

## TABLE OF CONTENTS

<b>List of Figures</b>	<b>iii</b>
<b>List of Tables</b>	<b>v</b>
<b>Chapter 1: Introduction</b>	<b>1</b>
1.1 Applications of Representation Matching . . . . .	1
1.2 Challenges of Representation Matching . . . . .	2
1.3 State of the Art . . . . .	3
1.4 Goals of the Dissertation . . . . .	5
1.5 Overview of the Solutions . . . . .	5
1.6 Contributions of the Dissertation . . . . .	11
1.7 Outline . . . . .	11
<b>Chapter 2: Problem Definition</b>	<b>13</b>
2.1 Data Representations . . . . .	13
2.2 Representation Matching . . . . .	15
2.3 A Semantics for Representation Matching . . . . .	19
2.4 Summary . . . . .	21
<b>Chapter 3: 1-1 Matching for Data Integration</b>	<b>23</b>
3.1 Problem Definition . . . . .	23
3.2 An Overview of Our Approach . . . . .	27
3.3 Multi-Strategy Learning . . . . .	31
3.4 The Base Learners . . . . .	36
3.5 Exploiting Domain Constraints . . . . .	38
3.6 Learning with Nested Elements . . . . .	43
3.7 Empirical Evaluation . . . . .	45
3.8 Discussion . . . . .	50
3.9 Summary . . . . .	51
<b>Chapter 4: Complex Matching</b>	<b>52</b>
4.1 Complex Matching for Relational Schemas . . . . .	52
4.2 The COMAP Approach . . . . .	54
4.3 The Similarity Estimator . . . . .	60
4.4 The Constraint Handler . . . . .	60
4.5 Empirical Evaluation . . . . .	61
4.6 Discussion . . . . .	63

4.7	Summary . . . . .	65
<b>Chapter 5:</b>	<b>Ontology Matching</b>	<b>66</b>
5.1	Introduction . . . . .	66
5.2	The GLUE Architecture . . . . .	70
5.3	Relaxation Labeling . . . . .	74
5.4	Empirical Evaluation . . . . .	78
5.5	Discussion . . . . .	82
5.6	Summary . . . . .	82
<b>Chapter 6:</b>	<b>Related Work</b>	<b>84</b>
6.1	Formal Semantics and Notions of Similarity . . . . .	84
6.2	Representation-Matching Algorithms . . . . .	85
6.3	Related Work in Learning . . . . .	90
6.4	Related Work in Knowledge-Intensive Domains . . . . .	91
<b>Chapter 7:</b>	<b>Conclusion</b>	<b>92</b>
7.1	Key Contributions . . . . .	92
7.2	Future Directions . . . . .	93
<b>Bibliography</b>		<b>95</b>
<b>Appendix A:</b>	<b>Data Processing for LSD Experiments</b>	<b>100</b>
A.1	Selecting the Domains . . . . .	100
A.2	Creating a Mediated DTD and Selecting Sources for Each Domain . . . . .	100
A.3	Creating Data and Manual Semantic Mappings for Each Source in a Domain . . . . .	102
A.4	Creating Integrity Constraints for Each Domain . . . . .	104
A.5	Pseudo Code of LSD . . . . .	105
<b>Appendix B:</b>	<b>Data Processing for COMAP Experiments</b>	<b>114</b>