

Ontology Alignment

Bridging the Semantic Gap

SEMANTIC WEB AND BEYOND

Computing for Human Experience

Series Editors:

Ramesh Jain
University of California, Irvine
<http://ngs.ics.uci.edu/>

Amit Sheth
University of Georgia
<http://lsdis.cs.uga.edu/~amit>

As computing becomes ubiquitous and pervasive, computing is increasingly becoming an extension of human, modifying or enhancing human experience. Today's car reacts to human perception of danger with a series of computers participating in how to handle the vehicle for human command and environmental conditions. Proliferating sensors help with observations, decision making as well as sensory modifications. The emergent semantic web will lead to machine understanding of data and help exploit heterogeneous, multi-source digital media. Emerging applications in situation monitoring and entertainment applications are resulting in development of experiential environments.

SEMANTIC WEB AND BEYOND

Computing for Human Experience

addresses the following goals:

- brings together forward looking research and technology that will shape our world more intimately than ever before as computing becomes an extension of human experience;
- covers all aspects of computing that is very closely tied to human perception, understanding and experience;
- brings together computing that deal with semantics, perception and experience;
- serves as the platform for exchange of both practical technologies and far reaching research.

Additional information about this series can be obtained from

<http://www.springer.com>

Additional Titles in the Series:

Semantic Web Services: Processes and Applications edited by Jorge Cardoso, Amit P. Sheth, ISBN 0-387-30239-5

Canadian Semantic Web edited by Mamadou T. Koné., Daniel Lemire; ISBN 0-387-29815-0

Semantic Management of Middleware by Daniel Oberle; ISBN-10: 0-387-27630-0

Ontology Alignment

Bridging the Semantic Gap

by

Marc Ehrig
University of Karlsruhe
Germany

 Springer

Marc Ehrig
University of Karlsruhe
76139 Karlsruhe
Germany

Library of Congress Control Number: 2006928852

Ontology Alignment: Bridging the Semantic Gap
by Marc Ehrig

ISBN-10: 0-387- 32805-X
ISBN-13: 978-0-387-32805-8
e-ISBN-10: 0-387-36501-X
e-ISBN-13: 978-0-387-36501-5

Printed on acid-free paper.

Dissertation, genehmigt von der Fakultät für Wirtschaftswissenschaften der Universität
Fridericiana zu Karlsruhe, 2005. Referent: Prof. Dr. Rudi Studer,
Korreferenten: Prof. Dr. Kuno Egle, Dr. (habil) Jérôme Euzenat

© 2007 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or
in part without the written permission of the publisher (Springer
Science+Business Media, LLC, 233 Spring Street, New York, NY 10013,
USA), except for brief excerpts in connection with reviews or scholarly
analysis. Use in connection with any form of information storage and
retrieval, electronic adaptation, computer software, or by similar or
dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks and
similar terms, even if they are not identified as such, is not to be taken as
an expression of opinion as to whether or not they are subject to
proprietary rights.

9 8 7 6 5 4 3 2 1

springer.com

Contents

Preface	xv
Acknowledgements	xvii
1 Introduction and Overview	1
1.1 Motivation	1
1.2 Contribution	3
1.2.1 Problem Outline	3
1.2.2 Solution Pathway	4
1.3 Overview	5
1.3.1 Structure	5
1.3.2 Reader's Guide	6

Part I Foundations

2 Definitions	11
2.1 Ontology	11
2.1.1 Ontology Definition	11
2.1.2 Semantic Web and Web Ontology Language (OWL) ...	14
2.1.3 Ontology Example	16
2.2 Ontology Alignment	19
2.2.1 Ontology Alignment Definition	19
2.2.2 Ontology Alignment Representation	20
2.2.3 Ontology Alignment Example	21
2.3 Related Terms	23
2.4 Ontology Similarity	25
2.4.1 Ontology Similarity Definition	25

- 2.4.2 Similarity Layers 26
- 2.4.3 Specific Similarity Measures 28
- 2.4.4 Similarity in Related Work 34
- 2.4.5 Heuristic Definition 34
- 3 Scenarios 37**
 - 3.1 Use Cases 37
 - 3.1.1 Alignment Discovery 38
 - 3.1.2 Agent Negotiation / Web Service Composition 38
 - 3.1.3 Data Integration 39
 - 3.1.4 Ontology Evolution / Versioning 40
 - 3.1.5 Ontology Merging 40
 - 3.1.6 Query and Answer Rewriting / Mapping 41
 - 3.1.7 Reasoning 42
 - 3.2 Requirements 42
- 4 Related Work 45**
 - 4.1 Theory of Alignment 45
 - 4.1.1 Algebraic Approach 45
 - 4.1.2 Information-Flow-based Approach 46
 - 4.1.3 Translation Framework 47
 - 4.2 Existing Alignment Approaches 47
 - 4.2.1 Classification Guidelines for Alignment Approaches 47
 - 4.2.2 Ontology Alignment Approaches 49
 - 4.2.3 Schema Alignment Approaches 53
 - 4.2.4 Global as View / Local as View 56

Part II Ontology Alignment Approach

- 5 Process 61**
 - 5.1 General Process 61
 - 5.2 Alignment Approach 64
 - 5.2.0 Input 64
 - 5.2.1 Feature Engineering 65
 - 5.2.2 Search Step Selection 67
 - 5.2.3 Similarity Computation 68
 - 5.2.4 Similarity Aggregation 69
 - 5.2.5 Interpretation 72
 - 5.2.6 Iteration 74
 - 5.2.7 Output 75
 - 5.3 Process Description of Related Approaches 76
 - 5.3.1 PROMPT, Anchor-PROMPT 76
 - 5.3.2 GLUE 78
 - 5.3.3 OLA 79

5.4	Evaluation of Alignment Approach	81
5.4.1	Evaluation Scenario	81
5.4.2	Evaluation Measures	82
5.4.3	Absolute Quality	88
5.4.4	Data Sets	88
5.4.5	Strategies	91
5.4.6	Results	92
5.4.7	Discussion and Lessons Learned	95
6	Advanced Methods	97
6.1	Efficiency	97
6.1.1	Challenge	97
6.1.2	Complexity	98
6.1.3	An Efficient Approach	100
6.1.4	Evaluation	104
6.1.5	Discussion and Lessons Learned	106
6.2	Machine Learning	107
6.2.1	Challenge	107
6.2.2	Machine Learning for Ontology Alignment	108
6.2.3	Runtime Alignment	113
6.2.4	Explanatory Component of Decision Trees	114
6.2.5	Evaluation	115
6.2.6	Discussion and Lessons Learned	117
6.3	Active Alignment	119
6.3.1	Challenge	119
6.3.2	Ontology Alignment with User Interaction	120
6.3.3	Evaluation	121
6.3.4	Discussion and Lessons Learned	123
6.4	Adaptive Alignment	124
6.4.1	Challenge	124
6.4.2	Overview	125
6.4.3	Create Utility Function	125
6.4.4	Derive Requirements for Result Dimensions	127
6.4.5	Derive Parameters	128
6.4.6	Example	131
6.4.7	Evaluation	132
6.4.8	Discussion and Lessons Learned	133
6.5	Integrated Approach	135
6.5.1	Integrating the Individual Approaches	135
6.5.2	Summary of Ontology Alignment Approaches	136
6.5.3	Evaluation	136
6.5.4	Discussion and Lessons Learned	138

Part III Implementation and Application

7	Tools	145
7.1	Basic Infrastructure for Ontology Alignment and Mapping – FOAM	145
7.1.1	User Example	145
7.1.2	Process Implementation	146
7.1.3	Underlying Software	147
7.1.4	Availability and Open Usage	148
7.1.5	Summary	149
7.2	Ontology Mapping Based on Axioms	149
7.2.1	Logics and Inferencing	150
7.2.2	Formalization of Similarity Rules as Logical Axioms ...	151
7.2.3	Evaluation	152
7.3	Integration into Ontology Engineering Platform	153
7.3.1	OntoStudio	153
7.3.2	OntoMap	154
7.3.3	FOAM in OntoMap	155
8	Semantic Web and Peer-to-Peer – SWAP	157
8.1	Project Description	157
8.1.1	Core Technologies	158
8.1.2	Case Studies	159
8.2	Bibster	159
8.2.1	Scenario	160
8.2.2	Design	160
8.2.3	Ontology Alignment / Duplicate Detection	163
8.2.4	Application	166
8.3	Xarop	167
8.3.1	Scenario	167
8.3.2	Design	169
8.3.3	Ontology Alignment	173
8.3.4	Application	174
9	Semantically Enabled Knowledge Technologies – SEKT	175
9.1	Project Description	175
9.1.1	Core Technologies	176
9.1.2	Case Studies	176
9.1.3	Ontology Alignment	176
9.2	Intelligent Integrated Decision Support for Legal Professionals	177
9.2.1	Scenario	177
9.2.2	Use Cases	177
9.2.3	Design	178
9.3	Retrieving and Sharing Knowledge in a Digital Library	179

9.3.1	Scenario	179
9.3.2	Use Cases	179
9.3.3	Design	180
9.4	Heterogeneous Groups in Consulting	180
9.4.1	Scenario	180
9.4.2	Use Cases	180
9.4.3	Design	181

Part IV Towards Next Generation Semantic Alignment

10	Next Steps	185
10.1	Generalization	185
10.1.1	Situation	185
10.1.2	Generalized Process	186
10.1.3	Alignment of Petri Nets	187
10.1.4	Summary	191
10.2	Complex Alignments	192
10.2.1	Situation	192
10.2.2	Types of Complex Alignments	193
10.2.3	Extended Process for Complex Alignments	194
10.2.4	Implementation and Discussion	195
11	Future	197
11.1	Outlook	197
11.2	Limits for Alignment	199
11.2.1	Errors	199
11.2.2	Points of Mismatch	200
11.2.3	Implications	201
12	Conclusion	203
12.1	Content Summary	203
12.2	Assessment of Contribution	205
12.3	Final Statements	207

Part V Appendix

A	Ontologies	211
B	Complete Evaluation Results	215

C	FOAM Tool Details	221
	C.1 Short description	221
	C.2 Download and Installation	221
	C.3 Usage	222
	C.4 Web Service	222
	C.5 Parameters	222
	C.6 Additional features of the tool	224
	References	227
	Index	245

List of Figures

1.1	Ontology Alignment	3
2.1	Ontology Example	17
2.2	Ontology Alignment Example	21
2.3	Similarity Layers	27
4.1	Morphisms on Ontologies	46
5.1	Alignment Process	62
5.2	Ontology Alignment Example	66
5.3	Sigmoid Function	71
5.4	Animals Ontology	90
5.5	Results of Label for Russia 1	93
5.6	Results of NOM Weighted for Russia 1	94
5.7	F-Measure for Labels and NOM	95
6.1	Quality over Time for Russia 1	105
6.2	F-Measure and Time for QOM	107
6.3	Detailed Process of APFEL	109
6.4	F-Measure for APFEL Learned Approach	118
6.5	F-Measure for Active Ontology Alignment	123
6.6	Overview Adaptive Alignment	125
6.7	Ranking of Different Strategies	135
6.8	F-Measure, Precision, and Recall for Different Strategies	140
6.9	Time and User Interaction for Different Strategies	140
7.1	Commandline Output	147
7.2	OntoMap	155
8.1	Semantic Web and Peer-to-Peer (SWAP)	158
8.2	Architecture of SWAP	162
8.3	Bibster System	164

8.4	Xarop System	171
8.5	Roles and Functions in Distributed Ontology Engineering	173
9.1	Semantically Enabled Knowledge Technologies (SEKT)	175
9.2	SEKT Architecture for the Legal Case Study	178
10.1	Petri Nets for Flight Reservation	188
11.1	Mismatches Occurring in Central Ontology Case	200
11.2	Mismatches Occurring in Distributed Ontology Case	201
A.1	Hotel Ontology	212
A.2	Bibliographic Ontology	213
C.1	Screenshot of FOAM Webpage	225

List of Tables

2.1	Alignment Table	22
4.1	Ontology Alignment Approaches	54
4.2	Schema Alignment Approaches	57
5.1	Features and Similarity Measures in NOM	68
5.2	Alignment Table with Similarity	76
5.3	Features and Similarity Measures in PROMPT/Label	77
5.4	Features and Similarity Measures in Anchor-PROMPT	78
5.5	Features and Similarity Measures in OLA	80
5.6	Alignment Approaches	81
5.7	Excerpt of Alignment Table from Evaluation	92
5.8	Evaluation Results for NOM	94
6.1	Complexity of Alignment Approaches	101
6.2	Features and Similarity Measures in QOM	104
6.3	Evaluation Results for QOM	106
6.4	Initial Alignments Returned for Validation	110
6.5	Generation of Additional Hypotheses	111
6.6	Training Data for Machine Learning	112
6.7	Evaluation Results for APFEL	116
6.8	Evaluation Results for Active Ontology Alignment	122
6.9	Requirements Based on a Maximum Utility and Use Cases ...	127
6.10	Optimal Parameters Based on Requirements	130
6.11	Parameters Based on Side Condition Constraints	130
6.12	Evaluation Results for Adaptive Alignment	134
6.13	Novel Approaches	137
6.14	Integrated Evaluation Results for Individual Data Sets	139
6.15	Integrated Evaluation Results, Average	139
7.1	Comparison of Procedural and Logics Alignment Approach ...	153

10.1	Features and Similarity Measures for Petri Nets	190
10.2	Results of Petri Net alignment	191
10.3	Features and Heuristics to Identify Subsumption	194
B.1	Complete Evaluation Results 1	216
B.2	Complete Evaluation Results 2	217
B.3	Complete Evaluation Results 3	218
B.4	Complete Evaluation Results 4	219
B.5	Complete Evaluation Results 5	220

Preface

In today's knowledge society, a large number of information systems use many different individual schemas to represent data. Ontologies are one promising approach for representing knowledge in a formal way. Many such ontologies have been developed in recent years. Semantically linking these ontologies is a necessary precondition to establish interoperability between agents or services, or simply humans working with them. Consequently, ontology alignment becomes a central issue, when building a world-wide Semantic Web. Integrating data per se is a billion dollar industry. As one can easily imagine, this cannot be done manually beyond a certain complexity, size, or number of, here, ontologies. Automatic or at least semi-automatic techniques have to be developed to reduce the burden of manual creation and maintenance of alignments.

The purpose of this book is to foster understanding in new semantic technologies, data integration, and the interaction between the two fields. In this application-driven work, the reader is presented a methodology and advice for a concrete tool for aligning ontologies. This is going to be done on theoretical and practical level for both research-focused audiences and developers. Goal is not to align ontologies by only integrating the syntax, but actually bringing together entities which have the same meaning, thus bridging the semantic gap.

The book begins with a short motivation, followed by a thorough investigation of the foundations including up-to-date related work on ontology alignment and application scenarios with their respective requirements. The six-step ontology alignment process consists of determining relevant features of individual entities, selection of promising alignment candidates, similarity assessment and aggregation, interpretation of the similarities for alignment, and, if applicable, several iterations thereof. As result one receives those pairs of entities which correspond to each other. Complex similarity considerations are claimed the key for identifying these alignments. The basic approach is extended through novel methods focusing on efficiency, machine learning optimization, active user inclusion, scenario-adaptive alignment, and an in-

tegrated strategy. The implementation and evaluation shows that both the (semi-) automatic ontology alignment process itself and its output improve significantly. Examples of running applications using the new strategies including one commercial product prove the practical value. Further pointers for next steps in ontology alignment are given including a generalization for other structures and schemas before a summary closes this work.

Using semantic features can help to reach levels of alignment which have never been possible before. The exploitation and application of these advantages is just starting. Methods in this work are basic elements of this development and are expected to be continuously enhanced. In this sense, they will lastingly affect future research and implementation. Therefore, the topic of ontology alignment coupled with the application-focused methodology is appropriate to excite interest of a broad readership.

Karlsruhe, April 2006

Marc Ehrig

Acknowledgements

This book would not have been possible without the invaluable support of many persons. Each of them guided me further towards the goal of creating a comprehensive work on ontology alignment, for which I thank every single one very much.

In first place, it was my advisor Prof. Dr. Rudi Studer, who gave me the challenge and the chance to do this research. He granted me the freedom, the trust, and the help I needed.

I also thank the reviewers comprising Prof. Dr. Kuno Egle, Dr. habil. Jérôme Euzenat, Prof. Dr. Hartmut Schmeck, and Prof. Dr. Ute Werner, who through their objective reflection provided valuable feedback. Further, I am grateful to Susan Lagerstrom-Fife and Sharon Palleschi from Springer for the successful and smooth collaboration on publishing this book.

My friends from Karlsruhe, Peter Haase, Jens Hartmann, Prof. Dr. Steffen Staab, Dr. York Sure, and Christoph Tempich were the closest co-workers, but especially main contributors and critics when I was developing new ideas. This list needs to be extended to all the great friends and colleagues at LS3WIM in Karlsruhe. The nice atmosphere was indeed an essential basis resulting in many professional and social events, which make my time in Karlsruhe unforgettable. Numerous people around the world continuously improved this work through helpful and detailed discussions.

I would like to acknowledge that this research would not have been possible without the funding of the European Union through the projects SWAP, SEKT, and Knowledge Web.

Finally, I thank my family: Aline, Helga, and Volker, who since the very beginning have always been reliable partners at my side motivating me to continue on my path.