

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Automatizované ověřování faktů daty z české Wikipedie
Jméno autora:	Tomáš Mlynář
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	KK
Oponent práce:	Gustav Šír
Pracoviště oponenta práce:	KP

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Uloha je sama o sobě velmi náročná, avšak práce staví na velkém množství již existujících řešení, náročnost lze tedy těžko posoudit.	

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 bez vyhrad.	

Zvolený postup řešení	vynikající
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Postup je výborný, student systematicky zkouma stavající state-of-the-art a informovaně nad ním staví vlastní pipeline.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Jedna se především o prakticky zaměřenou práci s metodami strojového učení a NLP, vhodná pro bakalářskou úroveň. Student však zjevně musel nastudovat i vědecké články a frameworky za hranice studijních osnov.	

Formální a jazyková úroveň, rozsah práce	A - výborně
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psána velmi precizně, skvělou angličtinou, inteligentně členěna, rozsahem vycerpávající, typograficky velmi povedena, i s citem pro grafiku/barvy. K dokonalosti už jen snad ze footnotes vždy až za interpunkci, a odkazy na Section/Chapter/Table/Figure s kapitálkami.	

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 též vycerpávající, student se zjevně i snažil odlišit vlastní přínos, ale stejně je těžké se v tom zorientovat, kvůli kombinování opravdu velkého množství stavajících zdrojů.	

Další komentáře a hodnocení
<i>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.</i>
Viz celkové hodnocení.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Jedna se o prakticky zameranou práci na pomezí ML/NLP. Práce postihuje opravdu širokou škálu činností v oblasti automatického overování faktu, a její zaber a rozsah jsou vyčerpávající. V zadné oblasti sice nejde výrazněji do hloubky, spis přímo stavi nad existujícími řešeními, které reviduje dle stavajících potřeb skupiny, avšak každá část je zpracována velmi přesně. K práci tak nemám žádné výhrady.

Souhlasím s vytyčeným future work, kvůli čemuž asi nemám ani dotazy. Porovnání vůči novým LLMs by mělo být velmi zajímavé, a to nejen z hlediska performance, ale i explainability odvození, za hranice jednoduchého matchování slov (kde tak zhruba končí použité feature attribution metody). V retrieval části by mi pak, v rámci posunu k dense metodám, přišlo zajímavé prozkoumat použití nové vznikajících vektorových databází.

Student si může zvolit, jestli chce reagovat.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 6.6.2023

Podpis: