně vyhodnocen. Z těchto důvodů navrhuji hodnocení stupněm F (49b).

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

V Brně dne: 29. května 2019

.....
podpis