



Posudek oponenta závěrečné práce

Oponent práce: Mgr. Petr Novák, Ph.D.
Student: Jan Peřina
Název práce: Automatická detekce přeložených textů
Obor / specializace: Znalostní inženýrství
Vytvořeno dne: 7. června 2021

Hodnotící kritéria

1. Splnění zadání

- ▶ [1] zadání splněno
- [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

2. Písemná část práce

60/100 (D)

Práce je dobře strukturovaná, jednotlivé části na sebe logicky navazují. Jazyková stránka je zdařilá méně, v některých pasážích se vyskytuje se český slovosled, neobratné formulace, překlapy nebo vynechaná slova znesnadňující čtení. Tabulky jsou občas relativně daleko od odkazujícího textu. Rozsah textu je při dolní doporučené hranici. Oceňuji naopak práci se zdroji a přehledné vzorce a tabulky.

Některé kroky autor popisuje zbytečně obšírně (casefolding, lemmatizace, stažení dat), někdy by naopak zasloužily hlubší diskusi (volba dat a příznaků, práce s hyperparametry, zhodnocení přesnosti jednotlivých metod). Postrádal jsem též analýzu kvality použitého lokálního automatického překladače - potenciální plagiátor by spíše využil dostupné state-of-the art technologie (Google Translate, DeepL).

3. Nepísemná část, přílohy

80/100 (B)

Jupyter notebooky použité pro přípravu a analýzu dat jsou přehledné a kód efektivní. Používají se externí knihovny i vlastní implementace. Chybí ale komentáře jednotlivých kroků a zastřešující readme, které by popisovalo posloupnost jejich spouštění. To znesnadňuje další využití kódu pro případné pokračování a rozšíření projektu.

4. Hodnocení výsledků, jejich využitelnost

80 /100 (B)

Jedná se o zdařilou analýzu dvou aktuálních témat. Představené metody pro detekci strojového překladu z angličtiny do češtiny byly vyzkoušeny jen na dvou datasetech, ale přesto jsou slibné a mohou sloužit jako základ pro budoucí publikaci. U druhého problému, hledání původního přeloženého textu pomocí webového vyhledavače, sice autor uspokojivé řešení nenalezl, nicméně prozkoumal vícero slepých uliček, což může pomoci dalšímu výzkumu.

Celkové hodnocení

75 /100 (C)

Práce se zabývá dvěma dosud málo prozkoumanými tématy, a to jednak detekcí textu přeloženého z angličtiny do češtiny a jednak nalezení původního textu pomocí vyhledavače. Autor provedl zevrubnou rešerši literatury týkající se obou témat, u prvního z nich úspěšně navrhl a implementoval metodu vykazující na zkoumaných datech dobré výsledky. Písemná zpráva je stručnější a práce má výše zmíněné nedostatky, nicméně se domnívám, že splňuje požadavky kladené na bakalářskou práci a doporučuji ji za ni uznat.

Otázky k obhajobě

- Jak byly zvoleny konkrétní použité datasety? Proč byly obě úlohy řešeny na různých datasetech?
- Čím by se dala vysvětlit vysoká efektivita metody Random forest (tabulky 2.1 a 2.5)?
- V závěru se autor domnívá, že i pro druhou úlohu existuje efektivní řešení. Jaké by k němu mohly vést cesty?

Instrukce

Splnění zadání

Posudte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posudte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

Posudte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

Nepísemná část, přílohy

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

Hodnocení výsledků, jejich využitelnost

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Celkové hodnocení

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.