

Table of Content

Table of Content XIII

List of FiguresXIX

List of TablesXXIII

List of Abbreviations XXV

PART I – Introduction, Economic Relevance, and Research Design1

1 Introduction1

1.1 Initial Problem Statement 1

1.2 Economic Relevance..... 3

1.3 Organization of this Thesis..... 6

1.4 Published Work..... 6

1.4.1 Book Chapters 7

1.4.2 Papers in Conference Proceedings..... 7

1.4.3 Other Publications..... 7

2 Research Design8

2.1 Semantic Technologies and Ontologies..... 8

2.2 Research Goal 9

2.3 Research Questions..... 11

2.4 Research Methodology 12

2.4.1 Design Science Research Methodology..... 13

2.4.2 Ontology Development Methodology 18

PART II – Foundations: Data Quality, Semantic Technologies, and the Semantic Web.....	20
3 Data Quality	20
3.1 <i>Data Quality Dimensions</i>	21
3.2 <i>Quality Influencing Artifacts</i>	24
3.3 <i>Data Quality Problem Types</i>	26
3.3.1 Quality Problems of Attribute Values	28
3.3.2 Multi-Attribute Quality Problems	30
3.3.3 Problems of Object Instances	32
3.3.4 Quality Problems of Data Models	34
3.3.5 Common Linguistic Problems	38
3.4 <i>Data Quality in the Data Lifecycle</i>	39
3.4.1 Data Acquisition Phase.....	40
3.4.2 Data Usage Phase.....	41
3.4.3 Data Retirement Phase	42
3.4.4 Data Quality Management throughout the Data Lifecycle.....	42
3.5 <i>Data Quality Management Activities</i>	43
3.5.1 Total Information Quality Management (TIQM).....	43
3.5.2 Total Data Quality Management (TDQM)	47
3.5.3 Comparison of Methodologies	49
3.6 <i>Role of Data Requirements in DQM</i>	49
3.6.1 Generic Data Requirement Types	50
3.6.2 Challenges Related to Requirements Satisfaction.....	54
4 Semantic Technologies	56
4.1 <i>Characteristics of an Ontology</i>	56
4.2 <i>Knowledge Representation in the Semantic Web</i>	58
4.2.1 Resources and Uniform Resource Identifiers (URIs)	58
4.2.2 Core RDF Syntax: Triples, Literal Triples, and RDF Links.....	59
4.2.3 Constructing an Ontology with RDF, RDFS, and OWL.....	60
4.2.4 Language Profiles of OWL and OWL 2.....	63
4.3 <i>SPARQL Query Language for RDF</i>	64

4.4	<i>Reasoning and Inferencing</i>	65
4.5	<i>Ontologies and Relational Databases</i>	67
5	Data Quality in the Semantic Web	69
5.1	<i>Data Sources of the Semantic Web</i>	69
5.2	<i>Semantic Web-specific Quality Problems</i>	71
5.2.1	Document Content Problems.....	72
5.2.2	Data Format Problems.....	72
5.2.3	Problems of Data Definitions and Semantics.....	73
5.2.4	Problems of Data Classification	74
5.2.5	Problems of Hyperlinks	75
5.3	<i>Distinct Characteristics of Data Quality in the Semantic Web</i>	76
PART III – Development and Evaluation of the Semantic Data Quality Management Framework		78
6	Specification of Initial Requirements	78
6.1	<i>Motivating Scenario</i>	78
6.2	<i>Initial Requirements for SDQM</i>	79
6.2.1	Task Requirements.....	80
6.2.2	Functional Requirements	82
6.2.3	Conditional Requirements	83
6.2.4	Research Requirements	85
6.3	<i>Summary of SDQM's Requirements</i>	86
7	Architecture of the Semantic Data Quality Management Framework (SDQM)	87
7.1	<i>Data Acquisition Layer</i>	88
7.1.1	Reusable Artifacts for the Data Acquisition Layer.....	89
7.1.2	Data Acquisition for SDQM	90
7.2	<i>Data Storage Layer</i>	91
7.2.1	Reusable Artifacts for Data Storage in SDQM	91
7.2.2	The Data Storage Layer of SDQM.....	92

7.3	<i>Data Quality Management Vocabulary</i>	94
7.3.1	Reuse of Existing Ontologies.....	95
7.3.2	Technical Design of the DQM Vocabulary	96
7.4	<i>Data Requirements Editor</i>	99
7.4.1	Reusable Artifacts for SDQM's Data Requirements Editor	100
7.4.2	Data Requirements Wiki	101
7.5	<i>Reporting Layer</i>	104
7.5.1	Reusable Artifacts for SDQM's Reporting Layer	105
7.5.2	Semantic Data Quality Manager	105
8	Application Procedure of SDQM	110
8.1	<i>Prerequisites</i>	110
8.2	<i>The Data Quality Management Process with SDQM</i>	111
9	Evaluation of the Semantic Data Quality Management Framework (SDQM)	122
9.1	<i>Evaluation of Algorithms</i>	122
9.1.1	Algorithm Evaluation Methodology	122
9.1.2	Application Procedure.....	123
9.1.3	Results.....	124
9.2	<i>Use Case 1: Evaluation of Material Master Data</i>	124
9.2.1	Scenario	125
9.2.2	Setup and Application Procedure of SDQM.....	125
9.2.3	Results and Findings	127
9.3	<i>Use Case 2: Evaluation of Data from DBpedia</i>	132
9.3.1	Scenario	132
9.3.2	Specialties of Semantic Web Scenarios	133
9.3.3	Setup and Application Procedure	133
9.3.4	Results and Findings	135
9.4	<i>Use Case 3: Consistency Checks Among Data Requirements</i>	141
9.4.1	Scenario	142
9.4.2	Application Procedure.....	142
9.4.3	Summary	144

9.5	<i>Comparison with Talend OS for Data Quality</i>	145
9.5.1	Representation and Management of Data Requirements	145
9.5.2	Data Quality Monitoring and Assessment Reporting	148
9.5.3	Summary	151
PART IV – Related Work		153
10 Related Work		153
10.1	<i>High-Level Classification Schema</i>	153
10.2	<i>Categorization Schema</i>	154
10.2.1	Supported Data Lifecycle Step	154
10.2.2	Supported Data Representation	155
10.2.3	Supported Data Quality Task	156
10.3	<i>Conventional Rule-Based Approaches</i>	157
10.4	<i>Ontology-based Approaches</i>	158
10.4.1	Information System-oriented Approaches	158
10.4.2	Web-oriented Approaches	165
10.5	<i>Summary</i>	168
PART V - Conclusion		171
11 Synopsis and Future Work		171
11.1	<i>Research Summary</i>	171
11.2	<i>Contributions</i>	173
11.3	<i>Conclusion and Future Work</i>	174
Appendix A – Comparison of TIQM and TDQM		177
Appendix B –Rules for the Evaluation of SDQM		182
Appendix C – Test Data for SDQM's Evaluation		187
Appendix D – Evaluation Results of SDQM's Data Quality Monitoring Queries		191

Appendix E – Evaluation Results of SDQM’s Data Quality Assessment	
Queries	193
References	195

Data Quality Management with Semantic Technologies

Fürber, C.

2016, XXVII, 205 p. 63 illus., Softcover

ISBN: 978-3-658-12224-9