Abstract

Title: *Information Retrieval Using Syntax Information*

Author: *Bc. Jana Kravalová*

Department: *Institute of Formal and Applied Linguistics*

Supervisor: *Mgr. Pavel Pecina, Ph.D.*

Supervisor's e-mail address: *pecina@ufal.mff.cuni.cz*

Abstract: In the last years, application of language modeling in information retrieval has been studied quite extensively. Although language models of any type can be used with this approach, only traditional n-gram models based on surface word order have been employed and described in published experiments (often only unigram language models). The goal of this thesis is to design, implement, and evaluate (on Czech data) a method which would extend a language model with syntactic information, automatically obtained from documents and queries. We attempt to incorporate syntactic information into language models and experimentally compare this approach with unigram and bigram model based on surface word order. We also empirically compare methods for smoothing, stemming and lemmatization, effectiveness of using stopwords and pseudo relevance feedback. We perform a detailed analysis of these retrieval methods and describe their performance in detail.

Keywords: information retrieval, language modelling, depenency syntax, smoothing