Automatic text categorization is a challenging task that involves machine learning, computational linguistics, and formal models of content representation. This dissertation focuses on both theoretical and empirical perpectives on the relationship between language and structure, in the context of text categorization. The purpose of the theoretical investigation is to study how text categorization can be defined and characterized by means of mathematics and formal linguistics. The purpose of the empirical study is to explore the usefulness of statistical semantic information for automatic text categorization. The empirical investigation is carried out by studying the classification performance of different semantic kernels for a machine learning algorithm called the support vector machine (SVM).

Johan Eklund is employed by the Swedish School of Library and Information Science at the University of Borås. With or without context: Automatic text categorization using semantic kernels is his doctoral dissertation.

Digital version: http://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-8949
ISBN 978-91-981654-8-7 (printed)
ISBN 978-91-981654-9-4 (pdf)
ISSN 1103-6990 Skrifter från Valfrid, nr. 60

DOCTORAL THESIS

Johan Eklund WITH OR WITHOUT CONTEXT: AUTOMATIC TEXT CATEGORIZATION USING SEMANTIC KERNELS

DOCTORAL THESIS
Library and Information Science

WITH OR WITHOUT CONTEXT: AUTOMATIC TEXT CATEGORIZATION USING SEMANTIC KERNELS

Johan Eklund

THE SWEDISH SCHOOL OF LIBRARY AND INFORMATION SCIENCE
UNIVERSITY OF BORÅS