

Abstrakt

Název práce: *Využití syntaxe v metodách pro vyhledávání informací*

Autor: *Bc. Jana Kravalová*

Katedra (ústav): *Ústav formální a aplikované lingvistiky*

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

e-mail vedoucího: *pecina@ufal.mff.cuni.cz*

Abstrakt: V posledních letech výzkumu v oblasti vyhledávání informací je věnována značná pozornost metodám založeným na jazykovém modelování. I přesto, že tento přístup dovoluje použití libovolného jazykového modelu, většina publikovaných experimentů byla prováděna s klasickým n-gramovým modelem (mnohdy pouze s unigramovým modelem). Cílem diplomové práce je navrhnout, implementovat a vyhodnotit (na českých datech) metodu, která by pravděpodobnostní model obohatila o použití syntaktické informace získané automaticky (strojově) z dokumentů i dotazů. V předkládané práci se pokusíme vhodným způsobem zavést syntaktickou informaci do jazykových modelů a experimentálně srovnáme navržený přístup s výsledky unigramového a bigramového povrchového modelu. Kromě využití syntaktické informace se zaměříme také na vliv vyhlazování, stemmingu, lemmatizace, použití stopwords a metody rozšiřování dotazů – pseudo relevance feedback. Provedeme také detailní analýzu použitých systémů vyhledávání informace a podrobně popíšeme jejich vlastnosti. Experimenty budou prováděny na české testovací kolekci z Cross Language Evaluation Forum 2007 Ad-Hoc Track ([1]) a předkládané výsledky lze tedy srovnat s výsledky publikovanými v [19] a [4].

Klíčová slova: vyhledávání informací, jazykové modelování, závislostní syntax, vyhlazování