

Oponentský posudek diplomové práce

Využití syntaxe v metodách pro vyhledávání informací

Autor: Bc. Jana Kravalová

Vedoucí diplomové práce: Mgr. Pavel Pecina, Ph.D.

Matematicko-fyzikální fakulta Univerzity Karlovy, Praha 2009

Hlavní téma předložené diplomové práce, aplikace počítačové lingvistiky v oboru vyhledávání informací v textech (dokumentech), patří nepochybně mezi témata aktuální. Autorka se zaměřila na zkoumání poměrně specifického jazykového modelu založeného na závislostních syntaktických vztazích uvnitř české věty a užití tohoto modelu ke zlepšení automatického vyhledávání relevantních dokumentů. Cílem práce bylo navrhnout, implementovat a vyhodnotit metodu, která obohatí pravděpodobnostní model pro vyhledávání informací o automaticky získané informace o syntaktických vztazích, a to jak v rámci dokumentů, tak v rámci dotazů (viz kap. 1).

Předložená práce má standardní strukturu. Autorka nejprve uvádí vybrané základní koncepty a technologie z oblasti jazykového modelování a vyhledávání informací (kap. 2). Potom čtenáře seznamuje s daty, která byla použita k experimentům (kap. 3). Následuje klíčová kap. 4, kde autorka popisuje metody experimentování. Výsledky experimentů jsou v kap. 5 podrobně analyzovány a je zhodnocen přínos jednotlivých modifikací či kombinací uvažovaných modelů.

Předložená práce neobsahuje žádné přílohy s programy či daty.

Drobné výtky mám ke stránce rešeršní. Diplomová práce obsahuje rozsáhlý seznam použitých literárních zdrojů. Lze říci, že autorka čerpala z nedávných či nejnovějších publikací relevantních pro daný obor, které ovšem pracují téměř výlučně s angličtinou. Je však poněkud překvapující, že nezmiňuje českou literaturu o vyhledávání v textech (např. a zejména [1]), už kvůli zavedené české terminologii. Rovněž co se týče pokusů využít závislostní syntax k obohacení popisu obsahu českých dokumentů, existují starší práce [2], [3], které autorka nezmiňuje.

Hlavní přínos práce vidím v tom, že autorka experimentálně srovnala a ověřila přínos jak jednotlivých pravděpodobnostních modelů pro vyhledávání, tak jejich kombinací. Poprvé tak zhodnotila užití závislostní syntaxe češtiny v rámci pravděpodobnostního modelu vyhledávání, v tomto smyslu se jedná o originální výsledek. Domnívám se, že tyto její experimentální výsledky budou pro vhodné zpracování publikovatelné i na relevantních mezinárodních odborných konferencích.

Jak formálně, tak věcně je předložená práce velmi pečlivě zpracována. Autorka se vyjadřuje jasně, přesně a bez gramatických chyb. Metodologicky postupuje správně a v souladu s dnes užívanými postupy pro vyhodnocování experimentálních výsledků. Hodnotím pozitivně, že práce není příliš dlouhá, není to totiž na újmu jejího obsahu.

Závěr

Předložená práce je zpracována adekvátně jejímu zadání a svou formou, rozsahem, náročností a kvalitou odpovídá požadavkům kladeným na diplomovou práci. Autorka jednoznačně dosáhla uložených cílů. Práci hodnotím velmi kladně a doporučuji ji k obhajobě. Doporučuji předloženou práci uznat jako diplomovou.

V Praze, 6. května 2009



RNDr. Martin Holub, Ph.D.

Reference

- [1] J. Pokorný, V. Snášel, D. Húsek: Dokumentografické informační systémy. Karolinum, Praha 1998.
- [2] M. Holub: *Context Information Mining I. Motivation and Syntactic Issues*. The Prague Bulletin of Mathematical Linguistics, 73—74:43--62, Charles University, Prague, 2000.
- [3] M. Holub: *Use of Dependency Microcontexts in Information Retrieval*. In V. Hlaváč, K. G. Jeffery, and J. Wiedermann, editors, SOFSEM'2000, Proceedings of the 27th Annual Conference on Current Trends in Theory and Practice of Informatics. Springer-Verlag, Berlin, 2000.