

Posudek oponenta bakalářské práce

Student: Hošťák Viliam Samuel
Téma: Shlukování slov podle významu (id 14835)
Oponent: Otrusina Lubomír, Ing., UPGM FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Zadání práce bylo obtížnější a to zejména z důvodu nutnosti nastudovat a pochopit velké množství netriviálních algoritmů. Práci neusnadňoval ani fakt, že trénování některých modelů trvalo desítky hodin.
- 2. Splnění požadavků zadání** **zadání splněno**
Zadání práce bylo splněno, student se seznámil s potřebnými metodami a implementoval systém, který dokáže k zadanému slovu automaticky najít a zobrazit slova sémanticky příbuzná.
- 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 skutečně nezbytné.
- 4. Prezentací úroveň předložené práce** **80 b. (B)**
Prezentací úroveň technické zprávy je dobrá. Logická struktura práce i rozsahy a návaznosti jednotlivých kapitol jsou také v pořádku. Přesto mám ale k technické zprávě několik drobných výhrad. Technická zpráva obsahuje několik sporných tvrzení. Obr. 3.1 by dával větší smysl, kdyby obsahoval česká slova. Obr. 3.3 je nedostatečně vysvětlen, a je tak pro čtenáře těžko pochopitelný. Rovněž část práce popisující deskovou hru Krycí jména, je nedostatečně vysvětlena a čtenář se tak může místy ztráct. Výsledky pro jednotlivé metody jsou uvedeny ve zvláštních tabulkách. Hodilo by se i nějaké srovnání použitých metod v jedné tabulce, aby to bylo přehlednější.
- 5. Formální úprava technické zprávy** **85 b. (B)**
Formální úprava technické zprávy je velmi dobrá. Hlavní výhradu mám k tomu, že student často nepřekládá některé anglické pojmy (např. part-of-speech, stoplist, stochastic gradient descent). Obr. 4.4 je zbytečně velký.
- 6. Práce s literaturou** **85 b. (B)**
Práce s literaturou je také na velmi dobré úrovni. Student cituje velké množství relevantních vědeckých článků. Výjimečně ale na některých místech chybí citace (např. na systém HAL nebo hru Krycí jména).
- 7. Realizační výstup** **90 b. (A)**
Realizační výstup je na velmi dobré úrovni. Student implementoval velmi zajímavý systém, který umožňuje určovat sémanticky blízká slova pro češtinu pomocí několika state-of-the-art metod.
- 8. Využitelnost výsledků**
Jedná se o práci, která rozšiřuje již publikované výsledky a to v tom smyslu, že student v práci využil několik state-of-the-art metod pro určování sémantické blízkosti slov a aplikoval je na češtinu. Práce poskytuje velmi dobrý základ pro zpracování sémantické podobnosti v češtině.
- 9. Otázky k obhajobě**
 - Při trénování modelů bylo pro každou metodu potřeba nastavit mnoho parametrů. V práci uvádíte, jaké hodnoty jste vybraným parametrům nastavil, ale příliš své rozhodnutí nezdůvodňujete. Můžete shrnout, jaké hodnoty parametrů jste při trénování volil a proč jste nezkoušel i jiné hodnoty?
 - Jak si vysvětlujete, že model GloVe dosahuje v některých testech podstatně nižší úspěšnosti než Word2Vec a FastText?
- 10. Souhrnné hodnocení** **85 b. velmi dobře (B)**
Zadání práce bylo obtížnější, student musel nastudovat a pochopit velké množství netriviálních metod. Prezentací úroveň technické zprávy je i přes veškeré připomínky dobrá, její formální úprava je velmi dobrá. Práce obsahuje velké množství relevantních citací. Realizační výstup práce je na vysoké úrovni. Z těchto důvodů hodnotím práci stupněm B.

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

V Brně dne: 1. června 2017

.....
podpis