
Posudek oponenta bakalářské práce

Jan Krajňák
Automatická detekce klíčových slov v textu

Cílem práce Jana Krajňáka bylo prozkoumat metody automatické extrakce klíčových slov, navrhnout vlastní metodu a porovnat jí s již existujícími. Autor nastudoval a s využitím knihoven implementoval 4 existující metody učení bez učitele pro extrakci klíčových slov a dále netrénoval model typu BERT pro řešení stejné úlohy.

Práce má vesměs dobrou logickou strukturu, ikdyž členění do kapitol a podkapitol není vždy úplně ideální (například nelogické umístění kapitol 3.2.4 a 5.1). V práci jsou často používané termíny, které v ní nejsou vysvětlené. Autor občas příliš zachází do detailů jednotlivých metod, kde je patrné, že danou metodu úplně nepochopil. Úroveň porozumění je ale na bakalářskou práci dostatečná. Největším nedostatkem práce je podle mě její gramatická úroveň. Práce obsahuje poměrně malé množství překlepů, obsahuje však i hrubé gramatické chyby a mnoho chyb ve stavbě vět. Text se proto poměrně špatně čte a některé části jsou s těžší pochopitelné.

Práce s literaturou je na dobré úrovni. Autor ve své práci cituje dostatečné množství relevantních zdrojů, tištěných i webových.

Autor provedl řadu experimentů a jejich výsledky správně kriticky zhodnotil. Pozitivně hodnotím, že experimenty nebylo těžké zreplicovat. Metodice provedených experimentů nelze nic vytknout. Dodaný program je dobře zdokumentovaný a logicky strukturovaný, takže je snadné se v něm orientovat. K programové části mám jen drobné výtky. V seznamu závislostí chybí několik knihoven (neúplný soubor *requirements.txt*), což komplikuje první spuštění aplikace. Implementace není příliš efektivní (například opakované předzpracování různých variant stejného datasetu).

Práce splňuje zadání ve všech bodech. Největší nedostatky vidím v teoretické části práce. Přes výše zmíněné výtky práci považuji za velmi dobrou. Autor prokázal, že je schopný samostatné vědecké činnosti. Proto práci doporučuji k obhajobě a hodnotím klasifikačním stupněm

„velmi dobře“.

Doplující otázky:

1. V práci zmiňujete, že metody extrakce klíčových slov často odstraňují stop slova. Použil jste v práci odstranění stop slov?
2. V teoretické části zmiňujete mikro a makro průměrování F1 metriky s tím, že pro klíčová slova se častěji používá mikro F1. Jakou F1 míru používáte vy při evaluaci?

Ing. Ondřej Pražák
KIV FAV ZČU

V Plzni 30. května 2022