ONTOMETRY-BASED APPROACH FOR RESOLVING SEMANTIC SCHEMA CONFLICTS IN THE INTEGRATION OF SEMI-STRUCTURED DATA SOURCES

ABDOLREZA HAJMOOSAEI

FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA, KUALA LUMPUR

JUNE 2010
Acknowledgements

I wish to express my appreciation and gratitude to Assoc. Prof. Datin Dr. Sameem Abdul Kareem for her support, guidance and patience during my thesis; I would also like to acknowledge the support of University of Malaya (UM) for my research work. To my parents, I own thanks for their wonderful love and encouragement.
Abstract

The web is the platform for information publishing; it is the biggest resource of information of any type. There are a lot of valuable data and business data on the web that organizations or users may need in order to improve the decision making process. It is therefore, very important and critical that this information are complete, precise and can be acquired on time. Most web data sources provide data in semi-structured forms on the internet. The process of combining semi-structured data from different sources on the internet often fails due to syntactic and semantic differences. The access, retrieval and utilization of information from different web data sources require the data to be integrated. Integration of web data is a complex process because of the heterogeneity nature of the web data and thus needs some kind of a web data integration system.

There are many types of heterogeneity and differences among web data sources that makes data integration a difficult process (e.g., different data model, different syntax and semantics in schema and data instance level among web data sources). Semantic schema heterogeneity, which refers to the misinterpretation of data at the schema level, is one major obstacle that needs to be overcome in web data integration process. Semantic schema heterogeneity has been identified as one of the most important problems when dealing with interoperability and cooperation among multiple data sources on the internet.

In this research, the major aim is to give a solution for resolving semantic schema heterogeneities in web data integration. For this purpose we first recommend an approach and system architecture for web data integration. The proposed web data integration system relies on the ontology technology for resolving of semantic heterogeneity among heterogeneous web data sources. Our proposed web data integration system covers all
abstraction levels of data heterogeneity conflicts between web data sources. The system applies:

- ontology as a solution for resolving schema heterogeneities;
- wrapper as solution for resolving data model heterogeneities;
- converter as solution for resolving data value heterogeneities;

In the second part of the work, we focus on semantic mapping module of proposed web data integration system and propose an approach and algorithm for resolving semantic conflicts between web ontologies. We use semantic ontology mapping as a solution for the reconciliation of semantic schema conflicts between web data. The proposed algorithm uses query path as a technique to enhance the quality of the mapping results and reduce the runtime of the algorithm. The algorithm searches domain ontology in order to find user query concept and its query attributes through query path. The query path gives two strength points to algorithm as follows:

1. *Reduce runtime of each achievement of mapping results:* It directs algorithm toward query concept and its attributes and causes to reduce the search domain of algorithm.

2. *Gain higher quality mapping results:* The query path possesses concepts which have some semantically relation with query concept. Therefore the algorithm has further information about meaning of query concept that helps to find corresponding term with query concept with higher quality mapping results.
Contents

Chapter 1 Introduction

1.1. Background...........................................................................................................2
1.2. Description of Problem.........................................................................................5
   1.2.1 Semantic schema heterogeneity........................................................................6
1.3. Solution for Problem.............................................................................................7
1.4. Outcome of Research...........................................................................................9
1.5. Research Objectives............................................................................................10
1.6. Research Methodology .........................................................................................11
1.7. Thesis Structure....................................................................................................12

Chapter 2 Data Integration System and Ontology

2.1 Introduction...........................................................................................................15
2.2 Overview of Data Integration Approached.............................................................16
2.3 Overview of the Semantic Schema Integration Approaches.................................17
2.4 Web Data Integration Projects...............................................................................18
   2.4.1 Ontology-based web data integration projects.................................................19
       1. COIN..............................................................................................................19
       2. MOMIS.........................................................................................................19
       3. KRAFT..........................................................................................................20
       4. OBSERVER..................................................................................................21
       5. SIMS.............................................................................................................22
   2.4.2 Non ontological web data integration projects...............................................22
1. TSIMMIS.................................................................22
2. Infomaster............................................................23
3. InfoSleuth.............................................................23

2.5 Shortcomings of Reviewed Ontology-based Web Data Integration Systems.................................................................24

2.6 Semantic Ontology Mapping Projects.................................................................26
1. GLUE.............................................................................27
2. QOM .............................................................................28
3. MAFRA.........................................................................29
4. Chimera........................................................................30
5. Anchor-PROMPT..........................................................31
6. Asco.............................................................................31

2.7 Comparison of Reviewed Semantic Mapping Approaches......................33

2.8 Shortcomings of Reviewed Semantic Ontology Mapping Approaches........37

2.9 Architecture of Ontologies in Data Integration System.............................38
2.9.1. Single ontology architecture........................................38
2.9.2. Multiple ontologies architecture.....................................39
2.9.3. Top-level ontology approach.........................................40
2.9.4. Shared vocabulary approach........................................41
2.9.5. Comparison of ontology architectures..............................43

2.10 Comparison of Reviewed Web Data Integration Systems..........................44

2.11 Summary......................................................................47
Chapter 3  Semantics and Ontology

3.1 Introduction..............................................................................................................50
3.2 Schema Data Heterogeneity .................................................................................51
  3.2.1 What is semantic...............................................................................................51
  3.2.2 Semantic heterogeneity in Schema level.........................................................52
3.3 Semantic Web ........................................................................................................55
3.4 Ontology ................................................................................................................58
  3.4.1 What is ontology...............................................................................................58
  3.4.2 Role of ontology...............................................................................................61
  3.4.3 Application area of ontology............................................................................62
3.5 Ontology Language................................................................................................66
  3.5.1 Formalisms mechanisms..................................................................................68
3.6 OWL (Ontology Web Language)............................................................................70
  3.6.1 Definition mechanisms in OWL.........................................................................72
3.7 Summary................................................................................................................76

Chapter 4  Ontology-based Web Data Integration System

4.1 Introduction..............................................................................................................80
4.2 Web Data Integration Approach ............................................................................81
  4.2.1 System approach.............................................................................................82
4.3 System Architecture...............................................................................................87
  4.3.1 Query Construction module............................................................................88
  4.3.2 Web ontology server........................................................................................88
  4.3.3 Semantic mapping module...............................................................................89
Chapter 5  Semantic Ontology Mapping

5.1 Introduction…………………………………………………………………96
5.2 Domain and Local Ontologies………………………………………………97
5.3 Ontology Uniform Representation Model…………………………………98
5.4 User Query Construction .................................................................102
5.5 GUI and Query Path .................................................................103
5.6 Ontology Mapping ........................................................................106
  5.6.1 Definition of ontology mapping ................................................107
  5.6.2 Ontology mismatches ..........................................................109
5.7 Query-based Ontology Mapping Approach and Architecture ............111
  5.7.1 Terminology ........................................................................112
  5.7.2 Mapping module architecture ..................................................113
    5.7.2.1 Transformer ...............................................................114
    5.7.2.2 Mapper .................................................................114
    5.7.2.3 Semantic mapping algorithm .......................................115
5.8 Evaluating of Ontology Mapping Approach ......................................128
  5.8.1 Quality of algorithm results ......................................................130
  5.8.2 Runtime complexity of algorithm ............................................132
5.9 Summary ......................................................................................136
Chapter 6 Implementation and Assessment of Ontology Mapping Approach

6.1 Introduction...........................................................................................................139
6.2 Building of University Domain Ontology.........................................................140
  6.2.1 Methodology for building domain ontology.................................................141
  6.2.2 Building of university ontology.................................................................143
6.3 Graphic User Interface.......................................................................................149
6.4 Sample Local Ontologies..................................................................................150
6.5 Implementation and Assessment of Mapping Algorithm...............................153
6.6 Conclusion.........................................................................................................161

Chapter 7 Conclusion and Discussion

7.1 Introduction.........................................................................................................163
7.2 Summery of the Research..................................................................................163
7.3 Results and Contributions of the Research.......................................................169
7.4 Future Works.....................................................................................................172

Appendix

Appendix A: OWL Coding of University Ontology..............................................176
Appendix B: OWL Coding of USM Ontology........................................................180
Appendix C: OWL Coding of UTM Ontology........................................................184
Appendix D: Implementation of Graphic User Interface......................................188
Appendix E: Implementation of Ontology Mapping Algorithm...........................192
Appendix E.1: The Coding of Mapping Function…………………………………….192
Appendix E.2: The Coding of Ontology Mapping Algorithm…………………………...195
Appendix F: Runtime Complexity of Algorithm…………………………………………211
Appendix G: Publications……………………………………………………………………215
References…………………………………………………………………………………………216
List of Figures

Fig. 1.1 Reservation system A.................................................................2
Fig. 1.2 Reservation system B.................................................................3
Fig. 1.3 Virtual approach architecture....................................................4
Fig. 1.4 Materialization approach architecture........................................4
Fig. 1.5 Data heterogeneity levels...........................................................6
Fig. 1.6 Research methodology.............................................................12
Fig. 2.1 Global schema generation........................................................18
Fig. 2.2 On the fly integration.................................................................18
Fig. 2.3 Comparison framework of reviewed mapping approaches .............35
Fig. 2.4 Single ontology architecture.....................................................39
Fig. 2.5 Multiple ontology architecture................................................39
Fig. 2.6 Top-level ontology architecture................................................41
Fig. 2.7 Shared vocabulary architecture................................................42
Fig. 3.1 Classification of symbols or terms in grammar............................53
Fig. 3.2 Part of grammar shown in syntax grammar.................................54
Fig. 3.3 Example of schema definition...................................................54
Fig. 3.4 Encoding and decoding information on web and semantic web........56
Fig. 3.5 Semantic web architecture........................................................57
Fig. 3.6 Stack of W3C recommendations related to the Semantic Web........71
Fig. 4.1 System approach for resolving semantic schema conflicts.............83
Fig. 4.2 Association between local ontologies and web sources..................84
Fig. 4.3 Partial hierarchical concepts from ontology A.............................85
Fig. 4.4 Partial hierarchical concepts from ontology B.............................85
Fig. 6.11 Mapping quality measure diagram 3........................................157
Fig. 6.12 Mapping quality measure diagram 4........................................158
Fig. 6.13 Mapping quality measure diagram 5........................................159
Fig. 6.14 Mapping quality measure diagram 6........................................159
Fig. 6.15 Mapping quality measure diagram 7........................................160
Fig. 6.16 Mapping quality measure diagram 8........................................160
List of Tables

Table 2.1 Shortcoming of ontology-based web data integration systems..................26
Table 2.2 Benefits and drawbacks of the different ontology-based Integration
approaches.................................................................44
Table 2.3 Comparison of ontology-based web data integration systems...................45
Table 5.1 Concept Hierarchy table.................................................................100
Table 5.2 Concept table.............................................................................101
Table 5.3 Attribute table...........................................................................101
Table 5.4 Relation table.............................................................................101
Table 5.5 Concept Relationship table........................................................101
Table 5.6 Key Properties table..................................................................102
Table 5.7 Mapping table for ontologies $O_1$ and $O_2$.....................................108
Table 5.8 Concept mapping table...............................................................125
Table 5.9 Attribute mapping table..............................................................126
Table 5.10 Concept mapping table2............................................................128
Table 5.11 Attribute mapping table2............................................................128
Table 5.12 Comparison framework of our mapping approach........................128
Table 5.13 Algorithm complexity.................................................................135
Table 6.1 Mapping algorithm results on Univeristy and USM ontologies..........155
Table 6.2 Mapping algorithm results on Univeristy and UTM ontologies...........158
Table 7.1 specification of proposed mapping algorithm.................................168