

Posudek oponenta diplomové práce

Autor práce: Bc. Karel Zíbar

Název práce: Automatická extrakce klíčových slov pomocí metod trénovaných bez učitele

Obsah práce:

Práce se zabývá automatickou extrakcí klíčových slov z textu pomocí strojového učení bez učitele. Práce nás seznamuje s teorií sémantické analýzy textu, dále pak s metodami extrakce klíčových slov a popisuje metody určování významnosti slov. Dále pak popisuje použité korpusy dat, testování, experimenty a dosažené výsledky.

Formální úroveň:

Rozsah je v očekávaném rozmezí (55 stran vlastního textu). **Logická struktura** má pouze zanedbatelné nedostatky. **Stylistická stránka** práce má prostor pro zlepšení (překlepy, hrubky a nejednotnost v názvosloví).

Kvalita řešení a dosažených výsledků je nadprůměrná především složitostí programu a časovou náročností experimentů a množstvím experimentů. Kód je srozumitelný a bohatě okomentovaný.

Práce s literaturou:

Uvedené zdroje jsou relevantní a dostatečné. Nalezl jsem zde pouze jeden překlep.


Dotazy k práci:

- 1) Vysvětlete blíže rozdělení metod (strojového učení) pro extrakci klíčových slov a přiřadte všechny metody zmíněné v práci do nějaké kategorie.
- 2) Vysvětlete tvrzení ze strany 18 třetí odstavce (*Nevýhodou této metody je, že upřednostňuje více čtená slova dokumentu před těmi méně čtenými. To může mít ale za následek ignorování slov významově důležitých, která se v textu objevila například jen jednou...*).

I přes výše zmíněné výtky převážně k formální úrovni textu, považuji práci za velmi dobrou a dosažené výsledky za výborné. Zadání bylo splněno s mírnou výhradou v bodě 2, kde dostupné datové kolekce byly omezeny jen na SemEval 2010 a Wikipedii, proto navrhuji hodnocení známkou **velmi dobře** a práci **doporučuji k obhajobě**.

V Plzni 6. 6. 2017


Ing. Tomáš Hercig

**SOUHLASÍ
S ORIGINÁLEM** 

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
katedra informatiky a výpočetní techniky