

Posudek vedoucího diplomové práce

Název DP: Shluky silně podobných textů
Diplomant: Jiří Diviš
Katedra: Ústav formální a aplikované lingvistiky
MFF UK, Praha 2006

Hlavním cílem diplomové práce bylo navrhnout a experimentálně implementovat prakticky použitelnou metodu pro automatické vyhledávání shluků silně podobných textů v přirozeném jazyce, a to v rozsáhlých textových kolekcích.

Diplomová práce je svým zaměřením výzkumná a experimentální. Úvodní druhá kapitola práce obsahuje stručný přehled známých metod shlukování a popis algoritmů, které pak autor využívá v dalších částech své práce. Ve třetí kapitole se autor podrobně zabývá definicí konceptotvorných shluků a algoritmickými postupy pro jejich vyhledávání, což je ústřední téma celé práce. Myšlenka konceptotvorných shluků byla vyvinuta a publikována již dříve (Holub, Semecký, Diviš: "Searching for Topics in a Large Collection of Texts", Proceedings of the ACL, 2004), a to v rámci projektu, na kterém diplomant aktivně participoval.

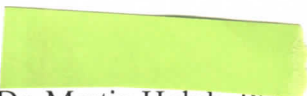
Těžiště práce a její největší přínos leží v kapitolách 4 a 5, kde se autor zabývá zpracováním experimentálních dat a provádí experimenty s hledáním konceptotvorných shluků v závislosti na volbě dílčích postupů a na nastavení řady parametrů celého procesu. Zabývá se také obtížným a samostatným problémem, jak hodnotit a srovnávat nalezené množiny shluků. Ke svým experimentům autor použil anotovaná data zpracovaná v diplomové práci Jana Pávka ("Testovací kolekce českých dokumentů a dotazů", MFF UK, 2004).

Po formální stránce předložená práce obsahuje vše, co obsahovat má, její struktura je přehledná a srozumitelná. Implementace a testovací data jsou obsaženy na přiloženém CD.

Závěr:

Předložená diplomová práce splňuje všechny požadavky, které jsou na ni kladené a hlavní její cíl vytyčený v zadání diplomant splnil. Práci doporučuji k obhajobě.

V Praze, 28.1. 2007


RNDr. Martin Holub, PhD