

## Posudek oponenta diplomové práce

**Student:** Pastorek Peter, Bc.

**Téma:** Příprava vyhodnocovací sady pro složité problémy rozpoznávání a zjednodučování pojmenovaných entit pomocí crowdsourcingu (id 21986)

**Oponent:** Otrusina Lubomír, Ing., UPGM FIT VUT

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

Zadání práce bylo splněno pouze částečně. Výhrady mám ke splnění bodu č. 1, kde se student nedostatečně seznámil s principy využití crowdsourcingu a gamifikace pro anotaci lingvistických dat a s dostupnými nástroji pro rozpoznávání a zjednodučování pojmenovaných entit, kvůli čemuž navržený systém nesplňuje očekávání. Dále mám výhrady k bodu č. 3, v práci zcela chybí jakékoliv vyhodnocení výsledků sběru dat a kvality získaných anotací. Z textu práce vyplývá, že žádné anotace od uživatelů získány nebyly.
- 3. Rozsah technické zprávy** nesplňuje minimální požadavky

Rozsah technické zprávy nesplňuje minimální požadavky. U některých částí práce navíc není jasné, proč tam jsou, a bylo by možné je vypustit.
- 4. Prezentační úroveň předložené práce** 50 b. (E)

Prezentační úroveň technické zprávy je velmi špatná. Práce nemá jasně stanovené cíle. Technická zpráva je v některých místech hůře srozumitelná (např. kapitola 3.2) a obsahuje mnohá nepravdivá a nepodložená tvrzení. Informace v některých kapitolách jsou hůře srozumitelné a nejsou dobře začleněny do kontextu. Mnohdy není jasné, proč je tam student vůbec uvádí. Obr. 4.1 a tabulky 6.1 a 6.2 nejsou v práci dostatečně vysvětleny.
- 5. Formální úprava technické zprávy** 65 b. (D)

Formální úprava technické zprávy je podprůměrná. Technická zpráva obsahuje několik překlepů a pravopisných chyb. Vyskytují se zde také nevhodně přeložené výrazy z angličtiny (např. kapitola 3.2), díky čemuž se zpráva stává místy nesrozumitelnou.
- 6. Práce s literaturou** 40 b. (F)

Seznam studijních pramenů obsahuje 31 položek a vypadá na první pohled velmi seriózně. Na většinu z nich se ale student v textu nijak neodkazuje, takže není vůbec jasné, jak (a jestli vůbec) je použil. V mnoha částech technické zprávy chybí citace, ačkoliv je jasné, že student musel odněkud čerpat (např. strana 2 nebo 12). V seznamu se nachází i několik odkazů na weby použitých technologií, které by se hodily spíše do poznámky pod čarou. V technické zprávě naopak chybí odkazy nebo reference na mnohé zmiňované technologie.
- 7. Realizační výstup** 30 b. (F)

Realizační výstup práce nenaplnuje očekávání. Student měl za úkol implementovat systém pro získávání dat s vyznačenými entitami a odkazy na Wikipedii pomocí crowdsourcingu a gamifikace. Studentem implementovaný systém je velmi jednoduchý a pravděpodobně není pro uživatele příliš zajímavý, a tak se pomocí něj nepovedlo sesbírat žádná data. Namísto toho student při tvorbě datové sady vychází z již existující datové sady Wikilinks. Z ní filtruje pouze určitá data na základě jím zvolených kritérií, tato kritéria ale dostatečně nezduodňuje. Student dále s velkými obtížemi zprovoznil dva systémy pro získávání dat s vyznačenými entitami a odkazy na Wikipedii (SEC a LingPipe), zpracoval jimi část vytvořené datové sady a výsledky vyhodnotil. Vyhodnocení ovšem nic neříká o kvalitě datové sady. Navíc je zkresleno tím, že v některých případech student trénoval a testoval systém na těch samých datech. Zdrojové kódy jsou velmi nepřehledné a téměř bez komentářů. Pokud tam nějaké komentáře jsou, tak většinou nedávají smysl. Dokumentace ke skriptům je ve formě nepřehledného README souboru.
- 8. Využitelnost výsledků**

Jedná se o práci kompilačního charakteru. Výsledky práce nejsou nijak využitelné v praxi.
- 9. Otázky k obhajobě**

-
- 10. Souhrnné hodnocení** 40 b. nevyhovující (F)

Jedná se o průměrně obtížné zadání, které student splnil jen z části. Student sice implementoval jednoduchý systém pro získávání dat dle zadání, nepodařilo se mu ale sesbírat žádná data, která by mohl vyhodnotit. Prezentační úroveň technické zprávy je velmi špatná. Technická zpráva rovněž obsahuje mnohá pochybení při práci s literaturou. Z těchto důvodů navrhuji hodnocení stupněm F (40b).

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

V Brně dne: 4. června 2019

Otrusina Lubomír, Ing.  
oponent