

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Ondřej Michálek
Název práce Biblické parafrázování
Rok odevzdání 2020
Studijní program Informatika
Studijní obor Obecná informatika

Autor posudku Mgr. Rudolf Rosa, Ph.D. Vedoucí
Pracoviště Ústav formální a aplikované lingvistiky

K celé práci

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	X			
Splnění zadání		X		
Rozsah práce <i>... textová i implementační část, zohlednění náročnosti</i>		X		
<p>Zadání vyžadovalo od studenta seznámení se s několika nástroji a koncepty počítačové lingvistiky v míře odpovídající spíše magisterskému studiu (word2vec, MorphoDiTa, slovní embedinky, morfologické značkování, tokenizace, lematizace, apod.) Autor praktickou práci s těmito nástroji zvládl, byť z textu práce je patrné, že ne vždy zcela porozuměl jejich fungování, pro účely bakalářské práce ale hodnotím jeho porozumění jako postačující.</p> <p>Během vývoje aplikace se ukázalo, že biblické texty jsou pro natrénování příslušných modelů (word2vec) pro danou úlohu příliš malé. Student proto navrhl, implementoval a odladil jednoduchou metodu váženého kombinování doménových dat (biblické texty), dat příbuzné domény (náboženské texty) a velkých dat obecné domény (zejména filmové titulky), tak aby natrénovaný model vhodně modeloval český jazyk, ale zároveň tak aby biblické texty měly dostatečnou váhu a nebyly v modelu převáženy výrazně většími obecnými daty. Jelikož pro tento problém nejsou dosud ustáleny přímočaré postupy zaručující dosažení dobrých výsledků, student v rámci práce samostatně provedl řadu experimentů za účelem nalezení vhodné metody kombinování datových zdrojů a nastavení hyperparametrů nástroje word2vec.</p> <p>Výsledné parafrázování nefunguje zdaleka dokonale, což autor v práci přiznává. Ukazuje se, že problém parafrázování je příliš obtížný na to, aby byl uspokojivě vyřešen jen za pomoci použitých nástrojů, zejména v situaci jisté datové nouze ohledně textů používajících podoně archaický jazyk jako bible. Přesto jsou výstupy v mnoha ohledech dobré, a v některých případech je lze i považovat za zdařilé parafráze. Zadání tak hodnotím jako v rozumné míře splněné.</p>				

Textová část práce

lepší OK horší nevyhovuje

Formální úprava	<i>... jazyková úroveň, typografická úroveň, citace</i>		X		
Struktura textu	<i>... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>		X		
Analýza		X			
Vývojová dokumentace			X		
Uživatelská dokumentace			X		
<p>Práce je psána čtivě a srozumitelně, byť některá vyjádření jsou poněkud vágní či mírně nepřesná. Text je dobře strukturován a je typograficky přijatelný. Text obsahuje určité množství chyb a překlepů.</p> <p>Zvláště oceňuji podrobné vyhodnocení výsledků práce v kapitole 7, kdy autor nechal různé výstupy aplikace ohodnotit několika nezávislými hodnotiteli podle několika vhodně zvolených kritérií. Oceňuji i upřímné přiznání částečného neúspěchu: „Kromě respondenta C si žádný z respondentů nemyslí, že by se modernizace libovolnou metodou výrazněji zdařila.“</p>					

Implementační část práce

lepší OK horší nevyhovuje

Kvalita návrhu	<i>... architektura, struktury a algoritmy, použité technologie</i>		X		
Kvalita zpracování	<i>... jmenné konvence, formátování, komentáře, testování</i>		X	X	
Stabilita implementace				X	
<p>Aplikace používá vhodné nástroje a programátorské postupy (C#, word2vec, REST API, Win Forms, apod.) a je multiplatformní (Windows, Linux). Zdrojový kód je dobře strukturován a částečně pokryt unit testy, nicméně bylo by vhodné jej ještě pročistit – odstranit nepoužívané části kódu, sjednotit formátování, zlepšit pokrytí komentáři, apod.</p> <p>Samotná aplikace je funkční a opatřena nápovědou, nicméně její ovládání by mohlo být pohodlnější a intuitivnější. Aplikace je poměrně stabilní, až na situaci, kdy po natrénování modelu a návratu do menu je toto deaktivované a pro další práci je nutné aplikaci restartovat (je možné, že problém nastává pouze na Linuxu).</p>					

Celkové hodnocení Velmi dobře

Práci navrhuji na zvláštní ocenění Ne

Datum

Podpis