

Posudek oponenta bakalářské práce

Autor/autorka práce: **Michel Soběhart**

Název práce: **Agregátor novinových zpráv**

Obsah práce

Práce se zabývá automatickým stahováním novinových zpráv z RSS zdrojů na internetu a jejich následným shlukováním podle podobnosti.

Kvalita řešení a dosažených výsledků

Navržené řešení a dosažené výsledky považuji za solidní a prokazují odbornou vyspělost studenta. Použité metody i způsob testování jsou vhodné pro danou aplikaci. Podle mě měl student věnovat podstatně víc času experimentům s různým nastavením systému. Jak se systém chová při použití jiného prahu (jak senzitivní je vůči danému prahu)? Jak se projevuje úspěšnost clusteringu při volbě jiných příznaků (např. se stemmingem a bez stemmingu)? V práci mi dále chybí statistiky trénovacího korpusu (počet dokumentů, počet slov, atd...)

Program na přiloženém CD funguje bez problémů.

Formální úroveň

Z formálního hlediska není práce příliš dobrá. Je zde spousta překlepů, vzorce nejsou číslované a navíc k nim velmi často chybí popis proměnných. Jednou jsou vektory značeny jako p, q a podruhé a, b . Kapitola 2.3.2 obsahuje uprostřed prázdnou stránku. Některé pasáže jsou nepřesně popsány (např. sekce 2.3.3, 2.3.4). Na několik obrázků v textu nevedou odkazy. Student nechápe význam slova dimenze.

Práce s literaturou

Student cituje převážně webové stránky nebo diplomové práce, což nepovažuji za šťastné. V referencích jsou pouze dvě skutečné publikace. V citaci číslo 7 jsou prohozená jména autorů.

Splnění zadání

Student splnil zadání.

Dotazy k práci

- 1) V práci popisujete několik metody clusteringu, přičemž jste použil úplně jinou, která shlukuje při překročení předem definovaného prahu. Proč?
- 2) Definujte pojem z lineární algebry – DIMENZE.

Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 28.5.2014

Ing. Tomáš Brychcín

