



POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Jméno studenta: Bc. Jakub Nevyhoštěný

Název práce: Text miningové nástroje pro tematickou analýzu česky psaných textů

Autor posudku: Ing. Martina Husáková, Ph.D.

Cíl práce: Nabídnout komplexní přehled metod a nástrojů určených pro předzpracování a vektorizaci anglických a českých textů za účelem klasifikace a modelování témat. Prakticky aplikovat tyto nástroje na reálných datech, výsledky porovnat a stanovit doporučení pro analýzu česky psaných textů pro potřeby výzkumu.

Povinná kritéria hodnocení práce	Stupeň hodnocení (známka)					
	A	B	C	D	E	F
Práce svým zaměřením odpovídá studovanému oboru	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vymezení cíle a jeho naplnění	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování teoretických aspektů tématu	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod, způsob jejich použití	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka a správnost provedené analýzy	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logická stavba a členění práce	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jazyková a terminologická úroveň	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vlastní přínos studenta	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Využitelnost výsledků práce v teorii (v praxi)	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Vyjádření k výsledku anti-plagiátorské kontroly

Anti-plagiátorská kontrola vykazuje 2% podobnosti s jinými pracemi v systému Odevzdej.cz.

Dílčí připomínky a náměty:

- Bylo by vhodné použít jinou strategii pro umísťování zdrojů, aby bylo zřejmé, k jaké části textu se zdroj přímo váže, resp. např. zdroj [13]/str. 4 se váže k větě předchozí nebo následující?: „...slovní zásobou jazyka. [13] Například slovo „jazyk“ může referovat k jazyku v ústech...“.

Celkové posouzení práce a zdůvodnění výsledné známky:

Diplomant vytvořil kvalitní práci, ve které se zabývá vybranými nástroji/knihovnamy/modely, které jsou určeny pro klasifikaci a modelování témat. Primárně se orientoval na analýzu jejich schopností předzpracovat české texty a provést pro ně tematickou analýzu. Diplomant aplikoval vhodné techniky pro předzpracování českých nestrukturovaných textů ve vybraných nástrojích. Výstupy předzpracování následně využil vybranými modely určenými pro klasifikaci a modelování témat.

Každý krok vhodně a detailně vysvětlil. Za výrazné pozitivum práce lze jistě považovat popsané zkušenosti autora z pohledu co daný nástroj/knihovna pro tematickou analýzu českých textů nabízí a co nenabízí. Pokud nenabízí, diplomant hledal alternativní řešení, která rovněž dobře dokumentoval. Prakticky tedy využil sadu modelů z knihovny Scikit-learn, Keras a NLTK pro klasifikaci témat. Dále pak opět knihovny Scikit-learn, Tomotopy a Gensim pro modelování témat. Výsledky získané jednotlivými modely vzájemně porovnal, čili si lze pak udělat určitou představu o tom, který model je do jaké míry vhodný pro daný český korpus. Porovnání je viditelné zejména v tab. 4 a 5 na str. 46 a 47. Práci považuji za velmi povedenou a užitečnou pro ty, kteří se chtějí o problematiku analýzy témat v českých textech více zajímat.

Otázky k obhajobě:

Pro účely klasifikace témat jste vybral recenze z ČSFD. Pro účely modelování témat jste vybral korpus obsahující české verše. Nebylo možné využít i pro modelování témat recenze z ČSFD?

Práci doporučuji k obhajobě.

Navržená výsledná známka: A

V Hradci Králové, dne 9. května 2024

podpis