

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Semantic Clustering of Twitter Data
Jméno autora:	Jan Petrov
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Vedoucí práce:	Ing. Jan Drchal, PhD.
Pracoviště vedoucího práce:	Katedra počítačů

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Práce vyžaduje dobrou znalost moderních metod strojového učení pro zpracování přirozeného jazyka, které nejsou systematicky pokryty výukou na FEL. Zadání čítá tvorbu specializovaného anotačního nástroje, samotnou anotaci dat a experimenty s běžnými shlukovacími metodami i moderní metodou supervizovaného shlukování.	

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>	
Zadání bylo splněno bez výhrad.	

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 po celou mimořádně aktivní a na všechny konzultace byl vždy připraven. Do tématu se rychle ponořil a samostatně vyhledal a nastudoval velký objem současně nejnovější literatury na téma NLP i mnohem méně pokryté téma supervizovaného shlukování. Byl důležitým členem studentské NLP skupiny, již jsem vedl.	

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>	
K odborné úrovni nemám námitek. Rozsah odvedené práce je nadstandardní, ať už se jedná o různé přístupy sémantického vyhledávání, které naimplementoval v rámci anotačního rozhraní, anotovaného datasetu (čítajícího cca 18,000 tweetů), či experimentů se shlukovacími metodami. Velmi oceňuji studentem navržené a otestované modifikace algoritmu DAC postavené na kontrastivním učení. Po menších rozšířeních má práce publikační potenciál.	

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 výbornou angličtinou. Text je dobře strukturovaný a čtivý. Práce je doplněna řadou originálních ilustrací. Výborná je úvodní část, ve které student detailně popisuje relevantní témata NLP i metod shlukování. Za velmi přínosný považuji text, popisující detaily metody DAC (a to nad úroveň původního článku). Kapitoly druhé části popisující vlastní řešení a experimenty, jsou opět detailní. Všechny kroky návrhu řešení jsou dobře motivovány. Z hlediska reprodukovatelnosti, je práce nadstandardní.	

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</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.

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.

Celkový rozsah odvedené práce je nadstandardní, přičemž některé výstupy se v textu, z důvodu zachování kompaktnosti, vůbec neobjevily: např. specializovaný sémantický model pro filtraci tweetů (fact-check-worthy tweets).

Výstupy práce využijeme zejména v rámci projektu Central European Digital Media Observatory (CEDMO), v rámci kterého pracujeme na asistivních nástrojích pro novináře.

III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE

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

S přístupem studenta i výstupy jeho práce jsem celkově mimořádně spokojen.

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

Datum: 8.6.2022

Podpis: