

Posudek oponenta bakalářské práce

Student: Drápela Karel

Téma: Strojové učení v oblasti stylometrie a určování autorství (id 18623)

Oponent: Škoda Petr, RNDr., UPGM FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Zadání vyžaduje pochopení, zprovoznění a odladění SVM a dále pak hlubší pochopení významu různých charakteristik textů v přirozeném jazyce.
- 2. Splnění požadavků zadání** **zadání splněno**
Student naplnil všechny body zadání.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
Rozsah práce odpovídá potřebám daného tématu, práce obsahuje dostatek diagramů tabulek i grafů.
- 4. Prezentací úroveň předložené práce** **80 b. (B)**
Práce má dobrou strukturu a čtenáře provádí problematikou bez závažných nedostatků. Práce navíc obsahuje všechna potřebná témata včetně škálovatelnosti a porovnání výsledků s podobnými systémy.
- 5. Formální úprava technické zprávy** **55 b. (E)**
V práci se vyskytuje větší množství překlepů a několik hrubých pravopisných chyb. Někdy překlepy a opakovaná slova komplikují pochopení vět. Jazyková kvalita napříč prací značně kolísá (kvalitnější jsou první kapitoly). V práci se dále vyskytuje množství typografických nedostatků. Některé obrázky mají nízkou kvalitu.
- 6. Práce s literaturou** **95 b. (A)**
Student použil na poměry BP větší množství odborných článků týkajících se předmětu práce a velmi dobře z nich těžil v teoretické i praktické části.
- 7. Realizační výstup** **85 b. (B)**
Student vytvořil funkční systém umožňující experimentovat s určováním autorství textů. Dále připravil skripty pro stahování experimentálních dat z Twitteru a Quora.
- 8. Využitelnost výsledků**
Práce experimentuje s kombinací určování autorství s pomocí sad delších textů (Quora) a krátkých textů (Twitter), což je nový přístup k této problematice a výsledný systém vykazuje dobré výsledky v porovnání s podobnými systémy.
- 9. Otázky k obhajobě**
 - Rozveďte význam charakteristiky písemných n-gramů a diskutujte jejich efekt na krátkých textech z Twitteru v porovnání s delšími texty.
 - Stručně zdůvodněte příčinu lepších výsledků dosažených s metodou výběru příznaků "informační přínos" oproti jiným metodám.
- 10. Souhrnné hodnocení** **85 b. velmi dobře (B)**
Pozitivně hodnotím především nadprůměrně odborně zpracované zadání s využitím velkého množství odborných článků. Práce zároveň dosáhla dobrých výsledků v přesnosti určování autorství anonymních textů v porovnání s obdobnými systémy. Spíše negativně pak hodnotím průměrnou typografií a méně obratné vyjadřování v některých částech práce.

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

V Brně dne: 1. června 2016

.....
podpis