

Dissertation
zur Erlangung des akademischen Grades
Doktorin der Naturwissenschaften
(Doctor rer. nat)

**A Contextual Approach to Ontology Reuse:
Methodology, Methods and Tools for the
Semantic Web**

eingereicht am
Fachbereich Mathematik und Informatik
der
Freien Universität Berlin
von
Dipl. Inform. Elena Pâslaru-Bontaş

Dekan: Prof. Dr.-Ing. Jochen Schiller

- 1. Gutachter:** Prof. Dr. Robert Tolksdorf
- 2. Gutachter:** Prof. Dr. Robert Meersman
- 3. Gutachter:** Prof. Dr. Dieter Fensel

Tag der Disputation: 28 Februar 2007

Abstract

The realization of the envisioned Semantic Web is conditioned by a real operationalized sharing and reuse of formally represented knowledge models, so-called ontologies. However, because of the open, distributed nature of Web environment and the inherent limitations of achieving highly reusable, commonly agreed domain conceptualizations, the current state of the art in the ontology engineering field states the need for additional techniques to aid ontology designers and users in evaluating and manipulating existing ontological sources for reuse purposes.

The first goal of this work is to identify the major factors which influence the feasibility of current ontology reuse processes. We perform this risk analysis by means of a comprehensive literature survey complemented by two case studies situated in typical application scenarios in the Semantic Web field. The results of this analysis are aligned to specific contextual dimensions: features of the ontologies to be reused and of the corresponding ontology application scenarios, as well as parameters of the reuse engineering process. Accounting for the conclusions of the feasibility study, we advocate that the success of reuse endeavors is fundamentally influenced by factors situated at these dimensions and design an *context-sensitive ontology reuse methodology* implementing this finding.

Our second goal is to provide methods and tools assisting ontology engineers and domain experts in operating reuse. On the basis of the requirements derived from the case studies and from the proposed methodology we develop a *metadata model* for the description of Semantic Web ontologies. This model represents reuse-relevant characteristics of ontologies in a semantically precise, machine-processable, interoperable and extendable manner. We elaborate on the *context-sensitive usage of metadata* information for human-driven *ontology evaluation* and exemplify it on a further sub-task of the reuse process, the *matching* between heterogeneous ontologies—as often encountered on the current Semantic Web. We also address the issue of *metadata management*: we propose a suite of heuristics to automatically derive and detect metadata information related to ontologies, thus minimizing the need for a manual annotation of the resources. The methods are prototypically implemented in a dedicated ontology reuse platform.

In order to validate our research from different perspectives, we apply multiple evaluation approaches. The case study methodology provides us the theoretical foundations for comparing the operation of reuse processes in real-world situations with and without our context-oriented methodological approach, and for demonstrating its practicability. Professional reviews estimate the technical quality, originality and impact of the proposed solution. Using the goal-free evaluation methodology we are able to situate the results of this thesis in the actual ontology reuse field, emphasizing the scenarios in which they can be optimally applied. The implementation is evaluated against method-specific quality criteria on representative test sets.

The experimental results demonstrate that for a representative class of Semantic Web applications—information retrieval and semantic annotation—our approach improves ontology reuse with respect to the invested efforts, the user-perceived process operation efficiency and the fitness of use of the reuse outcomes.

Zusammenfassung

Die Semantic Web Initiative verfolgt die Idee, Web Ressourcen unter Verwendung von Methoden der Wissensrepräsentation zu erweitern, damit sie maschinell leichter gefunden, besser genutzt und individuell zusammengestellt werden können. Die Realisierung des Semantic Web setzt voraus, dass Wissen einerseits explizit und maschinenlesbar dargestellt, andererseits nutzungsgerecht wiederverwendet und ausgetauscht werden kann. Ontologien, als *“explizite und formale Spezifikation einer gemeinsamen Konzeptualisierung”*, spielen dabei eine zentrale Rolle. Obwohl Ontologien mittlerweile in vielen Anwendungen aus Biowissenschaften, Wissensmanagement oder eCommerce verwendet werden, wird ihre Wiederverwendung gerade in offenen, verteilten Umgebungen wie das Web von einer Reihe von Faktoren entscheidend beeinflusst. Neben der allgemeinen Fragestellung der Realisierung einer allgemein gültigen und beliebig einsetzbaren Domänenkonzeptualisierung, wird der Prozess der Wiederverwendung durch den Mangel an ausgereiften Verfahren und Werkzeugen für die Evaluierung und Nutzung existierender ontologischer Quellen vor zusätzlichen Herausforderungen gestellt.

Ausgangspunkt der vorliegenden Arbeit ist die Identifizierung und Analyse der Hauptfaktoren, welche aktuell die Durchführbarkeit von Ontologiewiederverwendungsprozessen beeinflussen. Zu diesem Zweck werden existierende Ansätze in diesem Forschungsbereich untersucht und in Hinsicht auf angewendete Verfahren und Methoden, sowie erzielte Ergebnisse ausgewertet. Diese Literaturrecherche wird durch eigene umfassende Fallstudien vervollständigt. Die Machbarkeitsstudie deutet darauf hin, dass Kontext-Aspekte einen entscheidenden Beitrag bei der effizienten Wiederverwendung von Ontologien leisten. Das erste Ziel dieser Arbeit ist ein Vorgehensmodell für die Durchführung von Ontologiewiederverwendungsprozessen zu entwerfen, das auf den Ergebnissen dieser Studie aufbaut und explizit den kontext-sensitiven Charakter solcher Prozesse nutzt um ihre Effizienz zu steigern.

Die zweite Zielstellung der Arbeit ist Methoden und Verfahren bereitzustellen, welche den Ontologieentwickler und -nutzer darin unterstützen, Ontologien wiederzuverwenden und die systematische Anwendung unseres Vorgehensmodells erleichtern. Als Grundlage für die Entwicklung dieser Verfahren wird ein Metadaten-Schema für die Beschreibung von Ontologien vorgestellt. Dieses Modell erfasst wiederverwendungsrelevante Eigenschaften von Ontologien auf semantisch präzise, maschinenlesbare, und erweiterbare Art und Weise. Die kontext-sensitive Nutzung von Metadaten-Informationen wird in Zusammenhang mit zwei Kernschritten eines Ontologiewiederverwendungsprozesses, der Evaluierung von Ontologien und ihrer Zusammenführung zu einer Zielontologie, ausführlich behandelt. Weiterhin wird auf die Problematik der Metadaten-Erstellung und -verwaltung eingegangen. Es wird eine Reihe von Heuristiken eingeführt, welche eine automatische Generierung und Ableitung von ontologiebezogene Metadaten ermöglichen und damit den manuellen Aufwand für die Bereit-

stellung dieser Informationen reduzieren. Die Verfahren sind prototypisch umgesetzt.

Die Arbeit wird multiperspektivisch evaluiert. Das Fallstudien- Evaluierungsverfahren bildet die theoretische Grundlage um die Durchführung von Ontologiewiederverwendungsprozessen durch Anwendung unseres Lösungsvorschlags mit bestehenden Ansätzen zu vergleichen. Gleichzeitig zeigen die Fallbeispiele den praktischen Nutzen der Methodologie. Das expertenbasierte Evaluierungsverfahren bewertet die technische Qualität, Originalität und die Implikationen der vorgeschlagenen Lösung für das Forschungsgebiet. Die merkmalsbasierte Evaluierungsmethode ermöglicht es die Ergebnisse der Arbeit in das Forschungsumfeld der Ontologiewiederverwendung einzuordnen, und hebt hervor für welche Anwendungsgebiete die vorgeschlagene Arbeit am besten geeignet ist. Die prototypische Implementierung wird auf einem nutzungsrelevanten Datensatz anhand methodenspezifischer Qualitätsparameter getestet.

Die Evaluierungsergebnisse zeigen, dass der in dieser Arbeit verfolgten Ansatz Ontologiewiederverwendungsprozesse, wie sie im Kontext des Semantic Web auftreten, deutlich vereinfacht und verbessert.

Acknowledgements

First of all I would like to thank my advisor Prof. Dr. Robert Tolksdorf from the Free University of Berlin for giving me the opportunity to be part of a friendly working group and to work in an interesting and innovative field. He played an essential role in the accomplishment of my doctoral research with his advice and guidance during the last four years. I am also much obliged to Prof. Dr. Robert Meersman from the Free University of Brussels, who agreed to act as a referee of this thesis, and greatly contributed to its realization with his helpful comments and ideas.

Moreover I thank my colleagues from the Free University of Berlin, the Technical University of Munich and from many other institutions. They all had an important contribution to the improvement of this work with their valuable ideas and their research. My special thanks go to Malgorzata Mochol for our excellent collaboration and to Christoph Tempich for his support and helpful comments regarding this work and its further implications. I also have to thankfully mention my students, Yao Chen, Marko Harasic, Franziska Liebsch, Sebastian Schaepe and Sebastian Tietz for assisting me at the implementation of the thesis.

Furthermore I thank my family and friends for their patience and support during this challenging time. The completion of this thesis would be inconceivable without their help and understanding in the last years.

Acknowledgements

Contents

Abstract	iii
Zusammenfassung	v
Acknowledgements	vii
Acronyms	xvii
1 Introduction	1
1.1 Problem Description	1
1.2 Our Approach	4
1.2.1 Objectives	4
1.2.2 Research Methodology	4
1.2.3 Main Contributions	6
1.3 Terminology	6
1.3.1 General Engineering Process	7
1.3.2 Ontologies	7
1.3.3 Ontology Engineering	7
1.3.4 Ontology Reuse	8
1.3.5 Context and Metadata	9
1.4 Review of the Related Research Areas	9
1.4.1 Knowledge Reuse	9
1.4.2 Metadata	10
1.4.3 Context-sensitive Computing	11
1.5 Organization of the Thesis	12
2 Ontologies and Ontology Engineering	17
2.1 Ontologies	17
2.1.1 Definitions	18
2.1.2 Types of ontologies	19
2.2 Ontology Engineering Methodologies and Methods	23
2.2.1 Methodologies and Methods for Ontology Development	24
2.2.2 Methodologies and Methods for Ontology Use and Reuse	26
2.3 Application Scenarios for Ontologies	31
2.3.1 Information Retrieval	32
2.3.2 Knowledge Management	33
2.3.3 eCommerce	33
2.3.4 Semantic Web	34

2.4	Summary	40
3	Ontology Reuse Feasibility Study	41
3.1	Knowledge Reusability on the Semantic Web	41
3.2	Case Studies in Ontology Reuse	43
3.2.1	Gómez-Pérez and Rojas-Amaya’s Case Study	43
3.2.2	Uschold and Healy’s Case Study	44
3.2.3	Russ et al’ Case Study	45
3.2.4	Capellades’ Case Study	46
3.2.5	Arpírez at al’ Case Study	46
3.2.6	Laresgoiti et al’ Case Study	47
3.2.7	Bernaras et al’ Case Study	47
3.3	Case Study eHealth	48
3.3.1	A Semantic Web for Pathology	48
3.3.2	Reusing Medical Ontologies	50
3.4	Case Study eRecruitment	55
3.4.1	Knowledge Nets	55
3.4.2	Reusing Human Resources Ontologies	58
3.5	Requirements for Ontology Reuse	63
3.5.1	Requirements for Ontology Reuse Methodologies	64
3.5.2	Requirements for Ontology Reuse Support Methods and Tools	67
3.6	Summary	70
4	Ontology Reuse Methodology	71
4.1	Overview	71
4.1.1	General Process	71
4.1.2	Ontology Reuse Context	74
4.2	Ontology Discovery	79
4.3	Ontology Evaluation	80
4.3.1	Content Evaluation	83
4.3.2	Knowledge Representation Evaluation	84
4.3.3	Technical Evaluation	85
4.3.4	Application Evaluation	86
4.3.5	Availability Evaluation	88
4.4	Ontology Merging and Integration	88
4.4.1	Ontology Customization	90
4.4.2	Tool and Method Selection	91
4.4.3	Integration Preparation	92
4.4.4	Integration Execution	93
4.4.5	Integration Evaluation	93
4.5	Requirements for Methodology-Related Support Methods and Tools	94
4.6	Summary	96
5	Ontology Reuse Metadata Model	97
5.1	Model Engineering	97

5.1.1	Requirements Analysis	98
5.1.2	Conceptualization	101
5.1.3	Implementation	102
5.1.4	Evaluation	102
5.1.5	Evolution	103
5.2	Metadata Ontology	103
5.2.1	Class SemanticWebResource	106
5.2.2	Class Ontology	108
5.2.3	Class OntologyType	110
5.2.4	Class FormalityLevel	111
5.2.5	Class RepresentationLanguage	111
5.2.6	Class RepresentationParadigm	112
5.2.7	Class dmoz:Topic	113
5.2.8	Class wikipedia:Application	113
5.2.9	Class naics:IndustrialSector	116
5.2.10	Class ApplicationScenario	117
5.2.11	Class OntologyTask	118
5.2.12	Class MethodsMethodologiesAndTools	122
5.3	Methods for Metadata Creation and Management	122
5.3.1	Metadata Creation	124
5.3.2	Metadata Management	129
5.4	Summary	133
6	Methods Supporting Ontology Reuse	135
6.1	Ontology Evaluation Method	135
6.1.1	Method Overview	136
6.1.2	Context Specification	138
6.1.3	Features Selection	147
6.1.4	Usability Computation	147
6.1.5	Result Selection	159
6.2	Ontology Merging and Integration Method	159
6.2.1	Method Overview	160
6.2.2	Context Specification	161
6.2.3	Usability Computation	168
6.2.4	Result Selection	168
6.3	Summary	169
7	Tools Supporting Ontology Reuse	171
7.1	PROMI	171
7.1.1	Design and Architecture	172
7.1.2	User Guide	179
7.1.3	Technical Details	186
7.2	OntoMeta	195
7.2.1	Design and Architecture	195
7.2.2	Technical Details	198

7.3	Summary	204
8	Evaluation	205
8.1	Evaluation Methodology	205
8.1.1	Evaluation as a Field	205
8.1.2	Selection of Appropriate Evaluation Approaches	210
8.2	Professional Reviews	213
8.2.1	Ontology Reuse Methodology and Associated Methods	213
8.2.2	Ontology Reuse Metadata Model	214
8.3	Goal-Based Evaluation	216
8.3.1	Case Study eRecruitment	216
8.3.2	OntoMeta User Study	222
8.3.3	PROMI User Study	225
8.4	Goal-free Evaluation	228
8.4.1	Ontology Reuse Methodology	228
8.4.2	Ontology Reuse Metadata Model	229
8.5	Summary	231
9	Conclusions and Perspectives	233
9.1	Summary and Future Work	233
9.2	Outlook	235
	Appendices	239
A	Ontology Metadata	239
B	Ontology Merging and Integration Metadata	241
B.1	Description of the Composite Matching Service LSD	241
B.2	Description of the Hybrid Matching Service Cupid	242
C	Contextual Dependencies for Matcher Selections	245
C.1	Rule for Semi-formal Ontologies	245
C.2	Rule for Ontologies Using Different Natural Languages	246
C.3	Rule for Formal Axiomatized Ontologies without Instances	248
C.4	Rule for Formal Ontologies without Instances	251
	Bibliography	253

List of Figures

1.1	Research Narratives of Our Work and Its Results (cf. [217])	5
1.2	Organization of the Thesis	15
2.1	Classification of Ontologies as regards Their Reusability [111]	23
2.2	Ontology Engineering Process	25
2.3	High-Level Architecture of the Semantic Web [11]	35
2.4	Graph Notation for RDF Annotations	37
4.1	Ontology Reuse in Relation to Ontology Engineering	73
4.2	Ontology Reuse Process	74
5.1	Documentation Pattern for Ontology Classes	106
5.2	Overview of the Class <code>SemanticWebResource</code>	107
5.3	Overview of the Class <code>Ontology</code>	109
5.4	Overview of the Class <code>RepresentationLanguage</code>	112
5.5	Overview of the Top Categories of DMOZ Taxonomy	114
5.6	Overview of the Class <code>ApplicationScenario</code>	117
5.7	Overview of the Class <code>OntologyTask</code>	119
5.8	Overview of the Class <code>MethodsMethodologiesAndTools</code>	123
5.9	Core Metadata Services	124
5.10	Metadata Creation Support in Oyster	130
5.11	Metadata Creation Support in Onthology	131
6.1	Ontology Evaluation Method	137
6.2	Excerpt of the Context Ontology (Classes, Sub-Classes and Instances)	142
6.3	Ontology Merging and Integration Method	162
6.4	Taxonomy of Schema Matching Approaches cf. [184]	163
6.5	Classification of Individual Matchers with Some Instances	165
6.6	Classification of Combined Matchers with Some Instances	166
6.7	Implementation of Contextual Dependencies in Protégé	167
7.1	PROMI Layered Architecture	172
7.2	PROMI Ontology Reuse Process	175
7.3	PROMI Process Support: Working Progress Window (left) and Help Window (right)	176
7.4	PROMI Context Model	178
7.5	PROMI Evaluation Support: Context Specification	180

7.6	PROMI Evaluation Support: Context-sensitive Method Presentation for Domain Experts (top) and Ontology Engineers (bottom)	181
7.7	PROMI Evaluation Support: Feature Selection	182
7.8	PROMI Evaluation Support: Usability Computation and Result Selection	183
7.9	PROMI Translation Support	184
7.10	PROMI Merging Support: Context-based Selection of the Matching Algorithms	185
7.11	PROMI Merging Support: Concept Similarity Computation	186
7.12	PROMI Merging Support: Incremental Merging Process	187
7.13	PROMI Merging Support: Distribution of the Similarity Results	188
7.14	Overview of the PROMI packages	190
7.15	High-Level Architecture of the Evaluation Component	191
7.16	High-Level Architecture of the Merging and Integration Component	193
7.17	High-level Architecture of OntoMeta	196
7.18	OntoMeta Start Page	196
7.19	OntoMeta Result Page	197
7.20	Overview of the OntoMeta Classes	200
C.1	Results of Applying the Rule: “Use only linguistic matchers for semi-formal ontologies”	246
C.2	Results of Applying the Rule: “Use structure-based matchers for ontologies with different natural languages”	249
C.3	Results of Applying the Rule: “Use constraints-based matchers only for formal ontologies without instances and only if ontologies contain axioms”	252

List of Tables

3.1	Guidelines for Reusing Medical Ontologies	56
3.2	Guidelines for Building HR Ontologies by Reuse	62
3.3	Criteria for Analyzing and Comparing Ontology Engineering Methodologies cf. [63]	64
3.4	General Requirements for Ontology Reuse Methodologies	64
3.5	Reuse Process as Performed in the Analyzed Case Studies	65
3.6	Requirements for Ontology Discovery Support Methods and Tools	67
3.7	Requirements for Ontology Evaluation Support Methods and Tools	68
3.8	Requirements for Ontology Customization, Merging and Integration Support Methods and Tools	70
4.1	Ontology Roles in Relationship to Application Scenarios	76
4.2	Key Dimensions of the Context-sensitive Ontology Reuse Methodology	78
4.3	Requirements for Ontology Discovery Support Methods and Tools	95
4.4	Requirements for Ontology Evaluation Support Methods and Tools	95
4.5	Requirements for Ontology Customization, Merging and Integration Support Methods and Tools	96
5.1	Alignment of Ontology Roles to Ontology Tasks	121
6.1	Alignment of Ontology Features to Application Scenarios	146
6.2	Ontology Features Relevant for the Ontology Evaluation Step	148
6.3	Similarities between Types by Structure	152
6.4	Similarities between Types by Generality	152
6.5	Similarities between Ontology Tasks	154
6.6	Similarities between Application Systems	156
6.7	Similarities between Semantic Web Resources as regards Engineering Methods	157
8.1	Design Research Outputs and Their Evaluation	211
8.2	Evaluation Framework of Our Research	212
8.3	Evaluation Overview OntoMeta	223
8.4	Evaluation Results for Domain and View upon the Modelled Domain	224
8.5	Results of the Expert-based Evaluation on the Basis of the Ontological Content	227
8.6	Results of the Expert-based Evaluation on the Basis of the Ontology Metadata	227
8.7	Results of the Tool-based Evaluation	227
8.8	Results of the Goal-free Evaluation of the Methodology	230

List of Tables

Acronyms

AI	Artificial Intelligence
BKZ	Berufskennziffer
DC	Dublin Core
DEMO	Design Environment for Metadata for Ontologies
DMOZ	Directory at Mozilla
GO	GeneOntology
HR	Human Resources
IR	Information Retrieval
ISIC	International Standard of Industrial Classifications
IT	Information Technology
KM	Knowledge Management
KR	Knowledge Representation
MeSH	Medical Subject Headings
NAICS	North American Industry Classification System
NCI	National Cancer Institute Ontology
NLP	Natural Language Processing
NOC	National Occupational Classification
ODP	Open Directory Project
OE	Ontology/Ontological Engineering
OMV	Ontology Metadata Vocabulary
O*NET	Occupational Information Network
OWL	Web Ontology Language
OWL DL	OWL Description Logics

PROMI Platform for the **R**euse of **O**ntology through **M**erging and **I**ntegration

RDF Resource Description Framework

RDFS Resource Description Framework Schema

SNOMED Systematic Nomenclator of Medicine

SOC Standard Occupational Classification

SWRL Semantic Web Rule Language

UNSPSC United Nations Standard Products and Services Codes

UMLS Unified Medical Language System

WZ2003 Wirtschaftszweige 2003