



POSUDEK VEDOUcíHO BAKALÁŘSKÉ PRÁCE

Jméno studenta: Jakub Včelák
Název práce: Detekce pojmenovaných entit pomocí knihovny spaCy
Autor posudku: Martina Husáková
Cíl práce: Cílem práce je provést průzkum možností extrakce pojmenovaných entit z textů, které nabízí knihovna spaCy. Dále pak je cílem vytvořit vlastní extraktor ve zmíněné knihovně.

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 0% podobnosti s jinými pracemi v systému Odevzdej.cz.

Díličí připomínky a náměty:

Pouze na pár místech v práci se vyskytují interpunkční nedostatky nebo také překlepy, z nichž některé mají humorný podtón:

- str. 3: „Architektura *neutronových* sítí...“
- str. 4: „Dalším díličím cílem, je trénování detekce nové entity, kterou nebylo dosud možné vyhledávat, v knihovně spaCy.“
- str. 37: „... (jádér aplikace)...“
- str. 39: „Při využití nejpřesnějších modelu, které spaCy nabízí, se...“

Celkové posouzení práce a zdůvodnění výsledné známky:

Bakalantovi se podařilo dosáhnout cíle, který byl na počátku stanoven. Teoretická část práce je detailně zpracována na prvních 21 stranách. Srozumitelně zde vysvětluje klíčové techniky (postupy), které se týkají zpracování přirozeného jazyka a detekce pojmenovaných entit. Lehce se dotýká i

oblasti porozumění přirozeného jazyka. Stanovuje tři hypotézy, na které odpovídá pomocí studia literatury a vývojem vlastního extraktoru pojmenovaných entit, který je pojmenován jako NER_It, a detailně popsán v části praktické. Extraktor integruje schopnost detekovat pojmenované entity vybraných typů v široké škále jazyků, včetně jazyka českého, jehož jazykový model není knihovnou spaCy zatím nabízen. U jazyka českého se bakalant soustředil na detekci osob (pojmenovaný typ Person). Pro vývoj extraktoru dokázal vhodně propojit poměrně pestrou paletu vývojových prostředků, tj. kromě knihovny spaCy, dále Flask, React, GIT, Heroku CLI, Gunicorn (WSGI server), CNN, Transformer a Ner-annotator. Bakalant si poradil i s hardwarovou (časovou) náročností při použití Transformer modelu, který použil při trénování. Poslední zmíněný prostředek Ner-annotator bakalant vhodně zvolil z důvodu aktuálně ne zcela vyhovujícího komerčního řešení pro anotaci dat (Prodigy), které je v nabídce od tvůrců knihovny spaCy. Velmi kladně hodnotím prezentaci cenných zkušeností s vývojem extraktoru, prezentaci pozitivních a negativních stránek knihovny spaCy a její dokumentace. Dále pak je užitečné srovnání výsledků při použití CNN a Transformeru. Co se týká průběhu tvorby bakalářské práce, tak bakalant prezentoval dílčí kroky tvorby práce pravidelně. Velmi chválím jeho precizní přístup ke zpracování tématu. Předkládanou bakalářskou práci hodnotím jako velmi zdařilou po stránce teorie i praxe. Svým obsahem a náročností bych ji hodnotila na úrovni práce diplomové. Jak je ze závěru patrné, je tu další prostor pro pokračování v tématu.

Otázky k obhajobě:

1. Jak byla prováděna anotace 1408/81 entit, které byly následně použity pro fázi trénování/testování?
2. Doporučil byste využít knihovnu spaCy pro vývoj detektorů pojmenovaných entit s jejich následným nasazením do ostrého provozu na web?

Práci doporučuji k obhajobě.

Navržená výsledná známka: A

V Hradci Králové, dne 16. května 2023

podpis