

FREIE UNIVERSITÄT BERLIN
FACHBEREICH WIRTSCHAFTSWISSENSCHAFT

Quality-Driven Information Filtering in the Context of Web-Based Information Systems

INAUGURAL-DISSERTATION

zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaft des Fachbereichs
Wirtschaftswissenschaft der Freien Universität Berlin

vorgelegt von
Diplom Kaufmann Christian Bizer
geboren in Konstanz
wohnhaft in Berlin

März 2007

Dekan

Univ.-Prof. Dr. Michael Kleinaltenkamp
Freie Universität Berlin
Fachbereich Wirtschaftswissenschaft

Erstgutachter

Univ.-Prof. Dr. Uwe H. Suhl
Freie Universität Berlin
Fachbereich Wirtschaftswissenschaft
Institut für Produktion, Wirtschaftsinformatik und Operations Research

Zweitgutachter

Univ.-Prof. Dr.-Ing. Robert Tolksdorf
Freie Universität Berlin
Fachbereich Mathematik und Informatik
Arbeitsgruppe Netzbasierte Informationssysteme

Tag der Disputation: 14. Februar 2007

Abstract

Web-based information systems provide access to information originating from multiple information providers inside the company, from partner organizations, and from the public Web. The quality of provided information may vary as information providers have different levels of knowledge, different views of the world, and different intentions. Thus before information is used to accomplish a specific task, its quality should be assessed according to task-specific criteria. The goal of the thesis is to develop and evaluate a quality-driven information filtering framework which supports information consumers in their decision whether to accept or reject information. The main objective of the framework is to allow information consumers to apply a wide range of different subjective filtering policies. In order to facilitate the information consumers' understanding of filtering decisions, the framework can generate explanations why information satisfies a specific policy. The framework is integrated into a web browser and is applied within a financial information integration scenario.

Contents

Abstract	ii
1 Introduction	1
1.1 Problem Definition	3
1.2 Thesis Outline	4
1.3 Research Method	7
1.4 Published Work	8
I Information Quality and the Web	10
2 Information Quality	11
2.1 Related Work	11
2.2 Terminology	12
2.3 Information Quality Dimensions	14
2.3.1 Intrinsic Dimensions	17
2.3.2 Contextual Dimensions	18
2.3.3 Representational Dimensions	20
2.3.4 Accessibility Dimensions	21
2.4 Summary	21
3 Information Quality Assessment	23
3.1 Assessment Metrics	23
3.1.1 Content-Based Metrics	25
3.1.2 Context-Based Metrics	27
3.1.3 Rating-Based Metrics	31
3.2 Accuracy of Assessment Results	36
3.3 Information Filtering Policies	38
3.3.1 Policy Selection	39
3.3.2 Example Policies	40
3.4 Summary	42

II	Representation of Meta-Information	43
4	The Resource Description Framework	46
4.1	The RDF Data Model	47
4.1.1	Triples and Graphs	48
4.1.2	RDF Nodes	48
4.1.3	Example RDF Graph	51
4.2	RDF Schema	52
4.3	Syntaxes for RDF	54
4.3.1	The RDF/XML Syntax	54
4.3.2	The Turtle Syntax	57
4.4	Query Languages for RDF	57
4.5	RDF Reification	58
5	Named Graphs	62
5.1	The Named Graphs Data Model	63
5.2	Related Work	64
5.2.1	Quads	64
5.2.2	N3 Formula	65
5.2.3	RDF Dataset	66
5.3	Syntaxes for Named Graphs	67
5.3.1	The TriX Syntax	67
5.3.2	The TriG Syntax	68
5.4	Query Languages for Named Graphs	69
6	The Semantic Web Publishing Vocabulary	72
6.1	Authorizing Named Graphs	73
6.2	Signing Named Graphs	75
6.3	Related Work	80
6.3.1	Dublin Core Element Set	80
6.3.2	XML-Signature Syntax	81
7	Use Case: Financial Information Integration	83
7.1	Domain Model	83
7.2	Example Data Set	86
8	Summary	89

III	The WIQA Framework	91
9	Expressing Information Filtering Policies	95
9.1	Basic Grammar	95
9.2	Graph Pattern Matching	97
9.3	Accepting Triples	100
9.4	Context Variables	102
9.5	Filters	104
9.6	Functions	106
9.6.1	More Positive Ratings Function	109
9.6.2	Tidal Trust Function	110
9.6.3	Count Function	113
9.7	Summary	117
10	Explaining Assessment Results	119
10.1	Explaining Pattern Matches	121
10.1.1	Building the Graph Pattern Tree	122
10.1.2	Instantiating the Graph Pattern Tree	124
10.2	Explaining Extension Function Results	125
10.2.1	More Positive Ratings Function	127
10.2.2	Tidal Trust Function	129
10.3	RDF Explanations	129
11	Implementation	133
11.1	The NG4J - Named Graph API for Jena	133
11.1.1	Public Interface	134
11.1.2	Usage Example	137
11.2	The WIQA - Filtering and Explanation Engine	137
11.2.1	Public Interface	137
11.2.2	Usage Example	140
12	The WIQA Browser	143
12.1	Collecting Information	144
12.2	Importing and Exporting Information	145
12.3	Browsing Information	146
12.4	Applying Policies and Retrieving Explanations	147
13	Related Work	153

<i>CONTENTS</i>	vi
IV Conclusion	155
14 Conclusion	156
14.1 Contributions	157
14.2 Future Directions	159
Appendices	160
A List of Abbreviations	161
B Grammar of the TriG Syntax	163
C Grammar of the WIQA-PL Policy Language	165
D German Summary	168
Bibliography	170