

Posudek oponenta bakalářské práce

Student: Hajdin Martin**Téma:** Sémantická podobnost textů (id 18690)**Oponent:** Otrusina Lubomír, Ing., UPGM FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
2. **Splnění požadavků zadání** zadání splněno s drobnými výhradami

Cílem práce bylo navrhnout a realizovat systém, který dokáže utřídit zadané texty do skupin a navrhnout pojmenování identifikovaných skupin na základě společného obsahu. Zadání práce bylo splněno s drobnými výhradami. Student realizované řešení neporovnal s jinými metodami. Realizovaný systém je velmi jednoduchý.
3. **Rozsah technické zprávy** je v obvyklém rozmezí

Rozsah technické zprávy je v obvyklém rozmezí. Rozsah některých kapitol by ale bylo vhodné upravit. Např. část popisující časovou náročnost není příliš důležitá a přesto jí student věnuje několik stran. Na druhou stranu vyhodnocení přesnosti systému by mohlo být rozsáhlejší.
4. **Prezentační úroveň předložené práce** 50 b. (E)

Logická struktura technické zprávy je uspokojivá. Jednotlivé kapitoly na sebe navazují, ale občas by bylo potřeba upravit jejich rozsah (viz poznámka výše). Některé pasáže práce jsou hůře pochopitelné pro čtenáře. To je způsobeno zejména tím, že student často nevysvětluje rovnice a uvádí nepravdivá a nepodložená tvrzení. Zejména popisy složitějších algoritmů působí jakoby je student volně přebíral z odkazovaných zdrojů a nerozuměl jim.
5. **Formální úprava technické zprávy** 50 b. (E)

Po formální stránce má práce také několik nedostatků. Bezprostředně po nadpisu kapitoly 3 následuje nadpis kapitoly 3.1 a po něm hned nadpis kapitoly 3.1.1 (tento jev se v práci opakuje). Některá klíčová slova či názvy by stálo za to v textu zvýraznit. Tabulky působí často nepěkně. Kvalita přebíraných obrázků je velmi špatná. Občas chybí odkazy na zmiňované nástroje. Práce obsahuje číslování nadpisů až do 4. úrovně.
6. **Práce s literaturou** 50 b. (E)

Ačkoliv seznam literatury obsahuje relevantní vědecké publikace, student místo nich někdy cituje blogy. V některých pasážích student citace neuvádí, i když je zřejmé, že musel tyto znalosti odněkud čerpat.
7. **Realizační výstup** 40 b. (F)

Realizační výstup práce je velmi jednoduchý. Student nevytváří vlastní LDA model, stahuje již vytvořený anglický model a ten pak používá i na česká data, což nefunguje. Vytvoření vlastního modelu přitom není nijak složité. Student navíc na straně 18 sám přiznává, že aplikace obsahuje pouze základní funkcionalitu, některé její části nejsou dodělané a může obsahovat chyby. Implementované navrhování pojmenování identifikovaných skupin na základě společného obsahu je diskutabilní.
8. **Využitelnost výsledků**

Jedná se o práci kompilačního charakteru. V aktuálním stavu realizační výstup není příliš použitelný.
9. **Otázky k obhajobě**
 1. Nejsem si jistý, jestli správně rozumíte algoritmu SVD (viz např. věta "To je zabezpečené metodou SVD, která zredukuje rozmer vektorov na dve." - proč by měla být velikost vektoru pouze 2?). Můžete komisi v rychlosti objasnit, jak jste to myslel?
 2. V práci uvádíte: "Preto je vo výslednej aplikácii použitý algoritmus kmeans++, ktorého priebeh je v podstate lineárny.". Můžete prosím komisi objasnit, jaká je skutečná časová složitost algoritmu K-means++?
10. **Souhrnné hodnocení** 45 b. nevyhovující (F)

Student sice splnil zadání, ale pouze v nejmenší možné míře. Z práce je zřejmé, že student některým algoritmům nerozumí. V práci jsou mnohá nepravdivá tvrzení. Realizační výstup je velmi jednoduchý a jeho funkčnost je diskutabilní, což potvrzuje i provedené vyhodnocení systému. Z těchto důvodů navrhuji hodnocení stupněm F (45b).

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

V Brně dne: 1. června 2016

.....
podpis