Heiko Angermann
Naeem Ramzan

Taxonomy Matching Using Background Knowledge

Linked Data, Semantic Web and Heterogeneous Repositories

Springer

Taxonomy Matching Using Background Knowledge

Heiko Angermann · Naeem Ramzan

Taxonomy Matching Using Background Knowledge

Linked Data, Semantic Web and Heterogeneous Repositories

Springer

Heiko Angermann
University of the West of Scotland
Paisley
UK

Naeem Ramzan
University of the West of Scotland
Paisley
UK

ISBN 978-3-319-72208-5 ISBN 978-3-319-72209-2 (eBook)
https://doi.org/10.1007/978-3-319-72209-2

Library of Congress Control Number: 2017960928

© Springer International Publishing AG 2017
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.
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.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Taxonomies and its related technologies ontologies, schemas, directories, and electronic catalogs are using a hierarchical order to model a field of interest in a formal way and have become ubiquitous across domains and information management systems during the last years. Those are crucial to categorize customers according to their accompanying branch inside a Customer Relationship Management (CRM) system, are used in Product Information Management (PIM) systems to categorize goods of an enterprise according to the most related category, are necessary to classify the to the goods belonging assets inside a Media Asset Management (MAM) system, or are used in e-commerce systems to help customers finding the desired products.

At the same time, enterprises have the need to exchange data between the systems inside a complex data warehouse scenario, but also to exchange data between other providers. For example, the interaction between CRM and PIM can be used to filter for relevant customers, when setting up a new marketing strategy for specific branches. Or, when the own products and services have to be distributed in global marketplaces like Amazon or eBay, instead of only using the own marketplace, to increase sale. However, as the single systems and different providers usually use their own taxonomy to classify the instances according to the interpretation and embossing of the domain, the exchange is significantly made much more difficult. Correspondences according to the semantic similarity between the categories, i.e., more precisely concepts, have to be detected, to allow a (semi)-automatic exchange between the databases.

Aim of the Book

During last years, an increasing number of algorithms and systems has been proposed, which aim to find correspondences between taxonomies. Such taxonomy matching algorithms and systems detect the similarity between the taxonomies laborious or facile depending on the differences between the taxonomies

(i.e., taxonomic heterogeneity), and depending on the utilized matching techniques. In contrast to previous years, recent approaches now use an interaction between different matching techniques inside one matching strategy to overcome different types of heterogeneity inside one framework. In addition, recent attempts make use of various resources of background knowledge, to further help inferring similarities between the repositories, instead of merely using the knowledge derived from the two taxonomies. Both, the improved matching strategies and the usage of flexible background knowledge, have proven a significant improvement in the field according to various dynamic, as well as static benchmarks.

The book at hand is anticipated to be an assortment of existing and new matching techniques, a review of new introduced matching algorithms, an analysis of recently provided matching systems, as well as a comparison between matching evaluation approaches. In addition, it will investigate the different attempts according to the most important evaluation campaign in the field, namely the Ontology Alignment Evaluation Initiative (OAEI). Through this, the principal aspiration of this book is not only to cover the theoretical and practical fundamentals, but also to cover the state-of-the-art techniques and challenges of the subject area. A number of promising developments and innovative guidelines will also be explained in this book with the aim of motivating and informative opportunity for academic research and practice. In short, the following four objectives are followed by the book at hand:

- The main objective of this book is to provide an in-depth knowledge about the state of the art in the field of taxonomy matching, including its related fields ontology matching, and schema matching.
- It will address new generation matching attempts, including matching strategies, matching algorithms, matching systems, OAEI campaigns, as well as alternative evaluations.
- It will confer the way in which the recent and future attempts make use of different sources of background knowledge to allow a more precise matching between the repositories.
- It will cover research aspects, but also aspects to practitioners that make use of the matching approaches.

Outline of the Book

Part I is dedicated to the definition and motivation of the problem taxonomic heterogeneity. For this part, two chapters are included. The background to the research field is given in Chap. 1. Hereby, the problem of taxonomy matching is explained in detail. In addition, a categorization of works is provided. Afterwards, the different types of taxonomic heterogeneity are detailed. Finally, the main techniques for evaluating recent works are explained using also Chap. 1. A more fine-granular differentiation between the four different types of heterogeneity is

provided in Chap. 2. Hereby, a novel methodology is presented to clarify when and how taxonomies can be matched laborious or facile. Afterwards, the methodology is applied to terminological heterogeneity, conceptual heterogeneity, syntactical heterogeneity, as well as semiotic heterogeneity.

Part II is focusing on the works aiming to overcome taxonomic heterogeneity, and the progress made in the field. For this part, two chapters are included. Chapter 3 is reviewing the different types of matching techniques existing, before the book is reviewing the most recent matching algorithms. Hereby, algorithms focusing on one of the four types of heterogeneity are included. It is reviewed, in which techniques are used and combined, as well as, which sources of background knowledge are used. Finally, matching systems are reviewed that are able to overcome different types of taxonomic heterogeneity using a more comprehensive matching strategy. The systems included in Chap. 3 have all participated in different evaluation campaigns and have proven to outperform existing systems. Using Chap. 4, the most widely used evaluation methodology is reviewed. Hereby, the campaigns of the last 5 years are discussed in detail.

Part III is discussing related research fields making use of taxonomic heterogeneity. For this part, one chapter is included. In Chap. 5, related research fields aiming to evolve and adapt taxonomies according to insights and interaction are reviewed. This includes the fields of dynamic taxonomies, personalized directories, and catalog segmentation. For each field, the background is explained, concrete applications are discussed, as well as recent approaches are reviewed. Finally, the main techniques for evaluating works of those fields are explained. Afterwards, the research field making use of taxonomies to analyze preferences is reviewed, i.e., the field of recommender systems. Hereby, the different types of recommender systems making use of taxonomies are discussed.

Part IV concludes the book. Chapter 6 is summarizing the findings of the book and discussing the recent state of taxonomic taxonomy matching.

Readership and Lecture Guide

The target audience of this book will be composed of professionals and researchers working in the fields of taxonomy matching and its related fields ontology and schema matching in various disciplines, especially involved with artificial and business intelligence. The book should also serve as a solid and advanced-level course supplement to taxonomy matching for upper-level undergraduate and postgraduate students studying this subject. In addition, the book will offer insights and support to groundbreaking and innovative studies apprehensive with the development of taxonomy matching approaches within different types of research and working communities. In short, the target audience is as follows:

- Researchers with a focus on the fields of taxonomy matching, including taxonomy mapping and merging, or with a focus on the field of semantic similarity assessment.
- Lectures giving courses in the disciplines of information management, including the subdisciplines of electronic commerce, master data management, as well as content management.
- Graduated who specialize in the area of metadata management, including the techniques of formal metadata management as well as informal metadata management.
- Practitioners who specialize in the areas of data processing and analyzing, including the subareas of data science and business intelligence.

Paisley, UK
August 2017

Heiko Angermann
Naeem Ramzan

Contents

Part I Introduction to Taxonomy Matching

1 Background Taxonomy Matching 3
1.1 Taxonomy Principles 5
1.2 Matching Taxonomies 7
1.3 Taxonomic Heterogeneity 9
1.4 Matching Categorization 10
1.5 Matching Evaluation 10
1.6 Conclusions 13

2 Background of Taxonomic Heterogeneity 15
2.1 Methodology of Taxonomic Heterogeneity 17
2.1.1 Problem Statement 18
2.1.2 Degrees of Heterogeneity 19
2.2 Heterogeneity Investigation 19
2.2.1 Terminological Heterogeneity 20
2.2.2 Conceptual Heterogeneity 20
2.2.3 Syntactical Heterogeneity 21
2.2.4 Semiotic Heterogeneity 22
2.3 Experimental Evaluation 22
2.4 Conclusions 24

Part II Recent Matching Techniques, Algorithms, Systems, Evaluations, and Datasets

3 Matching Techniques, Algorithms, and Systems 27
3.1 Matching Techniques 28
3.1.1 Element-Level Matching Techniques 29
3.1.2 Structure-Level Matching Techniques 32
3.2 Matching Algorithms 33
3.2.1 Terminological/Conceptual Heterogeneity Matching Algorithms 34

3.2.2 Syntactical Heterogeneity Matching Algorithms 35
3.2.3 Semiotical Heterogeneity Matching Algorithms 37
3.2.4 Analytical Summary Matching Algorithms 38
3.3 Matching System 42
3.3.1 Not yet Another Matcher (YAM++) Matching System 42
3.3.2 Agreement Maker Light (AML) Matching System 43
3.3.3 Generic Ontology Matching (GOMMA) Matching System 44
3.3.4 Logic-Based Matching/Mapping (LogMap) Matching System 44
3.3.5 Ontology Matching Reasoner (OMReasoner) Matching System 45
3.3.6 Risk Minimization/Instance Matching (RiMOM) Matching System 46
3.3.7 Extensible Mapping (XMapxx) Matching System 46
3.3.8 Analytical Summary Matching Systems 47
3.4 Conclusions 48

4 Matching Evaluations and Datasets 51
4.1 OAEI Matching Evaluation Methodology 52
4.1.1 OAEI Matching Evaluation Participants 53
4.1.2 OAEI Matching Evaluation Methodologies and Datasets 53
4.2 OAEI Matching Evaluation Results 56
4.2.1 Benchmark Track 56
4.2.2 Ontology Track 56
4.2.3 Multifarm Track 59
4.2.4 Directories and Thesauri Track 61
4.2.5 Interactive Matching Track 61
4.2.6 Analytical Summary 63
4.3 Alternative Evaluation Attempts and Datasets 66
4.4 Conclusions 67

Part III Taxonomy Heterogeneity Applications

5 Related Areas 71
5.1 Dynamic Taxonomies 72
5.1.1 Dynamic Taxonomies Techniques 73
5.1.2 Dynamic Taxonomies Application 74
5.2 Personalized Directories 75
5.2.1 Personalized Directories in Logic Programming 75
5.2.2 Personalized Directories Techniques 76
5.2.3 Personalized Directories Application 78

5.3 Catalog Segmentation 79
5.3.1 Catalog Segmentation Techniques 79
5.3.2 Catalog Segmentation Application 80
5.4 Analyzing Customers' Requirements with Recommender Systems 80
5.4.1 Types and Characteristics of Recommendation Methods 81
5.4.2 Recommender Systems Techniques 82
5.5 Conclusion 83

Part IV Conclusions

6 Conclusions 87

Glossary 91

References 93

Index 103