

Posudek oponenta bakalářské práce

Student: Homola Ján
Téma: Získávání znalostí z textových dat v prostředí jazyka Python (id 24998)
Oponent: Hynek Jiří, Ing., Ph.D., UIFS FIT VUT

- Náročnost zadání** průměrně obtížné zadání
Cílem práce bylo vytvořit experimentální aplikaci vyhodnocující vybrané metody předzpracování textu a dolování znalostí z textu. Pro tento účel student prostudoval vybrané metody předzpracování textu (např. čištění dat, odstranění stopslov, stematizace a lematizace), převodu textu do vektorové podoby (např. TF-IDF) a klasifikace textových dat (k-NN, Naive Bayes, Random Forest, SVM). Dále musel prostudovat technologie pro implementaci aplikace (Python, PyQT, Pandas, NLTK nebo scikit-learn). Zadání hodnotím jako průměrně obtížné a považuji ho za splněné.
- Splnění požadavků zadání** zadání splněno
- Rozsah technické zprávy** je v obvyklém rozmezí
- Prezentační úroveň předložené práce** 65 b. (D)
Teoretická část technické zprávy je na dobré úrovni. Student v ní představuje jednotlivé metody dolování znalostí z textu použité v této práci. Druhá část popisující postup řešení práce má následující nedostatky. Student spojuje popis analýzy, návrhu a implementace do jedné čtyřstránkové kapitoly, což nepovažuji za dostatečné. Následná kapitola týkající se experimentů je podrobná, nicméně ocenil bych uvést nějaké srovnání měřených výsledků s jinými experimenty a publikacemi.
- Formální úprava technické zprávy** 85 b. (B)
Technická zpráva je pečlivě vysázená, občas se vyskytují překlepy, chybějící čárky, případně jsou nevhodně vysázeny vícenásobné citace.
- Práce s literaturou** 90 b. (A)
Student uvádí 27 zdrojů. Velkou část z nich představují tištěné publikace a knihy.
- Realizační výstup** 75 b. (C)
Realizační výstup působí dobrým dojmem. Jedná se o jednostránkovou aplikaci implementovanou pomocí jazyka Python a PyQT. Předzpracování textu je řešeno pomocí knihoven NLTK a Pandas, klasifikační metody pomocí scikit-learn, výsledná vizualizace pomocí Matplotlib a Seaborn. Prostor pro zlepšení vidím v lepším způsobu porovnání metod (možnost vybrat více konfigurací zároveň, které pak aplikace přehledně porovná).
- Využitelnost výsledků**
Výsledky aplikace mohou posloužit pro experimentální účely (např. kdy je žádané ověřit vhodnost vybraných klasifikačních metod nad vybranou datovou sadou se zvolenými konfiguracemi).
- Otázky k obhajobě**
 - Existují nějaké podobné nástroje, případně výsledky experimentů Vámi testovaných metod, s kterými byste mohl porovnat Vaše výsledky uvedené v kapitole 6 (Experimenty)?
 - Bylo by možné aplikaci upravit tak, aby umožňovala spouštět více testů různých metod zároveň a zobrazila porovnání jednotlivých výsledků přehledně na jedné obrazovce? Nastiňte možné řešení.
- Souhrnné hodnocení** 75 b. dobře (C)
Až na výše uvedené nedostatky hodnotím práci kladně. Student implementoval experimentální nástroj pro testování vybraných metod dolování znalostí z textu. Tyto metody dále s pomocí implementovaného nástroje analyzoval. Navrhuji hodnocení **stupněm C**.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 2. června 2022

Hynek Jiří, Ing., Ph.D.
oponent