



Posudek oponenta závěrečné práce

Student: Martin Vastl
Oponent práce: Ing. Karel Klouda, Ph.D.
Název práce: Analýza diskusních komentářů na českých zpravodajských serverech
Obor: Znalostní inženýrství

Datum vytvoření: 16. 6. 2019

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
1. Splnění zadání	<u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
<i>Popis kritéria:</i> Posuďte, 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, posuďte 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.	
<i>Komentář:</i> Všechny body zadání byly bezpochyby splněny. Za zpravodajský server si autor vybral aktuálně.cz a za dvě zkoumané metody pak BERT a Doc2Vec. Výsledky jsou pak přehledně prezentovány v textu práce, zejména v přílohách.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
2. Písemná část práce	88 (B)
<i>Popis kritéria:</i> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posuďte, 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. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3. Posuďte, 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.	
<i>Komentář:</i> Práce se dobře čte a je dobře uspořádána. Čtenář tak získá základní přehled o strojovém zpracování jazyka (NLP) tak i o konkrétních použitých metodách. V rešeršní části (kapitoly 1 a 2) se občas vyskytují formální nedostatky, výsledky jsou pak ale prezentovány přehledně a srozumitelně; tyto části naopak považují za velmi povedené. Z nedostatků zmíním následující: - Na str. 5 v prvním odstavci je poslední věta zamotaná. - Chybí interpunkce u vzorců (1.5), (1.6), (1.10), ... - Na str. 11 je použita nevysvětlená zkratka LSA a pojem "invertovaný index", což by mohlo méně poučeného čtenáře znejistit. - Pojem softmax je hojně zmiňován v kapitole 1 ale zaveden až na str. 17 v kapitole 2. - Na str. 20 je použit podivný zápis sumy, navíc písmeno N hned v následujícím odstavci značí něco jiného.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
3. Nepísemná část, přílohy	95 (A)
<i>Popis kritéria:</i> 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	
<i>Komentář:</i> I když jsem z pochopitelných důvodů nezkoušel veškerý přiložený kód (v jazyce Python), vše vypadá přehledně a je dobře uspořádaně. Kód je dobře okomentován, jak ten v Jupyter noteboocích, tak ten ve skriptech.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>

4. Hodnocení výsledků, jejich využitelnost

95 (A)

Popis kritéria:

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.

Komentář:

Hlavní výsledek práce je ukázka toho, že aktuální metody NLP se dají použít již na poměrně sofistikované úlohy, aniž byste měli zázemí velké firmy se spoustou dat a obrovským výpočetním clusterem. Větší výpočetní výkon (např. možnost vytrénovat model BERT na vlastních datech) by samozřejmě pomohl, ale i ve skromných "studentských" podmínkách autor ukázal, že výsledky jsou relevantní a použitelné.

Hodnotící kritérium:

Způsob hodnocení – nehodnotí se

5. Otázky k obhajobě

Popis kritéria:

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odřázkami).

Otázky:

1) Co Vám bránilo v tom, stáhnout i data ze serveru idnes.cz, kde je podle všeho více článků, komentářů i kategorií (např. sport)?

2) Co byste udělal jinak, kdybyste měl větší (libovolně velký) výpočetní výkon?

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

6. Celkové hodnocení

95 (A)

Popis kritéria:

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.

Text hodnocení:

Práce je po všech stránkách povedená a proto ji navrhuji hodnotit známkou A, jako výbornou.

Podpis oponenta práce: