

**I. IDENTIFIKAČNÍ ÚDAJE**

<b>Název práce:</b>	Metody zpracování přirozeného jazyka pro analýzu recenzí
<b>Jméno autora:</b>	Ondřej Jiří Beneš
<b>Typ práce:</b>	bakalářská
<b>Fakulta/ústav:</b>	Fakulta elektrotechnická (FEL)
<b>Katedra/ústav:</b>	KK
<b>Oponent práce:</b>	Gustav Šír
<b>Pracoviště oponenta práce:</b>	KP

**II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ**

<b>Zadání</b>	<b>průměrně náročné</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Jde o relativně standardní formát úlohy z oblasti NLP, zhruba odpovídající dovednostem pro BP, ač NLP není v osnovách.	

<b>Splnění zadání</b>	<b>splněno s menšími výhradami</b>
<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 je v principu splněno, avšak část s analýzou a evaluací je extrémně strohá, a část s deskriptory chybí.	

<b>Zvolený postup řešení</b>	<b>správný</b>
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Zvolený postup je v základu správný, použité metody dávají smysl, jen samotné provedení působí dosti nedbalým dojmem.	

<b>Odborná úroveň</b>	<b>C - dobře</b>
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Student používá několik pokročilých technik z oblasti NLP, které jsou za hranicí osnov bakalářského studia, a dávají tak práci slušný odborný rozměr, nicméně popis jejich použití, jakožto i celého postupu který je zahrnuje, působí naopak trochu ledabyle/amatérsky, a samotný odborný přínos studenta tak není moc zřejmý. Asi největší část práce je věnována scrapování a přípravě dat, což je samozřejmě činnost nutná a často problematická, nicméně nepřilíš odborná.	

<b>Formální a jazyková úroveň, rozsah práce</b>	<b>D - uspokojivě</b>
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Text práce je dosti strohý, což by nebyl sám o sobě problém, nicméně rovněž není příliš kvalitní a působí velmi nedbalým a uspěchaným dojmem - spíše jako záznam postupu pro soukromé účely, sloužící jako průběžný podklad, než výsledná odborná práce samotná. Myšlenky jsou v textu pospojovány občas dosti nahodile a nejasně, a pro skutečné pochopení studentova postupu by bylo zapotřebí ho rozvést mnohem podrobněji a formálněji.	

<b>Výběr zdrojů, korektnost citací</b>	<b>B - velmi dobře</b>
<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 taktéž trochu strohé, tvořené často internetovými články, což nicméně odpovídá SW charakteru práce.	

<b>Další komentáře a hodnocení</b>
<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.*

Předložená práce si kladla za cíl využití moderních NLP metod pro analýzu vlastností produktů z uživatelských recenzí. Student se zaměřil na konkrétní doménu farmakologie a přidružené reporty o lécích, což považuji za zajímavou a vhodnou volbu. Použil scrapovací nástroje pro získání dat, na jejichž analýzu využil výběr NLP technik, od velmi základních až po LLMs. Tento postup by zněl jako velmi pěkná BP, avšak samotný textový popis jeho provedení působí tak ledabylym dojem, že nemůžu hodnotit výše. Velmi strohé, či spíše chybějící, výsledky tomu také moc nepřidávají.

Můj dojem je, že se student zaseknul v části scrapování dat a na jejich NLP analýzu, jakož i textový popis postupu, už mu nezbylo moc času, což beru jako polehčující okolnost. Také trochu oceňuji, že student psal text zjevně sám, bez použití LLMs, jinak by asi neobsahoval věty typu "I have the molecules on which I am interested in data." či opakování slova „miliard“, a informace typu „...was consistently banning my accounts because I could not find any affable scraping delay. Therefore, I had to use different VPNs, which I was constantly switching.“, které by možná bylo naopak záhodno z odborné práce spíš vynechat :)

Otázky:

- Jaká úvaha stojí za konkrétním výběrem daných NLP metod?

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

Datum: 4.6.2024

Podpis: