

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Shlukování textů ze sociálních sítí
Jméno autora:	Sarana Egor
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra kybernetiky
Oponent práce:	Ing. Jiří Kubalík, Ph.D.
Pracoviště oponenta práce:	CIIRC, ČVUT

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Jedná se o standardní zadání bakalářské práce. Cílem je experimentálně otestovat a vzájemně porovnat několik existujících metod pro embedding a shlukování krátkých textů ze sociálních sítí.	

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 splněno bez výhrad.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Ke zvolenému postupu nemám výhrad. Student nastudoval danou problematiku přesně dle zadání. Z dostupných zdrojů vytvořil svoji testovací datovou množinu. Pozitivně hodnotím, že hodně pozornosti věnoval různým typům filtrace, aby byla výsledná datová sada co nejmysluplnější. Na těchto datech potom otestoval zvolené algoritmy. Výsledky experimentů přehledně prezentoval formou tabulek a obrázků. Škoda jen, že vlastní analýza pozorovaných výsledků je hodně stručná.	

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>	
Po odborné stránce je práce na dobré úrovni. Z textu je zřejmé, že se studentovi podařilo velice dobře zorientovat v dané oblasti.	

Formální a jazyková úroveň, rozsah práce	A - výborně
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Formální i jazyková úroveň je velmi dobrá. Práce je psána velmi dobrou angličtinou. Text je dobře strukturovaný a srozumitelný. Dobře se čte.	

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>	
Výběr i počet citovaných zdrojů je dostatečný.	

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, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

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

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

Do diskuze mám následující dotazy:

- V kapitole 6.2 se hodně věnujete filtraci dat. Můžete odhadnout, nakolik by se pozorované výsledky lišily, kdybyste data nefiltroval? Reálná data, na kterých by se vaše modely používaly také nemusí být takto „pročištěna“.
- U filtrace typu „Pruning by token length“ používáte pouze tokeny s délkou 3 až 10. Znamená to, že zahazujete všechny tweety, které obsahují běžná slova jako unbelievable, disappointed, significantly, apod.?

Datum: 31.1.2024

Podpis: