

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Automatizované ověřování faktů daty z české Wikipedie
Jméno autora:	Radek Štulc
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra kybernetiky
Vedoucí práce:	Ing. Herbert Ullrich
Pracoviště vedoucího práce:	Katedra počítačů – AIC

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadání požaduje trénink, evaluaci a zveřejnění českého modelu na rozpoznání pojmenovaných entit (NER), což je poměrně standardní úloha v NLP v jiných jazycích. Zadání tedy hodnotím jako průměrně náročné, ač citelně přesahující rámec bakalářského kurikula OI ZUI.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Splněno s pochvalou, modely RobeCzech natrénované na NER datech dosahují State-of-the-art přesnosti a jsou tedy přínosem pro naše další práce!	

Aktivita a samostatnost při zpracování práce	A - výborně
<i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	
Student byl během řešení BP velmi aktivní, účastnil se pravidelných konzultací dobře připraven, nadto udržoval komunikaci s garantem Janem Drchalem a kapacitou v oboru Janou Strakovou (ÚFAL MFFUK) která etablovala předchozí sota řešení	

Odborná úroveň	B - velmi dobře
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Student výborně využívá znalosti získané studiem OI ZUI i nad jeho rámec – statistika, programování v Pythonu, zpracování NLP dat, trénink neuronové sítě typu transformer. Drobné výhrady mám obecně k zmatečné argumentaci poznatků a autorových rozhodnutí směrem ke čtenáři, kde bych osobně preferoval větší rigor, srozumitelnost a transparentci (jako vedoucí ale mohu potvrdit, že postupováno bylo korektně). Celkově je ale odborná úroveň práce velmi dobrá!	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Text je psaný obstojnou angličtinou, některé části textu jsou poněkud matoucí, v některých jde poznat „czenglish“. Celkově je však text srozumitelný, forma je v pořádku a sazba krásná!	

Výběr zdrojů, korektnost citací	A - výborně
<i>Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.</i>	
Zdroje jsou relevantní, převažují konferenční články z oboru NLP, což je pro dnešní práce v tomto rychle krácejícím oboru typické. I když je pravda že jen málo z nich je z let 2023-2024 kdy BP probíhala a boom v oboru nadále	

pokračoval, mohu říct, že úkol rozpoznávání pojmenovaných entit není v tuto chvíli horkým tématem a práce tedy korektně využívá práce z doby předtím než se paradigma NLP posunulo z konkrétních úkolů do obecného generování textu. Je to na poptávku naší laboratoře, která hledá řešení na škále malých LMs jen pro tento specifický úkol a v ČJ.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.

Student pracoval po celý semestr pilně, výsledkem jsou SOTA modely pro rozpoznání pojmenovaných entit, které naplňují naše aplikační požadavky a budou dobře využitelné pro další výzkum. Nadto student probádal i rozpoznávání vnořených pojmenovaných entit (jako „Univerzita Jana Evangelisty Purkyně v Ústí nad Labem“) a i zde dosáhl slibných výsledků.

Text práce je (s drobnými jazykovými a argumentačními výhradami zmíněnými výše) dobře napsaný i vysázený, experimenty jsou reprodukovatelné pomocí publikovaných Jupyter notebooků.

S předloženou prací jsem velmi spokojen a hodnotím ji klasifikačním stupněm **B - velmi dobře**.

Jako její vedoucí přirozeně nemám žádné dotazy.

Datum: 4. června 2024

Podpis:

