

Posudek oponenta bakalářské práce

Student: Smutka Miloslav

Téma: Určování blízkost pojmů v oblasti informačních technologií (id 18688)

Oponent: Otrusina Lubomír, Ing., UPGM FIT VUT

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

Cílem práce bylo vytvořit systém, který dokáže z kolekce relevantních textů v oblasti informačních technologií extrahovat termíny a navrhnout hierarchii pojmů vzhledem k jejich sémantice. Zadání práce bylo splněno s vážnějšími výhradami. Systém nevytváří žádnou hierarchii pojmů a nebyl srovnán s alternativními přístupy.
3. **Rozsah technické zprávy** je v obvyklém rozmezí

Rozsah technické zprávy je v obvyklém rozmezí, jen některé popisy algoritmů jsou příliš strohé. Práce obsahuje cca 30 stran příloh, které nejsou pro práci nezbytné a pro čtenáře nejsou zajímavé.
4. **Prezentační úroveň předložené práce** 50 b. (E)

Logická struktura práce je dobrá. Některé části práce jsou ovšem hůře pochopitelné a není jasné, jak to student myslel. Práce rovněž obsahuje mnohá sporná a nepravdivá tvrzení.
5. **Formální úprava technické zprávy** 60 b. (D)

Po formální stránce práce obsahuje několik prohřešků. Student nedělá mezery před otevírací závorkou, což působí dost rušivě. Některé obrázky jsou dost nepřehledné. V práci se také vyskytují překlepy a gramatické chyby.
6. **Práce s literaturou** 50 b. (E)

Seznam studijních pramenů je zvolen dobře, jen některé odkazy na použité technologie by se spíše hodily do poznámky pod čarou. Problémem ale je, že student v mnoha podkapitolách žádné zdroje necituje a není jasné, odkud čerpal.
7. **Realizační výstup** 50 b. (E)

Student vytvořil systém, který dokáže z kolekce relevantních textů v oblasti informačních technologií extrahovat termíny. Jedná se však o velmi jednoduchou implementaci, jejíž funkčnost je diskutabilní, protože nebylo provedeno řádné vyhodnocení systému. Student hodnocení shrnuje s tím, že všechny modely dosáhly průměrné úspěšnosti 90 %.
8. **Využitelnost výsledků**

Jedná se o práci kompilačního charakteru. Výsledky práce pravděpodobně nejsou dále využitelné.
9. **Otázky k obhajobě**
 1. Na straně 14 uvádíte, že optimální nastavení prahu pro zpracování korpusu se pohybuje v závislosti na počtu klíčových slov od 0.004 do 0.02. Jak jste na tato čísla přišel?
 2. V tabulce 5.1 uvádíte náhodně vybraná slova a sousloví z oblasti IT. Jejich výběr ale nevypadá příliš náhodně. Jak jste tato slova přesně vybíral?
10. **Souhrnné hodnocení** 50 b. dostatečně (E)

Zadání práce bylo splněno s vážnějšími výhradami. Systém nevytváří žádnou hierarchii pojmů a nebyl srovnán s alternativními přístupy. Implementovaný systém je velmi jednoduchý a jeho vyhodnocení je velmi subjektivní. V práci občas není jasné, z jakých zdrojů student čerpá. Z těchto důvodů navrhuji hodnocení stupněm E (50b).

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

V Brně dne: 2. června 2016

.....
podpis