

Posudek oponenta diplomové práce

Student: Bolješik Michal, Bc.

Téma: Analýza postojů českých uživatelů k obchodním řetězcům na základě dat ze sociálních sítí a webových diskusí (id 18873)

Oponent: Otrusina Lubomír, Ing., UPGM FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
2. **Splnění požadavků zadání** zadání splněno
Zadání práce bylo splněno. Student navrhl a implementoval systém, který dokáže pravidelně získávat, indexovat a analyzovat data ze sociálních sítí.
3. **Rozsah technické zprávy** je v obvyklém rozmezí
Rozsah technické zprávy je v obvyklém rozmezí. Všechny části technické zprávy jsou informačně bohaté a pro práci většinou nezbytné. Konkrétní připomínky jsou popsány níže.
4. **Prezentační úroveň předložené práce** 65 b. (D)
Prezentační úroveň technické zprávy je spíše špatná. V kapitole 2 "Rozbor riešenej problematiky" student uvádí informace o několika systémech či přístupech pro analýzu sentimentu a indexování dat. Občas ale těmto informacím chybí kontext či jsou neúplné, což ubírá práci na srozumitelnosti. To se týká zejména podkapitoly 2.2 "Analýza postojov", ve které také postrádám podrobnější popis klasifikátorů, které student v práci používá. Očekával bych, že se v kapitole 3 "Návrh systému" dozvím, jaká data bude student v práci používat. To je ale uvedeno až v kapitole 4 "Implementácia systému". Kapitola 4 "Implementácia systému" obsahuje příliš podrobný popis nepřiliš zajímavých implementačních detailů, což ubírá práci na čtivosti. V kapitole 5 "Vyhodnotenie systému" student uvádí dosažené hodnoty správnosti natrénovaných klasifikátorů, ale není příliš jasné, jak tyto hodnoty určil. V kapitole 6 "Záver" jsou uvedeny podrobné informace týkající se implementace (např. kterým skriptem se systém spouští), což považuji za nevhodné.
5. **Formální úprava technické zprávy** 80 b. (B)
Formální úprava technické zprávy je velmi dobrá. Vytkl bych snad jen, že student občas nepřekládá některé anglické pojmy a že ukázky kódu nejsou v práci nijak číslovány. Jazykovou stránku práce nemohu zcela posoudit. V odevzdaném pdf souboru nefungují odkazy na reference, což znesnadňuje čtení.
6. **Práce s literaturou** 50 b. (E)
Práce obsahuje přiměřené množství studijních pramenů. V kapitole 2.2 "Analýza postojov" student čerpá mnohé informace ze zdroje [4], což je přehledová studie přístupů pro analýzu sentimentu. Student pak ve své práci uvádí popisy systémů zmíněných v tomto zdroji, ale ty už necituje. V některých místech také není jasné z jakého zdroje informace pochází. Seznam studijních pramenů není řazen dle abecedy. Některé položky ze seznamu referencí by se hodily spíše do poznámky pod čarou.
7. **Realizační výstup** 80 b. (B)
Realizační výstup práce je zdařilý. Student vytvořil systém, který dokáže pravidelně získávat, indexovat a analyzovat stahovaná data. Získaná data jsou vhodně vizualizována.
8. **Využitelnost výsledků**
Jedná se o práci kompilačního charakteru. Vytvořený systém umožňuje získávat zajímavé informace o postojích uživatelů k potravinovým obchodním řetězcům v České republice. Systém dokáže odhalit kauzy, které se daných řetězců týkají.
9. **Otázky k obhajobě**
 - Proč jste systém implementoval v jazyce Python 2 a ne Python 3?
 - V práci uvádíte, že používáte několik indexů v Elasticsearch. Kolik jich celkem je a proč nestačí jen jeden?
 - Můžete komisi objasnit jak jste přesně určil hodnoty správnosti systému, uvedené v kapitole 5 "Vyhodnotenie systému"?
10. **Souhrnné hodnocení** 55 b. dostatečně (E)
Prezentační úroveň technické zprávy je podprůměrná, naopak její formální úprava je velmi dobrá. Kvalita realizačního výstupu je velmi dobrá. Práce s literaturou obsahuje výše zmíněné prohřešky, které jsou hlavním důvodem k hodnocení stupněm E. Nebýt jich, bylo by hodnocení zcela jistě lepší.

V Brně dne: 7. června 2017

.....
podpis