

Posudek oponenta bakalářské práce

Student: Matyáš Šimon
Téma: Metavyhledávání recenzí na českém webu (id 21574)
Oponent: Doležal Jan, Ing., UPGM FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání

2. **Splnění požadavků zadání** zadání splněno

3. **Rozsah technické zprávy** splňuje pouze minimální požadavky

Rozsah technické zprávy splňuje pouze minimální požadavky. Všechny části technické zprávy jsou skutečně nezbytné. Schází mi popis pojmů "metavyhledávání recenzí", "aspekt" a "sentiment", jenž jsou použity v technické zprávě. Dále mi schází formální popis formátu vstupních dat a konfiguračního souboru.

4. **Prezentační úroveň předložené práce** 55 b. (E)

Některé informace z kapitoly "Návrh řešení" patří spíše do kapitoly "Realizace", protože student popisuje chování výsledného nástroje. Také v kapitole "Návrh řešení" popisuje, jakým způsobem vytvořil datovou sadu pro testování. Naopak některé informace z kapitoly "Vyhodnocení" postrádám v návrhu. Student např. v kapitole "Vyhodnocení" uvádí, že první verze jeho nástroje využívaly jiné vstupní soubory než bylo uvedeno v kapitole "Návrh řešení".

Student popisuje model OpenAI GPT jako "před trénovaný jazykový model založený na transformátorech", ale nikterak nevysvětluje, že "trasformátorem" myslí právě model Transformer, jenž popisuje na stejné stránce technické zprávy.

Často schází slovní uvedení před odkaz na obrázek, kapitolu či zdroj. Zvláště na straně 6 pokládám za matoucí uvést odkaz na zdroj [38] slovy "začínající písmenem P a číslem[38]", na straně 12 odkaz na obrázek 3.1 slovy "jednotlivé části celé práce 3.1" a na straně 14 odkaz na obrázek 3.3 slovy "nadpisy a obsah článku3.3".

V závěru práce je uvedeno, že "Sentiment je určován s průměrnou přesností 72,6%." Jedná se ale o chybné zobecnění, neboť se jedná o průměrnou přesnost určení sentimentu u pěti testovaných názvů produktu ze strany 33.

5. **Formální úprava technické zprávy** 59 b. (E)

Ohledně typografické stránky technické zprávy bych vytkl uvádění desetinné tečky místo desetinné čárky a použití spojovníku místo pomlčky. Kresby vytvořené studentem jsou bitmapové a text v nich je psán malým bezpatkovým písmem (písmo se odlišuje od zbytku práce). Přitom často by bylo jednoduché je vytvořit ve vektorovém formátu.

Co se týče jazykové stránky, technická zpráva obsahuje velké množství chyb. Často se jedná o chybnou koncovku. Občas se vyskytují "věty" bez přísudku. Místy se v českém textu objevuje anglický slovosled a chybí interpunkce.

6. **Práce s literaturou** 60 b. (D)

V teoretické části technické zprávy student často uvádí více zdrojů u jedné věty. Extrémem je část "Klasifikace pomocí slovníků" na straně 8, kde student napsal pouze tři věty a u té třetí věty uvedl tři odkazy zdroj.

Formát citace závěrečných prací není konzistentní. Míchá se angličtina s češtinou. Poprvé student napíše "Bakalářská práce. Vysoké učení technické v Brně", podruhé "Bachelor's thesis. Vysoké učení technické v Brně" a potřetí "Bachelor's thesis. Brno University of Technology".

7. **Realizační výstup** 60 b. (D)

Plakát obsahuje zkopírovaný text z technické zprávy (včetně chyb) doplněn o několik obrázků, což mi nepřipadá jako šťastné řešení. Očekával bych, že student věnuje tvorbě plakátu více svého času a umu.

Studentem vytvořený nástroj funguje. Dlouho ale trvá získání odpovědi na dotaz. Hledání recenzí probíhá paralelně na uživatelem definovaném počtu vláken. Následné vyhodnocení sentimentu v nalezených recenzích je

ale zpracováváno pouze jedním vláknem. Mohlo by tedy být jednoduše urychleno souběžným zpracováváním recenzí definovaným počtem vláken.

Ve zdrojovém kódu bych uvítal větší množství komentářů a implementaci kontroly konfiguračního souboru, jenž se v odevzdané verzi vyhodnocuje jako zdrojový kód.

8. **Využitelnost výsledků**

Jedná se o práci kompilačního charakteru. Student vytvořil nástroj pro příkazový řádek, jenž slouží k vyhledávání recenzí a jejich klasifikaci. Práce s nástrojem je sice jednoduchá, ale kýženého výsledku uživatel dosáhne pouze tehdy, zadá-li správně typ produktu, který hledá. Např. při hledání produktu "smartphone" či "chytrý telefon" v časovém horizontu od 2021-03-10 do 2021-03-20 nebyla nalezena žádná recenze, ale při hledání produktu "telefon" ve stejném časovém horizontu bylo nalezeno dost recenzí na "chytrý telefon" a alespoň jedna z recenzí obsahovala slovo "smartphone". Uživatelskou přívětivost zhoršuje i dlouhé čekání na odpověď.

9. **Otázky k obhajobě**

-

10. **Souhrnné hodnocení**

60 b. uspokojivě (D)

Technická zpráva se hůře čte a obsahuje velké množství chyb. Některé informace z kapitoly "Návrh řešení" by měly být spíše v kapitole "Realizace" a kapitola "Návrh řešení" by měla být rozšířena o skutečný návrh. Pro správné pochopení schází vysvětlení zmíněných pojmů. Z těchto důvodů navrhuji hodnocení stupněm D (60 bodů).

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

V Brně dne: 2. června 2021

Doležal Jan, Ing.
oponent