



POSUDEK VEDOUcíHO DIPLOMOVÉ PRÁCE

Jméno studenta: Bc. Jakub Nevyhoštený

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

Autor posudku: Mgr. Jiří Haviger, Ph.D.

Cíl práce: Poskytnout komplexní přehled metod a nástrojů pro předzpracování a vektorizaci textů psaných v češtině a angličtině, jakož i pro jejich klasifikaci a modelování témat. Tyto nástroje aplikovat na skutečných datech a porovnat jednotlivé výsledky. Nakonec na základě výsledků navrhnout postup, kterým by se měl ubírat výzkumník, jenž chce provést tematickou analýzu česky psaných textů.

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

Vyjádření k výsledku anti-plagiátorské kontroly

Celková podobnost 2%

Dílní připomínky a náměty:

Struktura práce je v pořádku, čtenář se v práci snadno orientuje. Teoretická část 21 stran, metody 13 stran a výsledky + diskuze 8 stran.

V práci se objevují pouze drobné nedostatky v terminologii (např. str. 24 neviděl jsem používat český výraz „batoh slov“, běžně se používá anglické bag-of-words nebo jen zkratka BoW, což autor v dalším textu také používá).

Při popisu autor využívá spíše textu než vzorců, na druhou stranu je text vhodně doplněn ukázkami implementace v jednotlivých knihovnách.

Praktické testování provádí autor na datasetu českých filmových recenzí (klasifikace s učitelem) a na souboru básní (modelování bez učitele).

V závěru autor příliš doporučení nedává, neboť jednotlivé metody se ukazují jako specializované a nelze tedy provést snadné doporučení jediné z nich

Celkové posouzení práce a zdůvodnění výsledné známky:

Práce splňuje požadavky na diplomovou práci v plném rozsahu.

Otázky k obhajobě:

Popište možné důvody, proč hluboké učení s Keras nemělo v klasifikaci tak dobré výsledky, jako support vector machine.

Práci doporučuji k obhajobě.

Navržená výsledná známka: A

V Hradci králové, dne 15. května 2024

podpis