

Posudek bakalářské práce

„Metody dolování dat pro analýzu textů“

(Ondřej Kozák)

Práce představuje metodiku a specifické možnosti textové analýzy v rámci dolování a třídění lingvistických dat z odborných studií v elektrotechnice. Autorské řešení jednotlivých praktických úkonů vychází z programovacího jazyka Python, včetně příslušných knihoven, a prostředí PyCharm; pro srovnání je nicméně vymezen i potenciál platformy Matlab.

Z hlediska struktury i vlastního uspořádání obsahu lze výklad označit za systematický a přehledný; teoretická část postupuje od obecného (byť velmi stručného) úvodu do problematiky ke konkrétním metodám předzpracování a analýzy textu, zatímco praktické kapitoly obdobně přecházejí od jednodušších aplikačních úloh ke složitějším. Rozbor analytických přístupů a prvků je celkově vyvážený: autor ve většině případů posuzuje funkční vlastnosti i vhodnost metody pro daný účel a na tomto základě pak provádí volbu prostředků k zajištění experimentu. Praktické partie přinášejí popis a hodnocení nejen standardních experimentálních úkonů, k nimž náleží především určení četnosti a významu slov, ale také komplexnějších zadání, jež zahrnují mj. stanovení podobnosti textu a definování jeho témat.

Bakalářská práce p. Kozáka se zakládá na rozsáhlém a kvalitním rešeršním výzkumu; autor prostudoval značnou část dostupné literatury k tématu a získané poznatky následně využil při experimentech, které vykazují aplikační potenciál při tvorbě a úpravách specializovaných písemných materiálů ve vědě a technice. Stanovené cíle a úkoly byly vhodným způsobem splněny, a tedy předložené výsledky lze hodnotit kladně; ucelené a vyrovnané pojetí celé práce, společně s perspektivními experimentálními výstupy, mohou s klidným svědomím označit za její nosný prvek.

Současně je ovšem třeba uvést, že i odevzdaná finální podoba textu k mé nelibosti zahrnuje řadu gramatických, syntaktických i stylistických nedostatků (např. „Převedení... se dělá, protože...“, „U velkých písmen by mohlo dojít, že...“, „tyto slova“, „v euklidovském prostoru“, „každý slovo textu“, „vyplívá“); na dané formální úrovni tedy text významně zaostává za prezentovaným obsahem.

Do budoucna zůstává prostor pro další relevantní náměty, např. zpracování chybovosti v odborných překladech.

Vzhledem k výše představeným skutečnostem, s přihlédnutím ke snaze a zájmu autora, po delší úvaze navrhuji hodnocení „A“.

V Brně dne 5. 6. 2022

Přemysl Dohnal