

Posudek oponenta bakalářské práce

Student: Růček Peter
Téma: Analýza dějových linií na základě shrnutí obsahu knih a uživatelských recenzí (id 24752)
Oponent: Dočekal Martin, Ing., UPGM FIT VUT

- 1. Náročnost zadání** průměrně obtížné zadání
Zadání považuji za průměrně náročné.
- 2. Splnění požadavků zadání** zadání splněno
Zadání této práce bylo splněno.
- 3. Rozsah technické zprávy** je v obvyklém rozmezí
Technická zpráva je v obvyklém rozmezí. Všechny části jsou vzhledem k tématu práce relevantní.
- 4. Prezentací úroveň předložené práce** 60 b. (D)
Kapitoly na sebe dobře logicky navazují až na podkapitulu 6.1 Dátová sada, která by lépe zapadala do třetí kapitoly Data.

Jednotlivé části textu svým obsahem dobře zapadají do tématu práce. Bohužel však některé pasáže jsou popsány příliš povrchně až vágně, místy s nepravdivými informacemi. Text například tvrdí, že BERT model nebyl speciálně vytvořen pro práci s textem. Dále například autor nedostatečně popisuje vytvořený model.

Bylo by také vhodné více doplnit text o matematické vztahy.

- 5. Formální úprava technické zprávy** 80 b. (B)
Celkově formátování textových pasáží je na dobré úrovni. Text obsahuje pouze menší množství překlepů a typografických chyb.

Ilustrace jsou povětšinou rastrové obrázky v horší kvalitě a bylo by vhodné je vyměnit za vektorové.

Autor občas používá nevhodné hovorové fráze.

- 6. Práce s literaturou** 75 b. (C)
Z textu je dobře patrné odkud autor čerpal a zdroje jsou vesměs dobře zvolené. Ovšem například volba zdroje [11] je problematická, jelikož tento přehledový článek je dosti povrchní. Dále například zdroj [12] je poměrně starý (2001) a autor jej používá ke zhodnocení aktuálního stavu poznání v NLP.

Je použito celkem 26 zdrojů.

- 7. Realizační výstup** 65 b. (D)
Za realizační výstup považuji vytvořené datové sady, systém pro analýzu dějových linií a experimenty s ním.

Autor vytvořil datovou sadu z portálu IMDb vhodnou pro naučení a evaluaci modelů analyzujících klíčové dějové linie. Dále vytvořil datovou sadu z portálu Goodreads, která ale neobsahuje anotace klíčových dějových linií a tady její použití k evaluaci je omezeno pouze na lidského hodnotitele.

V rámci práce byl natrénován systém pro analýzu dějových linií založený na modelu RoBERTa. Systém je realizován jako soubor skriptů, které nemají možnost uživatelské konfigurace a používají množství předdefinovaných konstant.

Systém by si zasloužil lepší vyhodnocení výsledků. Čtenáři by mělo být poskytnuto alespoň porovnání s nějakým základním modelem či porovnání s výsledky z ostatních prací na příbuzných datových sadách.

Kvantitativní vyhodnocení je pouze poskytnuto pro datovou sadu IMDb a to jen pro extrakci dějových linií a ne pro určování podobnosti.

- 8. Využitelnost výsledků**
Podařilo se vytvořit datovou sadu (z portálu IMDb) vhodnou pro naučení a evaluaci modelů analyzujících klíčové dějové linie.

Mimoto autor nashromáždil velké množství knižních dat, které mohou najít využití v dalších projektech.

9. Otázky k obhajobě

Jaké jiné přístupy k určování podobnosti příběhů jste zvažoval?

10. Souhrnné hodnocení

70 b. dobře (C)

Autor vytvořil systém schopný analyzovat dějové linie a datovou sadu pro jeho natrénování a vyhodnocení.

Textová část práce má několik nedostatků a vyhodnocení systému by si zasloužilo dodatečné úsilí.

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

V Brně dne: 24. května 2022

Dočekal Martin, Ing.
oponent