

Posudek oponenta diplomové práce

Autor práce: Bc. Jakub Hain

Název práce: Analýza sentimentu na sociální síti Twitter

Student Jakub Hain v rámci své diplomové práce prozkoumal možnosti získávání příspěvků ze sociální sítě Twitter, metody pro klasifikaci příspěvků dle obsaženého sentimentu a metody pro získání extraktivních souhrnů pozitivních a negativních příspěvků. Tyto poznatky student využil při implementaci aplikace, která umožňuje získat aktuální příspěvky odpovídající zadanému dotazu, klasifikovat je dle sentimentu a extrahovat z nich souhrn pozitivního a negativního mínění.

Text diplomové práce je logicky členěn, má očekávaný rozsah (59 stran) a obsahuje všechny informace potřebné pro pochopení řešeného problému, který není triviální. Použité zdroje informací jsou relevantní, je jich rozumné množství a v práci jsou řádně odkazovány. Po formální stránce je práce v pořádku. Nicméně textu práce musím vytknout pravopisné chyby či překlepy (min. 30 pravopisných chyb a min. 30 chyb ve větných čárkách) a časté nepoužívání větných spojek (věty jsou často spojovány jen využitím čárky). Také zkratky jsou často psány malými písmeny nebo velikost jejich písmen není v textu konzistentní. Za zbytečný považuji velmi podrobný popis komponent použitých pro tvorbu GUI (viz část 8.7). S přihlédnutím nejen k těmto nedostatkům hodnotím text diplomové práce studenta Jakuba Haina jako dobrý až velmi dobrý.

Kvalitu implementovaného řešení student stanovil využitím anotovaných datových kolekcí a své závěry diskutoval v textu práce. Vytvořená aplikace je funkční a její zdrojové kódy jsou komentovány. Dle mého názoru tímto student splnil bez výhrad zadání práce. Protože ale student nenavrhoval žádnou vlastní úpravu algoritmů použitých pro klasifikaci sentimentu nebo extrakci souhrnu, tak při hodnocení vytvořené aplikace se přikláním ke známce velmi dobře.

Na základě zde popsaných skutečností hodnotím diplomovou práci studenta Jakuba Haina známkou velmi dobře.

Dotazy k práci:

- Na str. 40 textu práce píšete, že klasifikátor Komplementární Naivní Bayes určuje třídu podle nejméně vhodné třídy pro vzorek a Vy jste toto upravil tak, že je naopak využívána třída nejméně vhodnější. Vysvětlete, prosím, důvod této úpravy.
- Řekněme, že máme v kolekci dat použito 10 klasifikačních tříd, z nichž jedna svým objemem tvoří např. 90% dat kolekce. Jak takto objemná třída ovlivní výpočet mikro a makro průměrů?
- Čím je ovlivněno, zda Twitter streaming API poskytne 1% nebo 40% z celkového objemu zpráv Twitteru?

Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 30. 5. 2019

Ing. Michal Nykl, Ph.D.