

# An Unsupervised Ontology Matching System Using Outlier Analysis

Master Thesis

presented by  
Alexander Mller  
Matriculation Number 1376818

submitted to the  
Chair of Information Systems V  
Prof. .Dr. Heiko Paulheim  
University Mannheim

Mai 2015

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overcoming the Disparate Data Space . . . . .	1
1.2	Contributions . . . . .	1
<b>2</b>	<b>The Ontology Matching Problem</b>	<b>2</b>
2.1	Definitions . . . . .	2
2.1.1	Ontologies . . . . .	2
2.1.2	Ontology Matching . . . . .	4
2.2	Motivating Example . . . . .	8
2.3	State-of-the Art . . . . .	8
2.3.1	Bibliography Benchmark . . . . .	9
2.3.2	Conference . . . . .	9
2.3.3	Anatomy . . . . .	10
2.3.4	Library . . . . .	10
2.4	Challenges . . . . .	13
<b>3</b>	<b>Ontology Matching Approaches (Related Work)</b>	<b>15</b>
3.1	Classification of Approaches . . . . .	15
3.2	Individual Matchers . . . . .	15
3.2.1	Element-level Matchers . . . . .	15
3.2.2	Structure-level Matchers . . . . .	16
3.3	Matching Combination Techniques . . . . .	17
3.4	Analysis Combination Techniques . . . . .	17
<b>4</b>	<b>Selected Approaches to Outlier Analysis (Related Work)</b>	<b>18</b>
4.1	Definition Outlier Analysis . . . . .	18
4.2	Approaches . . . . .	18
<b>5</b>	<b>An unsupervised Ontology Matching System using Outlier Analysis</b>	<b>19</b>
5.1	Motivation for using Outlier Analysis for Ontology Matching . . .	19
5.2	Ontology Matching as an Outlier Detection Problem . . . . .	19
5.2.1	Creating the Feature Vector . . . . .	19
5.2.2	Significance of Outliers for Ontology Matching . . . . .	19
5.2.3	Transforming the Outlier Analysis Result to a Matching . .	19

<b>6</b>	<b>A Matching Pipeline using Outlier Detection</b>	<b>20</b>
6.1	Overview . . . . .	20
6.2	Individual Matchers used . . . . .	20
6.3	Feature Selection . . . . .	20
6.4	Outlier Analysis to Combine the Results . . . . .	20
<b>7</b>	<b>Evaluation</b>	<b>21</b>
7.1	Datasets . . . . .	21
7.1.1	Conference . . . . .	21
7.1.2	Benchmark . . . . .	21
7.1.3	Anatomy . . . . .	21
7.1.4	Library . . . . .	21
7.2	Experimental Setup . . . . .	21
7.3	Used Baselines . . . . .	21
7.4	Results . . . . .	21
<b>8</b>	<b>Discussion</b>	<b>22</b>
8.1	Flexibility towards changing Data Domains . . . . .	22
8.2	Runtime Considerations . . . . .	22
8.3	Comparison with current OAEI Participants . . . . .	22
<b>9</b>	<b>Conclusion</b>	<b>23</b>