

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Evaluation of Readers' Reactions to the Content of Media News
Jméno autora:	Jakub Ambroz
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Department of Cybernetics
Oponent práce:	Ing. Jan Drchal, PhD.
Pracoviště oponenta práce:	Department of Computer Science

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Pro bakaláře je toto zadání náročnější. Metody NLP nejsou součástí bakalářské výuky.	

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 výhrad.	

Zvolený postup řešení	částečně vhodný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Se zvoleným řešením práce souhlasím, je však postaven na starších, méně přesných metodách. Pro analýzu sentimentu jsou použity pouze metody založené na lexikonu a v potaz jsou brány pouze nekontextové reprezentace (Word2Vec). Předpokládám, že lepších výsledků by bylo dosaženo některým z volně dostupných předučených multilinguálních modelů typu Transformer, které berou v potaz kontext celého textu, případně naučením vlastního modelu (viz např. publikace ÚFAL http://hdl.handle.net/11234/1-4601). Důležitou částí práce je scrapper, ale o jeho architektuře a vlastnostech je v textu uvedeno velmi málo (viz otázka na konci posudku). Rozsah analytické (experimentální) části je menší – student se zaměřuje téměř výjimečně na pozitivní a negativní slova.	

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>	
K odborné úrovni nemám větší výhrady. Vybrané metody jsou používány správně.	

Formální a jazyková úroveň, rozsah práce	D - uspokojivě
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psána obstojnou angličtinou s občasnými překlepy, které nemají na srozumitelnost textu zásadní vliv. Je dobře strukturována. Rešeršní část je v některých, s prací relevantních, částech příliš stručná (Word2Vec), jindy zbytečná (model neuronu, sekce 3.4 popisující metody jako LDA, které nejsou dále použity). Problémem je to, že většina sekcí je psána na základě jediného zdroje. Za nedostatek rovněž považuji chybějící část rešerše, která by shrnula existující přístupy k analýze komentářů metodami NLP. Implementační část popisuje jednotlivé nástroje, ale postrádám informace o jejich integraci. Formální zápis není takřka přítomen. Z hlediska typografie nemám k práci námítky, kromě chybějících popisů os u grafů. U tabulky 7.4 není zcela zřejmý význam čísel, každopádně jejich přesnost je zbytečně veliká.	

Výběr zdrojů, korektnost citací	B - velmi dobře
<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</i>	

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.

Zdroje jsou citovány správně, často se však jedná o starší publikace.

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.

Vytvoření scrapperu pro dva velké české zpravodajské portály, stažení a přípravu dat považuji za velký přínos této práce, na kterém se dá dále stavět.

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.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **C - dobře**.

K práci mám následující otázky:

- 1) Jaká je architektura scrapperu? Řešil jste paralelní stahování více stránek? Jak řešíte chyby stahování? Jak dlouho trvalo získat korpusy uvedené v tabulce 6.1?
- 2) Velikost stažených dat (počty článků a komentářů) je relativně malá – jednalo se o nějaké specifické období, případně rubriku?
- 3) Jakou strukturu mají stažená data? Obsahují například identifikátory komentujících uživatelů?
- 4) V sekci 3.4 „Semantic Analysis“ prezentujete metody topic detection jako LSA a LDA. Dále v textu jsem nenarazil na jejich použití. Prováděl jste s nimi nějaké experimenty, které se již nedostaly do textu práce?

Datum: 31.5.2023

Podpis: