

# Ontology Matching

Jérôme Euzenat • Pavel Shvaiko

# Ontology Matching

Second Edition

 Springer

Jérôme Euzenat  
INRIA and LIG  
Grenoble, France

Pavel Shvaiko  
Informatica Trentina SpA,  
while at Department of Engineering  
and Computer Science (DISI),  
University of Trento,  
while at Web of Data, Bruno Kessler  
Foundation - IRST  
Trento, Italy

ISBN 978-3-642-38720-3

ISBN 978-3-642-38721-0 (eBook)

DOI 10.1007/978-3-642-38721-0

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013952732

© Springer-Verlag Berlin Heidelberg 2007, 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

*À mes parents et grand-parents qui ont poussé derrière,  
à Anton et Johanna qui tirent par devant,  
et à Jutta qui marche à mes côtés.*

*Jérôme*

Моим родителям, Ларисе и Леониду,  
за понимание и поддержку.  
*A Marlene, il mio amore, che mi è sempre stata vicino.  
To Tomas and Emma for cheering up and joyful smiles.*

*Pavel*

# Preface

An ontology typically provides a vocabulary describing a domain of interest and a specification of the meaning of terms in that vocabulary. Depending on the precision of this specification, the notion of ontology encompasses several data or conceptual models, e.g., classifications, database schemas, fully axiomatised theories. Ontologies tend to be everywhere. They are viewed as the silver bullet for many applications, such as database integration, peer-to-peer systems, e-commerce, semantic web services, social networks (Fensel 2004). They are, indeed, a practical means to conceptualise what is expressed in a computer format (Brodie et al. 1984; Guarino 2009). However, in open or evolving systems, such as the semantic web, different parties would, in general, adopt different ontologies. Thus, merely using ontologies, like using XML, does not reduce heterogeneity: it raises heterogeneity problems to a higher level.

For instance, imagine two organisations dealing with books: one is a cultural product electronic commerce site (which sells books, music, movies, etc.) and the other is a university library. Both organisations deal with some related products, the books, but are concerned with different aspects of these: the seller is concerned by the margin, the publisher or the type of binding; the library, in turn, pays more attention to the topic, the size and the year of publication. Both are concerned by the price and the author. Yet they may consider these differently, because the price can include tax and shipping fees or not and being expressed in different currencies or because the authors can be denoted by individual objects or by the character string of their names. Moreover, the seller may organise books according to their commercial types and the library according to their literary types. In summary, these two organisations will obviously have different and heterogeneous ontologies.

The book seller and the library may have to interact, for example, because the latter wants to order books to the former or because the former wants to digitise the collections of the latter. In order to do so seamlessly, they need to find the correspondences between the entities in their respective ontologies. The correspondences may express that what is called a book in the ontology of the seller stands for what is called a volume in that of the library. Furthermore, the price in the seller ontology should be multiplied by a tax rate for obtaining the corresponding price in the

library ontology. The process of finding these correspondences is called ‘ontology matching’.

This book is devoted to ontology matching as a solution to the semantic heterogeneity problem faced by computer systems. Ontology matching aims at finding correspondences between semantically related entities of different ontologies. These correspondences may stand for equivalence as well as other relations, such as consequence, subsumption, or disjointness, between ontology entities. Ontology entities, in turn, usually denote the named entities of ontologies, such as classes, properties or individuals. However, these entities may also be more complex expressions, such as formulas, concept definitions, queries or term building expressions. Ontology matching results, called alignments, can thus express with various degrees of precision the relations between the ontologies under consideration.

Alignments can be used for various tasks, such as ontology merging, query answering, data translation or for browsing the semantic web. In the above mentioned example, the library can take advantage of alignments for automatically ordering a book and the seller can use them for checking the availability of a reference by the library. Matching ontologies enables the knowledge and data expressed in the matched ontologies to interoperate. It is thus of utmost importance for the above mentioned applications whose interoperability is jeopardised by heterogeneous ontologies.

Many different matching solutions have been proposed so far from various viewpoints, e.g., databases, information systems, artificial intelligence. They take advantage of various properties of ontologies, e.g., structures, data instances, semantics, or labels, and use techniques from different fields, e.g., statistics and data analysis, machine learning, automated reasoning, and linguistics. These solutions share some techniques and tackle similar problems, but differ in the way they combine and exploit their results. As a consequence, they are quite difficult to compare and describe, lacking a uniform framework.

## **About *Ontology Matching***

*Ontology Matching* aims at being a reference book that presents currently available work in the topic in a uniform framework. In particular, though we use the word ontology, the work and the techniques considered in this book can equally be applied to database schema matching, catalogue integration, XML schema matching and other related problems. The objectives of the book include presenting (i) the state of the art and (ii) the latest research results in ontology matching by providing a detailed account of matching techniques and matching systems in a systematic way from theoretical, practical and application perspectives. The main emphasis of this book is thus on technical solutions for matching.

We have aimed at a sufficiently comprehensive and documented book so that readers can find and learn about almost any subject related to ontology matching and be referred to further reading. Several topics are not covered in full depth but presented only in some salient details for completeness purpose.

It is not the goal of this book to advocate one approach to ontology matching against the others, but rather to show the variety of approaches and their adequacy in different contexts. We are convinced that there is not one unique approach to ontology matching. We concentrate, however, on automatic solutions for matching. Many applications require submitting matching results to user scrutiny and control before using them, but the better the automated part of the task, the easier the control.

This book provides a comprehensive coverage of ontology matching for the researcher and the practitioner. In particular, it reconsiders former frameworks and classifications, broadening their scope and accounting for more solutions. It goes as far as describing in detail basic techniques used in matching systems, reviewing available systems, providing a framework for their evaluation and discussing their applications. This unified view of ontology matching techniques and solutions aims at being the starting point to implementing matching solutions dedicated to a particular application context or developing new techniques. So readers should find in this book inspiration for implementing and understanding matching, they should not expect the ultimate matching solution to be unveiled.

*Ontology Matching* is not meant to be a textbook, though it features exercises for a selected number of chapters. These exercises can help readers in evaluating their understanding of some technical concepts. This book is also complemented by a web site<sup>1</sup> which features additional information and resources.

## Novelty of the Second Edition

Six years have passed since the first edition of *Ontology Matching*, during which the field has made considerable progress Shvaiko and Euzenat (2013). Although, this did not affect the relevance of the first edition, we felt the need to update its content and introduce several novel topics.

A new trend that has risen during the few past years is *linked data* and the subsequent need for data interlinking. Data interlinking falls technically under the definition that we gave for ontology matching and shares enough similarity, so we address this topic to a certain extent (see, in particular, Sects. 1.3, 5.4.2, and 12.4). In fact, ontology matching and data interlinking can be used for improving each other. However, we think that the subject deserves a full autonomous treatment as such.

The new Chap. 3 provides methodological guidelines for people wanting to start a project involving ontology matching. This chapter introduces the alignment life cycle and presents the articulation of the various techniques presented in this book.

Due to the development of matching techniques, the former Chaps. 4 ('Basic techniques') and 5 ('Matching strategies') have been reorganised into three chapters: Chap. 5 covering most of the former Chap. 4 and concerned with local comparison measures, Chap. 6 gathering matching methods working globally on ontologies,

---

<sup>1</sup><http://book.ontologymatching.org>.

which were dispatched in the two former chapters, and Chap. 7 covering most of its former counterpart. The proposed ‘Classification of ontology matching techniques’ (Chap. 4) has also been revised in light of new techniques. The respective number of matching systems and frameworks overviewed in Chaps. 8 and 10 increased from 50 to over 100.

We renamed Chap. 11, ‘Explaining alignments’, into ‘User involvement’. This is to account for the various ways in which users can participate in the matching task beyond being simply explained the output of the process.

Finally, several topics have been worth a new specific section due to their importance in the current matching theory and practice, such as partitioning and pruning (Sect. 7.1.1), context-based matching (Sect. 7.3), matcher tuning (Sect. 7.6), and alignment metadata (Sect. 10.2), to mention a few.

## *Outline of the Book*

This book is organised in five parts.

Part I is dedicated to the motivation and the definition of the ontology matching problem. The motivation is given in Chap. 1 through the presentation of various applications that can take advantage of matching ontologies and the presentation of how matching contributes to these applications. In Chap. 2, the ontology matching problem is technically defined in various instances of ontology matching occurring in different contexts, such as folksonomies, classifications, databases, XML and entity–relationship schemas and finally formal ontologies. It justifies the emphasis of this book on ontology matching and provides definitions for the vocabulary used. Finally, it technically defines the ontology expression languages, the ontology matching process and its result: the alignment. Chapter 3 provides methodological guidelines for carrying out an ontology matching project through the whole alignment life cycle: from matching ontologies to evolving alignments. It articulates most of the remaining chapters in a rational process plan.

Part II provides a comprehensive coverage of the techniques currently used for ontology matching. It is the main part of the book. Chapter 4 defines a classification of matching approaches. Chapter 5 presents the basic similarity or dissimilarity measures that can be used for comparing ontology entities. These techniques are the basis of most ontology matchers. Chapter 6 discusses more elaborate techniques, which match ontologies by comparing them globally. This may involve propagating similarities globally, from basic measures, to reach an equilibrium. The composition of ontology matching systems from these techniques is considered in Chap. 7, which presents techniques that do not perform matching themselves, but rather manipulate matchers and alignments.

Part III is devoted to packaged matching systems that can be manipulated in applications. Chapter 8 presents a large panel of state-of-the-art matching systems. The reader will find that the basic techniques presented before can lead to a large diversity of systems. Chapter 9 is dedicated to the evaluation of matching solutions.

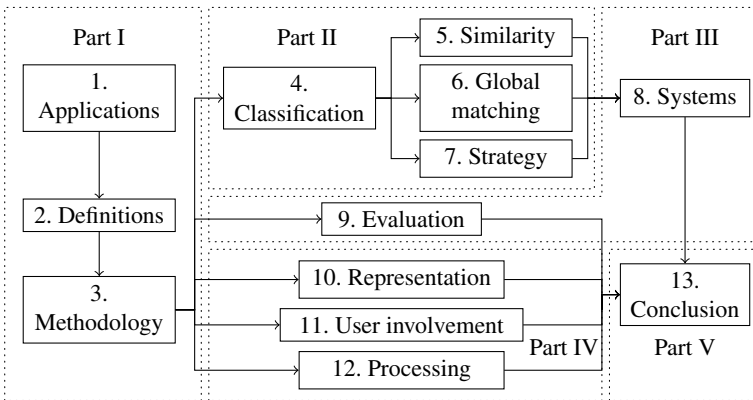


It presents techniques for discriminating empirically among these systems and evaluating their suitability to a particular application.

Part IV is devoted to the use of the ontology matching results in applications once they have been obtained. Chapter 10 considers how alignments can be expressed either for being stored or for being communicated between systems. This chapter also presents frameworks in which alignments may be both obtained and used in various ways. Chapter 11 deals with user involvement. This is important when matching is not expected to be automatic. Finally, Chap. 12 addresses the ultimate use of ontology matching results through their implementation as effective procedures, e.g., rules, articulation axioms, mediators that can be used within applications.

Part V concludes the book, summarising the current state of ontology matching and emphasising remaining problems that will have to be addressed by further research.

A graphical representation of this organisation is presented below. The arrows offer different independent reading paths through the book.



## Readership and Lecture Guide

This book is intended for researchers and practitioners of information and ontology engineering.

The book outline provides a progressive presentation of the ontology matching field and can be read in its entirety. However, each chapter considers ontology matching under a different perspective and can be read in isolation (though it is advised to read the first part before any other). Those who are only interested in getting acquainted with ontology matching can start by reading Chaps. 1, 2 and 13.

For researchers and students dealing with the problem of semantic heterogeneity, we provide not only a comprehensive overview of the state of the art in ontology matching, but also present in detail recent research developments. They show how

ontology matching technologies are going to evolve, indicating which research topics are in the academic agenda and which of them represent the scientific challenges. A course on ontology matching should take some motivations from Chap. 1, explain the concepts introduced in Chap. 2, get inspiration from the guidelines presented in Chap. 3, use the classification of Chap. 4 for exposing Chaps. 5, 6 and 7 and certainly provide some insights from Chap. 9.

For information technology practitioners, both from industry and academia, who want to implement an ontology matching component, this book will help take advantage of state-of-the-art solutions. These readers will take more profit in Chaps. 5, 6, 7, 8, 10, 11 and 12.

For professionals in the areas of e-commerce and knowledge management, the book provides decision support on the use of ontology matching technologies, information about potential problems, and guidelines for the successful application of existing approaches. These readers will take more profit in Chaps. 1, 2, 3, 4, 8, 9, 10 and 12.

We only expect from readers a basic knowledge about data and conceptual modelling and graph theory. Knowledge about logics can also be helpful, though not strictly necessary.

## Acknowledgements

The work presented in this book has been partly supported by the Knowledge Web network of excellence (IST-2004-507482) of the European Commission 6th Framework Programme for Research and Technological Development. We emphasise the crucial role played by the European networks of excellence in providing support for cooperative research on important and emerging topics such as this one. This book testifies to the rich working atmosphere these networks contributed to create.

We thank all the participants of the Heterogeneity workpackage of Knowledge Web and, in particular, Than-Le Bach, Jesus Barrasa, Paolo Bouquet, Jan De Bo, Jos De Bruijn, Rose Dieng-Kuntz, Enrico Franconi, Raúl García Castro, Manfred Hauswirth, Pascal Hitzler, Mustafa Jarrar, Markus Krötzsch, Ruben Lara, Malgorzata Mochol, Amedeo Napoli, Luciano Serafini, François Sharffe, Giorgos Stamou, Heiner Stuckenschmidt, York Sure, Vojtěch Svátek, Valentina Tamma, Sergio Tessaris, Paolo Traverso, Raphaël Troncy, Sven van Acker, Frank van Harmelen, and Ilya Zaihrayeu.

Some people had a specific impact on the book through many fruitful discussions, detailed technical feedback on various ontology matching themes, joint work and continuous support during the time we have been elaborating on it. We are very grateful for this to Marc Ehrig, Fausto Giunchiglia, Chan Le Duc, Loredana Laera, Angela Locoro, Diana Maynard, Deborah McGuinness, Petko Valchev, Mikalai Yatskevich, and Antoine Zimmermann.

We also thank Amedeo Napoli for his careful reading, as well as Vincent Englebert, Janina Fengel, Jake Ge, Karl Hammar, Mahboobeh Houshmand, Wei Hu,

Sajjad Hussain, Antoine Isaac, Jun Liang, Tobias Rafreider, Tuukka Ruotsalo, Juan Sequeda, Ondřej Šváb-Zamazal, Sinan Yurtsever, and Antoine Zimmermann for pointing out mistakes in the first version of this book. We are indebted to Fiona McNeill for the time she kindly spent on a first complete draft of this book and her insightful suggestions. Jérôme specially thanks Fred Freitas for hosting him at Porto de Galinhas to quietly work on the second edition of the book.

Finally, we are grateful to our Springer Verlag editor, Ralf Gerstner, for his belief that we had material for such a book and for his kind patience during its production.

Grenoble, France  
Trento, Italy

Jérôme Euzenat  
Pavel Shvaiko

# Contents

## Part I The Matching Problem

<b>1</b>	<b>Applications</b>	3
1.1	Ontology Engineering	3
1.2	Information Integration	5
1.3	Linked Data	11
1.4	Peer-to-Peer Information Sharing	13
1.5	Web Service Composition	16
1.6	Autonomous Communication Systems	18
1.7	Navigation and Query Answering on the Web	20
1.8	Summary	23
<b>2</b>	<b>The Matching Problem</b>	25
2.1	Vocabularies, Schemas and Ontologies	25
2.2	Ontology Language	33
2.3	Types of Heterogeneity	37
2.4	Terminology	39
2.5	The Ontology Matching Problem	41
2.6	Summary	53
<b>3</b>	<b>Methodology</b>	55
3.1	The Alignment Life Cycle	56
3.2	Identifying Ontologies and Characterising Needs	57
3.3	Retrieving Existing Alignments	60
3.4	Selecting and Composing a Matcher	61
3.5	Matching Ontologies	63
3.6	Evaluating Alignments	64
3.7	Enhancing Alignments	66
3.8	Storing and Sharing	67
3.9	Rendering and Processing Alignments	68
3.10	Summary	68

**Part II Ontology Matching Techniques**

**4 Classifications of Ontology Matching Techniques** . . . . . 73

4.1 Matching Dimensions . . . . . 73

4.2 Classification of Matching Approaches . . . . . 75

4.3 Classes of Concrete Techniques . . . . . 79

4.4 Other Classifications . . . . . 82

4.5 Summary . . . . . 83

**5 Basic Similarity Measures** . . . . . 85

5.1 Similarity, Distances and Other Measures . . . . . 85

5.2 Name-Based Techniques . . . . . 87

5.3 Internal Structure-Based Techniques . . . . . 106

5.4 Extensional Techniques . . . . . 112

5.5 Summary . . . . . 120

**6 Global Matching Methods** . . . . . 121

6.1 Relational Techniques . . . . . 121

6.2 Iterative Similarity Computation . . . . . 130

6.3 Matching as Optimisation . . . . . 137

6.4 Probabilistic Matching . . . . . 140

6.5 Semantic Techniques . . . . . 145

6.6 Summary . . . . . 148

**7 Matching Strategies** . . . . . 149

7.1 Ontology Partitioning and Search-Space Pruning . . . . . 149

7.2 Matcher Composition . . . . . 153

7.3 Context-Based Matching . . . . . 156

7.4 Similarity and Alignment Aggregation . . . . . 160

7.5 Matching Learning . . . . . 172

7.6 Matcher Tuning . . . . . 180

7.7 Alignment Extraction . . . . . 186

7.8 Alignment Improvement . . . . . 192

7.9 Summary . . . . . 196

**Part III Systems and Evaluation**

**8 Overview of Matching Systems** . . . . . 201

8.1 Schema-Based Systems . . . . . 203

8.2 Instance-Based Systems . . . . . 233

8.3 Mixed, Schema-Based and Instance-Based Systems . . . . . 242

8.4 Metamatching Systems . . . . . 262

8.5 Summary . . . . . 269

- 9 Evaluation of Matching Systems** . . . . . 285
  - 9.1 Evaluation Principles . . . . . 285
  - 9.2 Data Sets for Evaluation . . . . . 291
  - 9.3 Evaluation Measures . . . . . 300
  - 9.4 Application-Specific Evaluation . . . . . 314
  - 9.5 Summary . . . . . 317
- Part IV Representing, Explaining, and Processing Alignments**
- 10 Frameworks and Formats: Representing Alignments** . . . . . 321
  - 10.1 Alignment Formats . . . . . 321
  - 10.2 Alignment Metadata . . . . . 337
  - 10.3 Alignment Frameworks . . . . . 340
  - 10.4 Summary . . . . . 350
- 11 User Involvement** . . . . . 353
  - 11.1 Individual Matching . . . . . 353
  - 11.2 Collective Matching . . . . . 357
  - 11.3 Explaining Alignments . . . . . 360
  - 11.4 Alignment Editors and Visualisers . . . . . 369
  - 11.5 Summary . . . . . 375
- 12 Processing Alignments** . . . . . 377
  - 12.1 Ontology Merging . . . . . 378
  - 12.2 Ontology Transformation . . . . . 380
  - 12.3 Data Translation . . . . . 381
  - 12.4 Data Interlinking . . . . . 385
  - 12.5 Mediation . . . . . 387
  - 12.6 Reasoning . . . . . 389
  - 12.7 Alignment Services and Repositories . . . . . 390
  - 12.8 Alignment Evolution . . . . . 394
  - 12.9 Summary . . . . . 395
- Part V Conclusions**
- 13 Conclusions** . . . . . 399
  - 13.1 A Brief Outlook of the Trends in the Field . . . . . 399
  - 13.2 Future Challenges . . . . . 401
  - 13.3 Final Words . . . . . 404
- Appendix A Legends of Figures** . . . . . 407
- Appendix B Running Example** . . . . . 409
- Appendix C Exercises** . . . . . 423
- Appendix D Solutions** . . . . . 431
- References** . . . . . 463
- Index** . . . . . 497