

Posudek vedoucího diplomové práce

Autor práce: **Bc. Martin Váňa**

Název práce: **Incremental News Clustering**

Cílem práce bylo seznámit se s moderními algoritmy shlukování textů, zejména s Chinese restaurant process a také distance-dependent variantou (ddCRP) a aplikovat je na úlohu shlukování zpravodajských textů.

V teoretické části autor popisuje nemalý matematický základ potřebný pro popis a implementaci algoritmů. Čerpáno bylo převážně z knihy K. Murphye [14]. Nastudování potřebné literatury (viz reference v DP) bylo časově i intelektuálně velmi náročné. Práce dále popisuje tři možnosti vektorové reprezentace dokumentu a metody vyhodnocení shlukování. Praktická část se věnuje především implementaci a dosaženým výsledkům.

Zadání bylo s ohledem na nutnost nastudování pokročilé matematiky v pozadí algoritmů složité a diplomant se s ním popral velmi dobře. Celou dobu řešení byl velice aktivní, až měl vedoucí pocit, že jej občas brzdí. Práci párkrát také konzultoval s T. Brychcínem, který má v problematice související s ddCRP největší přehled. Celá práce má výzkumný charakter a určitě převyšuje průměrnou diplomovou práci.

Ačkoliv je kapitola s výsledky rozsahem kratší, diplomant provedl velkou řadu experimentů (obširnější výsledky jsou v příloze a na CD). Nakonec dosáhl velmi dobrých výsledků s implementací ddCRP a vektory doc2vec. Bonusem bylo vytvoření evaluační metriky NV-measure. Experimenty limituje velikost korpusu. Diplomant pracoval s anglickou částí korpusu, vytvářením naší NLP skupinou, který bude sloužit pro vyhodnocení krosjazyčného shlukování zpravodajských textů.

Práce je psána v angličtině, což považuji za další plus, navíc v dobré angličtině. Formálně je v pořádku.

Diplomant ukázal schopnosti samostatné výzkumné práce. Práce by po rozšířeních diskutovaných v závěru mohla být publikována a téma by bylo i dezertabilní.

Navrhuji hodnocení známkou **výborně** a práci doporučuji k obhajobě.

V Plzni 5. 6. 2018

Doc. Ing. Josef Steinberger, Ph.D.



**SOUHLASÍ
S ORIGINÁLEM**



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

①