Carstens · Ontology Based Query Expansion

Schriften zur Informationswissenschaft
Band 60

Herausgegeben vom Hochschulverband für Informationswissenschaft (HI) e.V. Konstanz

Mitglieder des wissenschaftlichen Beirates für die Schriftenreihe:

Nicolas Belkin	Rutgers University
Hans Peter Frei	ETH Zürich
Rainer Hammwöhner	Universität Regensburg
Ilse M. Harms	Universität des Saarlandes
Norbert Henrichs	Universität Düsseldorf
Josef Herget	Donau-Universität Krems
Gerhard Knorz	Hochschule Darmstadt
Jürgen Krause	Universität Koblenz-Landau IZ Sozialwissenschaften
Rainer Kuhlen	Universität Konstanz
Klaus-Dieter Lehmann	Präsident der Stiftung Preußischer Kulturbesitz
Hans-Jürgen Manecke	TU Ilmenau
Achim Oßwald	Fachhochschule Köln
Wolf Rauch	Universität Graz
Harald Reiterer	Universität Konstanz
Marc Rittberger	Hochschule Darmstadt/DIPF
Christian Schlögl	Universität Graz
Wolfgang Semar	HTW Chur
Dagobert Soergel	University of Maryland
Wolfgang G. Stock	Universität Düsseldorf
Christian Wolff	Universität Regensburg
Christa Womser-Hacker	Universität Hildesheim
Harald Zimmermann	Universität des Saarlandes

Carola Carstens

Ontology Based Query Expansion

Retrieval Support for the Domain of Educational Research

vwh
Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

C. Carstens: Ontology Based Query Expansion

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter http: //d-nb.de abrufbar.

© Verlag Werner Hülsbusch, Boizenburg, 2012

vwh Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

www.vwh-verlag.de

Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch, Boizenburg.
Eine weitere Verwertung im Sinne des Urheberrechtsgesetzes ist nur mit Zustimmung des Autors möglich.

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen, Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und als solche den gesetzlichen Bestimmungen unterliegen.

Druck und Bindung: Kunsthaus Schwanheide

Printed in Germany

Zugleich: Diss., Univ. Hildesheim, 2011

– Als Manuskript gedruckt –

ISSN: 0938-8710
ISBN: 978-3-86488-011-7

Ontology Based Query Expansion – Retrieval Support for the Domain of Educational Research

Dissertation accepted by
Department III (Linguistics and Information Science)
of the University of Hildesheim

for obtaining the degree of
Dr. phil. in Information Science

by

Carola Carstens
born in Walsrode (03.04.1981)

Supervisors
Prof. Dr. Christa Womser-Hacker (University of Hildesheim)
Prof. Dr. Thomas Mandl (University of Hildesheim)
Prof. Dr. Marc Rittberger (German Institute for International Educational Research)

Date of disputation
31.08.2011

Disclaimer

The views expressed in this thesis are the sole responsibility of the author and in no way represent the view of the European Commission and its services.

Acknowledgements

This thesis was realised during my work as a research assistant at the Information Center for Education of the German Institute for International Educational Research (DIPF) in Frankfurt am Main, Germany. I am grateful to my colleagues at the Information Center for creating a very supportive atmosphere for my research.

Table of Contents

1	**Introduction**	**11**
2	**Ontologies**	**15**
2.1	Basic Concepts	15
2.1.1	Definition of Ontologies	15
2.1.2	Basic Elements of Ontologies	16
2.1.3	Ontologies and Other Knowledge Structures	17
2.1.4	Types of Ontologies	21
2.2	Ontologies in the Semantic Web	23
2.2.1	Semantic Web Technologies	23
2.2.2	Semantic Web Tools	28
2.2.3	Research Context-Related Data on the Semantic Web	29
3	**Retrieval**	**33**
3.1	Information Retrieval	33
3.1.1	Definition of Information Retrieval	33
3.1.2	Information Retrieval versus Data Retrieval	36
3.1.4	Traditional versus Interactive Information Retrieval Models	42
3.2	Evaluation of Information Retrieval Systems	45
3.2.1	Design of Retrieval Experiments	45
3.2.2	Automatic versus Interactive Retrieval Experiments	46
3.2.3	The TREC Initiative	50
3.2.4	Evaluation Measures	59
4	**Query Expansion**	**69**
4.1	Definition of Query Expansion	69
4.2	Query Expansion Methods	72
4.3	Effectiveness of Query Expansion Methods	74
4.3.1	Query Expansion Based on Search Results	75
4.3.2	Query Expansion Based on Collection Dependent Knowledge Structures	76
4.3.3	Query Expansion Based on Collection Independent Knowledge Structures	77
4.4	Knowledge Structures for Collection Independent Query Expansion	79
4.4.1	Query Expansion Based on Thesauri	79

4.4.2 Query Expansion Based on WordNet 83
4.4.3 Query Expansion Based on Ontologies 85
4.5 Evaluation Methods for Collection Independent Query Expansion 87

5 Search Behavior 97

5.1 Query Characteristics 98
5.2 Search Strategies 99
5.3 Search Tactics 101
5.4 Strategic Search Support 102
5.5 Studies on Search Behavior 104
5.6 Search Behavior in the German Education Index 107
5.6.1 Document Corpus and Indexing Vocabulary 108
5.6.2 Search Functionalities 109
5.6.3 Users 110
5.6.4 Findings 111

6 Research Interest and Experimental Framework 117

6.1 Summary of the State of the Art 117
6.2 Overview of Experimental Design 122
6.3 Research Questions 123
6.4 Experimental Variables 131
6.5 Ontological Knowledge Base 132
6.5.1 The RESCON Ontology Schema 133
6.5.2 Integration of Data from Original Data Sources 136
6.6 Prototypical Ontology Based Query Expansion System 141
6.7 Document Corpus 144

7 Automatic Query Expansion Experiments 145

7.1 Experiments for Concept Queries 146
7.1.1 Test Queries 147
7.1.2 Relevance Judgments 151
7.1.3 Evaluation Measures 154
7.2 Query Expansion Effects for Concept Queries 157
7.2.1 Semantic Relations' Expansion Effects on Baseline Queries 158
7.2.2 Semantic Relations' Average Query Expansion Effects 164
7.2.3 Significance of Differences between Semantic Relations' Query Expansion Effects 166

7.2.4 Semantic Relations' Evaluation Measure Value Distributions 172
7.3 Term Frequency-Dependent Analysis of the Relations' Query Expansion Effects for Concept Queries 182
7.3.1 Influence of Expansion Terms' Document Frequencies 183
7.3.2 Relation-Specific Populations of Frequency Categories 189
7.3.3 Frequency Category-Specific Expansion Effects of Semantic Relations 192
7.3.4 Semantic Relations' Suitability for the Automatic and Interactive Query Expansion Modes 207
7.4 Experiments for Organization and Project Queries 212
7.4.1 Test Queries 212
7.4.2 Relevance Judgments 213
7.4.3 Evaluation Measures 214
7.5 Query Expansion Effects for Organization and Project Queries 215
7.5.1 Expansion with Preferred Labels 216
7.5.2 Expansion with Alternative Labels 222
7.5.3 Disambiguated Expansion with Alternative Labels and Context Information 227
7.6 Conclusion for Automatic Query Expansion Experiments 233
7.6.1 Semantic Relations' Effects for the Expansion of Concept Queries 234
7.6.2 Semantic Relations' Effects for the Expansion of Project and Organization Queries 237
7.6.3 Methodological Discussion 238
8 Interactive Query Expansion Experiments 243
8.1 Experimental Design 245
8.2 Data Collection Methods 247
8.3 Test Persons 252
8.4 Topics 257
8.5 Relevance Judgments 259
8.6 Evaluation Measures 261
8.7 Significance Tests for System Comparisons 262
8.8 Retrieval System Versions 264
8.8.1 Baseline System 264
8.8.2 Automatic System 267
8.8.3 Interactive System 270
8.9 Experimental Results 279
8.9.1 Subjective Ratings of the Systems' Shared Functionalities 280
8.9.2 Subjective Ratings of the Query Expansion Functionalities 288

8.9.3 Recall 294
8.9.4 Number of Collected Documents 298
8.9.5 DistinctDocuments Measure 301
8.9.6 Precision 306
8.9.7 Pre-Click Confidence 309
8.9.8 Number of Query Reformulations 311
8.9.9 Fraction of Interactive Reformulation Steps 314
8.9.10 Effects of Interactive Reformulation Steps 316
8.9.11 Types of Query Reformulations 320
8.9.12 Query Complexity 327
8.10 Conclusion for Interactive Query Expansion Experiments 330
8.10.1 Potential of Ontology Based Query Expansion 330
8.10.2 Effectiveness of Ontology Based Query Expansion 331
8.10.3 Query Expansion Behavior 337
8.10.4 Methodological Discussion 341

9 Interpretation of Experimental Results 347

9.1 Novelty 347
9.2 Related Work 351
9.3 Further Research 357
9.4 Applicability of Experimental Results to the GEI Environment 364

10 Conclusion 369

References 371

List of Abbreviations 385

List of Tables 387

List of Figures 391

Appendices *) WWW

*) Der Anhang ist online abrufbar unter http://www.vwh-verlag.de/vwh/?p=719.