A SEMANTIC DATA MEDIATOR FRAMEWORK TO SUPPORT AUTOMATION OF WEB SERVICES DATA MEDIATION

KANMANI A/P MUNUSAMY

A thesis submitted in fulfilment of the requirements for the award of the degree of Doctor of Philosophy (Computer Science)

Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia

JULY 2013
To my husband Mr. Sugumaran and
sons Sarveshaah and Avhineshah
ACKNOWLEDGEMENT

I would like to express my deepest appreciation my principle supervisor, Associate Prof. Dr. Suhami Ibrahim and co-supervisor Prof. Dr. Sapiyan Baba for their continued support and invaluable constructive comments which have helped me in completing this research on time. I am also indebted to Dr. Sayed Gholam Hassan Tabatabaei who inspired me in selecting this research area, and Dr. Mohd Naz’ri Mahrin for his vivacious discussions that kept my focus on the research problems at hand.

I also thank my two special colleagues Ms. Hazlifah Mohd Rusli and Dr. Mansoor Abdullah for their stimulating conversations and taking interest in my research work. I am also thankful to University Malaya and Ministry of Higher Education for funding my studies for this PhD programme. My special gratitude goes to my mother, parents-in-law, sisters, brothers and friends, Shanthi Marie and Thana for their encouragement at all times. My special thanks also go to all the staff at AIS, co- researchers at UTM and colleagues at PTM, University Malaya who have helped me in my ways complete this research.

At last and most importantly, this programme would not have been possible without the support of my husband, Mr. Sugumaran. I am thankful for his endless patience and love, consistent support and motivation for me to be where I am at present.
ABSTRACT

Most businesses nowadays use Web Services (WS) technology as a platform to facilitate interaction between service providers and requestors. Data mediators in these services play an important role in ensuring successful interactions; however, the Semantic Web Service (SWS) still faces great challenges in providing the mediation actions that are necessary for smooth WS interactions and is thus open for further research exploration and automation in discovery, selection and composition. Many existing data mediation approaches focus on automated ontology mapping that provides only limited discussions on mediating actual instances. As such, current approaches suffer from insufficient mediation knowledge for related domains to mediate messages correctly at run time. The objective of this thesis is to construct a data mediation framework for the SWS and its associated processes that can establish data mediation automatically for WS interactions at run-time. The Semantic Data Mediator Framework (SDMF) is proposed to manage interactions between source messages and target messages by expressing the data mediation knowledge of developers in the form of semantic knowledge representation. The research steps in engineering method research methodology are adopted to identify the required improvement and design the SMDF data mediation solution. A data mediation component that mediates messages during a WS interaction between scholarly databases and local universities is developed using the proposed SDMF. The evaluation results on the semantic data mediator component using the SDMF are benchmarked with an existing middleware application that is used to support the data mediation. The evaluation results prove that semantic descriptions of the Web service message content through the SDMF are able to enhance the correctness and automation during the run time of a WS interaction.
ABSTRAK

Kebanyakan perniagaan kini menggunakan teknologi Perkhidmatan Web (PW) sebagai platform untuk membolehkan interaksi antara pembekal dan peminta perkhidmatan. Pengantara data dalam PW memainkan peranan penting untuk memastikan interaksi yang berkesan. Perkhidmatan Web Semantik (PWS) telah memberi ruang yang lebih banyak kepada penerokaan penyelidikan ke arah mengautomasikan pemilihan, penemuan dan komposisi PW. Walaubagaimanapun, PWS masih menghadapi cabaran yang besar dalam menyediakan tindakan pengantaraan yang diperlukan bagi melancarkan interaksi PW. Kebanyakan pendekatan pengantaraan data yang sedia ada, memberi tumpuan kepada pemetaan ontologi secara automatik dan hanya menyediakan perbincangan secara terhad bagi pengantaraan data dengan nilai sebenar. Oleh yang demikian, pengetahuan pengantaraan data semasa bagi sesuatu domain tidak mencukupi untuk menjana pengantara mesej secara automatik pada masa larian. Objektif tesis ini adalah untuk membina rangka kerja pengantaraan data bagi PWS dan proses yang berkaitan dengannya bagi mewujudkan pengantara data secara automatik semasa interaksi PW. Rangka kerja Pengantara Data Semantik (RPDS) adalah dicadangkan untuk menguruset interaksi antara mesej sumber dan mesej sasaran dengan menyatakan data pengantaraan pengetahuan untuk pembangun sistem mewakilkannya sebagai pengetahuan semantik. Langkah-langkah yang terkandung dalam kaedah kejuruteraan telah digunakan sebagai kaedah penyelidikan untuk mengenalpasti penambahbaikan yang diperlukan dan merekabentuk pengantaraan data RPDS. Satu komponen pengantaraan data yang berfungsi sebagai pengantara mesej PW semasa interaksi antara pangkalan data ilmiah dan universiti tempatan telah dibangunkan menggunakan RPDS. Hasil penilaian ke atas komponen pengantaraan semantik data menggunakan RPDS diukur dengan membandingkan dengan aplikasi perisian tengah yang sedia ada untuk menyokong pengantaraan data. Keputusan penilaian membuktikan bahawa perwakilan pengetahuan semantik bagi kandungan mesej PW melalui RPDS dapat dipertingkatkan ketepatannya dan mengautomasikan interaksi PW.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td></td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td></td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td></td>
<td>vi</td>
</tr>
<tr>
<td>TABLE OF CONTENT</td>
<td></td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>xvii</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td></td>
<td>xxi</td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td></td>
<td>xxii</td>
</tr>
</tbody>
</table>

1 INTRODUCTION

1.1. Background of the Problem
1.1.1. Issues in Web Service Interactions
1.1.2. Issues of Instance Transformation in Data Mediation
1.2. Problem Statement
1.3. Research Questions
1.4. Objectives of the Study
1.5. Scopes of the Study
1.6. Significance of the Study
1.7. Glossary
1.8. Thesis Outline

1
2. LITERATURE REVIEW

2.1. The Need for Semantics in Web Services

2.1.1. Definition of Web Service

2.1.2. Definition of the Semantic Web

2.1.2.1. Semantic Representation using the Ontology

2.1.2.2. Semantic Languages

2.1.2.3. Rules Languages

2.1.3. Evolution of the Semantic Web and Web Services

2.1.4. Web Service Modelling Ontology (WSMO)

2.2. Message Mismatches in the WS Interactions

2.2.1 Level of Message Mismatches in the WS Interaction

2.2.2 Types of Message Mismatches in WS Interaction

2.2.3 Role of the Mediator in Supporting Message Mismatches

2.2.3.1 Adaptation

2.2.3.2 Mediation

2.3. Data Mediation to Support Interaction Message Mismatches

2.3.1 Interaction Message Mismatches

2.3.2 Levels in Data Mismatches

2.3.2.1 Data Mismatches in Semantic Web

2.3.2.2 Data Mismatches in the Ontology Alignment

2.3.3 Components of the Semantic Data Mediator

2.3.3.1 Acquisition Component

2.3.3.2 Unification Component

2.3.3.3 Transformation Component

2.4. Instance Transformation in WSMO

2.4.1 Definition of Instance Transformation

2.4.2 Methods Used in Instance Transformation
2.4.2.1 Instance Unification Techniques using the SPARQL 37
2.4.2.2 Instance Transformation using Rules and the Reasoner 37

2.4.3 The Required Quality Issues for Instance Transformation 38
  2.4.3.1 Automation 38
  2.4.3.2 Correctness 38

2.4.4 Elements in WSMO that Supports Instance Data Transformation 38
  2.4.4.1 The Mediator as a WSMO component 39
  2.4.4.2 The Web Service Element to Describe the Mediator as a Service 41

2.5. Semantic Description of Message Content 44
  2.5.1 Association Analysis in Describing Message Content 45
  2.5.2 Role of Ontology in Describing Message Content 45
  2.5.3 Existing Approaches that Extracts Ontology from Database 46
  2.5.4 Message Manipulation to Create Mediation Action 47
    2.5.4.1 Message Manipulating using the SPARQL 47
    2.5.4.2 Message Manipulating using the SWRL 48
    2.5.4.3 Message Manipulating using the WSML 48

2.6. Chapter Summary 49

3. COMPARATIVE EVALUATION ON STATE-OF-ART DATA MEDIATION APPROACHES 51
  3.1 A Survey on the Need of Semantics and Reasoning to Support Message Mismatches 51
3.1.1 Systematic Review on Message Mediation Approaches
3.1.2 Comparative Evaluation Criteria
3.1.3 Message Mapping
  3.1.3.1 Semantic Support
  3.1.3.2 Mapping Elements
  3.1.3.3 Mapping Rules
  3.1.3.4 Automation in Message Mapping
3.1.4 Identifying Interaction Mismatches
  3.1.4.1 Data and Control Flow Analysis
  3.1.4.2 Analysis Method
  3.1.4.3 Automation in Identifying Message Mismatches
3.1.5 Identifying the Message Mediator
  3.1.5.1 Resolving Interaction Mismatches
  3.1.5.2 Automation Method in Identifying Interaction Mismatches
3.1.6 Generating the Message Mediator
3.1.7 Discussion on the Comparative Evaluation
3.2 Message Mediation Approaches in the SWS
  3.2.1 Message based Mediation Approach
  3.2.2 Message Mediation Algorithm in the SWS Challenges
  3.2.3 Message Mediator as Goal in IRS-III
  3.2.4 Space based Message Mediation in Triple Space Computing
3.2.5 Discussion on Message Mediation Approaches
3.3 Chapter Summary

4. RESEARCH METHODOLOGY
4.1 Research Design
4.2 Research Procedure and Activities
  4.2.1 Literature Review
5. PROPOSED SEMANTIC DATA MEDIATOR FRAMEWORK

5.1 Motivation of Research 94
5.2 Overview of Proposed Framework
   5.2.1 Extending Current Instance Level Data Mediation 96
   5.2.2 Semantic Data Mediator Framework 97
5.3 Process and Interactions within the SDMF
   5.3.1 Knowledge Extraction 102
   5.3.2 Knowledge Representation and Evaluation 102
   5.3.3 Knowledge Modelling into the SWS 103
5.4 Challenges of Research Motivation Addressed by the SDMF
   5.4.1 Semantic Description of Message Content 104
   5.4.2 Applying Message Manipulation 106
   5.4.3 Implementing Description of Message Content within the SWS 107
5.5 Evaluation Model 107
5.6 Chapter Summary 109

6. DESIGN AND IMPLEMENTATION OF THE SEMANTIC DATA MEDIATOR FRAMEWORK 110

6.1 Extracting Data Mediation Knowledge
   6.1.1 Analysing Data Mediation Knowledge 112
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1.1.1</td>
<td>Defining Required Elements in Data Mediation Knowledge</td>
<td>113</td>
</tr>
<tr>
<td>6.1.1.2</td>
<td>Mapping Data Mediation Elements to Database Elements</td>
<td>114</td>
</tr>
<tr>
<td>6.1.2</td>
<td>Extracting DMRO Elements from Database Elements</td>
<td>116</td>
</tr>
<tr>
<td>6.1.2.1</td>
<td>Extracting Concepts</td>
<td>118</td>
</tr>
<tr>
<td>6.1.2.2</td>
<td>Extracting Properties</td>
<td>118</td>
</tr>
<tr>
<td>6.1.2.3</td>
<td>Extracting Instances of Concepts</td>
<td>121</td>
</tr>
<tr>
<td>6.1.2.4</td>
<td>Extracting Taxonomy Based on Related Concepts</td>
<td>123</td>
</tr>
<tr>
<td>6.2</td>
<td>Representing and Evaluating the Mediation Knowledge</td>
<td>127</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Algorithm Representing the Ontology Elements for DMRO</td>
<td>128</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Defining Data Mediation Rules for the DMRO</td>
<td>130</td>
</tr>
<tr>
<td>6.2.2.1</td>
<td>Description of a Taxonomy</td>
<td>131</td>
</tr>
<tr>
<td>6.2.2.2</td>
<td>Finding Equivalent and Disjointed Concepts</td>
<td>132</td>
</tr>
<tr>
<td>6.2.2.3</td>
<td>Creating String Manipulations</td>
<td>133</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Evaluating the Represented DMRO</td>
<td>134</td>
</tr>
<tr>
<td>6.2.3.1</td>
<td>Pellet Reasoner Test</td>
<td>134</td>
</tr>
<tr>
<td>6.2.3.2</td>
<td>Retrieving Pellet Reasoner Results</td>
<td>135</td>
</tr>
<tr>
<td>6.3</td>
<td>Modelling the Mediation Knowledge into the SWS</td>
<td>136</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Defining Ontology Elements in the WSMO</td>
<td>137</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Defining Web Service Elements in the WSMO</td>
<td>140</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Interaction of the DMSWS with the WSMO</td>
<td>142</td>
</tr>
<tr>
<td>6.4</td>
<td>Chapter Summary</td>
<td>144</td>
</tr>
<tr>
<td>7.</td>
<td>EVALUATING THE SEMANTIC DATA MEDIATOR FRAMEWORK</td>
<td>146</td>
</tr>
<tr>
<td>7.1</td>
<td>Benchmarking the Evaluation Method</td>
<td>146</td>
</tr>
<tr>
<td>7.1.1</td>
<td>Evaluation Procedures Based on the DESMET guidelines</td>
<td>147</td>
</tr>
</tbody>
</table>
7.1.2 Research Evaluation Model 148

7.2 Evaluation Preparation 150

7.2.1 Motivating Comparison 150

7.2.1.1 Data Mismatches between the BSD and the HLI 150

7.2.1.2 Service Description given by the provider 152

7.2.1.3 Requestor Data Format 153

7.2.1.4 Comparing Middleware Applications 154

7.2.2 Task Samples 156

7.2.3 Performance Measure 157

7.2.3.1 Evaluating Correctness Using Precision, Recall and F-measure 157

7.2.3.2 Evaluating Protégé Results 158

7.3 Evaluation Execution 160

7.3.1 Executing Middleware Application 161

7.3.2 Executing Semantic Data Mediator using the SDMF 162

7.4 Building Instance Level Data Mediator using the SDMF 163

7.4.1 Extracting Data Mediation Knowledge 163

7.4.1.1 Brainstorming with Domain Experts 163

7.4.1.2 Analyse Internal Database 164

7.4.2 Representing Data Mediation Knowledge as the DMRO 169

7.4.2.1 Extracting Ontology Elements in the DMRO 172

7.4.2.2 Applying the Semantic Data Mediation Rules 183

7.4.2.3 Assigning the String Manipulation Functions to the DMRO 187

7.4.3 Modelling the DMRO into the SWS using the WSMO Framework 189

7.4.3.1 Building the Ontology Component for the SWS 190
7.4.3.2 Building the Web Service Component 194

7.5 Analysing Evaluation Results and Discussion 197

7.5.1 Results from the SDMF 198

7.5.1.1 Evaluating the Extracted DMRO elements 198

7.5.1.2 Evaluating the Taxonomy 199

7.5.1.3 Evaluating the Retrieved Target Data 203

7.5.2 Results from the Existing Middleware Application 205

7.5.3 Discussion on the Evaluation Results 209

7.6 Chapter summary 211

8. CONCLUSION AND FUTURE RESEARCH 212

8.1 Research Summary and Achievement 212

8.2 Summary of the Contributions 214

8.3 Recommendation for Future Research 215

REFERENCES 217

Appendices A-D 226-240
### LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Five ways to address interaction mismatches</td>
<td>30</td>
</tr>
<tr>
<td>2.2</td>
<td>Levels in Data Mismatches [4]</td>
<td>32</td>
</tr>
<tr>
<td>2.3</td>
<td>Level of mismatches in the ontology alignment</td>
<td>32</td>
</tr>
<tr>
<td>2.4</td>
<td>Example of string manipulation functions that are supported by SPARQL [83]</td>
<td>48</td>
</tr>
<tr>
<td>2.5</td>
<td>Example of string manipulation functions that are supported by the WSML [84]</td>
<td>49</td>
</tr>
<tr>
<td>3.1</td>
<td>Refinement Process on Search Results</td>
<td>53</td>
</tr>
<tr>
<td>3.2</td>
<td>Evaluation based on message mapping criteria</td>
<td>58</td>
</tr>
<tr>
<td>3.3</td>
<td>Various patterns that supports protocol mismatches</td>
<td>61</td>
</tr>
<tr>
<td>3.4</td>
<td>Evaluation based on the criteria for identifying interaction mismatches</td>
<td>63</td>
</tr>
<tr>
<td>3.5</td>
<td>Evaluation based on the criteria in identifying the message mediator</td>
<td>65</td>
</tr>
<tr>
<td>3.6</td>
<td>Evaluation based on the generation of the message mediator</td>
<td>68</td>
</tr>
<tr>
<td>3.7</td>
<td>Summary of the comparative evaluation</td>
<td>69</td>
</tr>
<tr>
<td>4.1</td>
<td>Summary of Design Activity</td>
<td>82</td>
</tr>
<tr>
<td>4.2</td>
<td>Operational Framework</td>
<td>86</td>
</tr>
<tr>
<td>6.1</td>
<td>Generated subconcepts for each level</td>
<td>125</td>
</tr>
<tr>
<td>6.2</td>
<td>Definition of the SWRL [111]</td>
<td>131</td>
</tr>
<tr>
<td>6.3</td>
<td>SWRL Rules describing Taxonomy</td>
<td>132</td>
</tr>
<tr>
<td>6.4</td>
<td>SWRL Rules Defining Equivalent and Disjointed concepts</td>
<td>133</td>
</tr>
<tr>
<td>6.5</td>
<td>Relationship between the OWL and the WSML Conceptual Syntax</td>
<td>137</td>
</tr>
<tr>
<td>7.1</td>
<td>An Example of a Publication Record Stored in a UPS Database</td>
<td>154</td>
</tr>
<tr>
<td>7.2</td>
<td>Number of Authors according to Religion and Race</td>
<td>165</td>
</tr>
<tr>
<td>7.3</td>
<td>Axioms that Describe each Sub-concept in Level_1</td>
<td>184</td>
</tr>
<tr>
<td>7.4</td>
<td>Axioms used in Retrieving First and Last Names</td>
<td>189</td>
</tr>
<tr>
<td>7.5</td>
<td>Data Mediation Rules in the WSML Syntax</td>
<td>193</td>
</tr>
<tr>
<td>7.6</td>
<td>Number of Extracted OWL Elements for the DMRO</td>
<td>198</td>
</tr>
<tr>
<td>7.7</td>
<td>Precision, Recall and F-measure for the Extracted Instances</td>
<td>199</td>
</tr>
<tr>
<td>7.8</td>
<td>Precision, Recall and F-measure for DMRO Taxonomy</td>
<td>199</td>
</tr>
<tr>
<td>7.9</td>
<td>Precision, Recall and F-measure for retrieved target data from the DMRO</td>
<td>203</td>
</tr>
<tr>
<td>7.10</td>
<td>Precision, Recall and F-measure for the retrieved target data from the existing middleware application</td>
<td>207</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Ontology mapping that supports data mediation</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Role of Developer in Assisting Instance Level Matching in Data Mediation</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>Interaction between a service provider, a requestor and the repository</td>
<td>16</td>
</tr>
<tr>
<td>2.2</td>
<td>Evolution of SWS</td>
<td>19</td>
</tr>
<tr>
<td>2.3</td>
<td>The WSMO Core Elements [40]</td>
<td>21</td>
</tr>
<tr>
<td>2.4</td>
<td>Language and Tools in WSMO</td>
<td>22</td>
</tr>
<tr>
<td>2.5</td>
<td>Hierarchy of Message Mismatches</td>
<td>24</td>
</tr>
<tr>
<td>2.6</td>
<td>Overview of mediator in web services</td>
<td>27</td>
</tr>
<tr>
<td>2.7</td>
<td>Types of message mismatches</td>
<td>29</td>
</tr>
<tr>
<td>2.8</td>
<td>Components of the Data Mediator [60]</td>
<td>33</td>
</tr>
<tr>
<td>2.9</td>
<td>Instance Transformation Scenario in Data Mediation [9]</td>
<td>36</td>
</tr>
<tr>
<td>2.10</td>
<td>Attributes in the WSMO Top Level Elements</td>
<td>39</td>
</tr>
<tr>
<td>2.11</td>
<td>WSMO Mediator Typology [68]</td>
<td>40</td>
</tr>
<tr>
<td>2.12</td>
<td>Ontology precision by [28]</td>
<td>46</td>
</tr>
<tr>
<td>3.1</td>
<td>Selected State-of-the-art Message Mediation Approaches</td>
<td>54</td>
</tr>
<tr>
<td>3.2</td>
<td>Overview of message mediation</td>
<td>55</td>
</tr>
<tr>
<td>3.3</td>
<td>Evaluation criteria for each component</td>
<td>56</td>
</tr>
<tr>
<td>3.4</td>
<td>Data and control flow analysis in Web service interaction</td>
<td>60</td>
</tr>
</tbody>
</table>
3.5 Data flow analysis in Web service interaction 60
3.6 Gap analysis using spider graph 71
3.7 Message-based Mediation Approach 72
3.8 Message Mediation in the SWS Challenge 74
4.1 Research Procedures 84
4.2 Research Flow Chart 85
4.3 Instruments in Building DMRO and DMSWS 89
5.1 Extending current instance-level data mediation using the SDMF 96
5.2 Overview of Semantic Data Mediator Framework 98
5.3 SDMF Components That Supports Message Understanding Processes 100
5.4 Overall Procedures in Developing the Proposed SDMF 101
5.5 Transferring the DMRO (OWL + SWRL) into the DMSWS (WSML) 104
5.6 Steps involved in describing message content 105
5.7 Evaluation model of the proposed SDMF 108
6.1 Overview of SDMF Design Process 110
6.2 Data mediation knowledge extraction process 111
6.3 DM definition for a) splitting messages; b) merging messages 114
6.4 Role of the related columns in describing mediation knowledge between a source and a target data. 116
6.5 Example of Data Mediation Scenario 117
6.6 Extracting objectProperty, Domain and Range. 119
6.7 Extraction of datatypeProperty 121
6.8 Defining a structure of taxonomy using the frequent itemset generation 124
6.9 Data mediation knowledge representation and evaluation process 128
6.10 Algorithm Generating OWL Components Automatically 130
6.11 Data mediation knowledge modeling process 136
6.12 OWL-WSML Translator Tool 139
6.13 Relationship between the SWRL and the WSML 140
6.14 precondition and postcondition for DMRWS 141
6.15 Describing Mediator Functionality using the WSML Elements 142
6.16 Interaction of the DMSWS with the WSMO Components 143
6.17 Steps involved in designing and implementing SMDF 145
7.1 Interaction between the HLI and the BSD using a Middleware 151
7.2 XML data scheme for Authors extracted from the WSDL file 153
7.3 Categories based on gender, religion and race. 156
7.4 Inferred Members for FEMALE_ISLAM_AUTHOR 159
7.5 DL Query to test the Objectproperty of each instance 159
7.6 Inferred first name, middle name and last name. 160
7.7 Results retrieved from the existing middleware application 161
7.8 Building the Semantic Data Mediator using the SDMF Procedures 162
7.9 Classification of authors according to religion and race. 166
7.10 Database that Contains the Authors’ Data 170
7.11 Extracted DMRO Concepts 173
7.12 Extracted objectProperties for the DMRO 174
7.13 Assigned Domain and Range for the hasGender objectProperty 175
7.14 Extracted *datatypeProperties* for the DMRO 176

7.15 Generated Instances for Related Concepts (a) Gender; (b) Religion (c) Race 178

7.16 a) Extracted instances of authors; b) Property assertions 180

7.17 Summary of extracted instances from database 180

7.18 Overview of the Created *Subconcepts* 181

7.19 Extracted *Subconcepts* for Level_1 182

7.20 (a) Subconcepts that need to be removed; (b) The refined subconcepts 183

7.21 Defining the Data Mediation Rules using the SWRLTab Tool. 186

7.22 Defining the Equivalent Concepts for DMRO 187

7.23 Translating the DMRO into the WSML Syntax 190

7.24 The DMRO in the WSML Syntax using the WSMT Tool 191

7.25 Testing the Translated DMRO using Axioms in the IRIS Reasoner 192

7.26 A Visualization of the DMRO in using the WSML-Visualizer 193

7.27 Evaluating the Retrieved Target Data from the DMRO based on a) Precision b) Recall and c) F-Measure 205

7.28 Results retrieved from the existing middleware application 206

7.29 Evaluating the Retrieved Target Data from the Middleware Application based on a) Precision b) Recall and c) F-Measure 208

7.30 Average precision, recall and F-measure for a) DMRO component b) middleware application 210
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARM</td>
<td>Association Rule Mining</td>
</tr>
<tr>
<td>ASM</td>
<td>Abstract State Machine</td>
</tr>
<tr>
<td>DMRO</td>
<td>Data Mediation Rule Ontology</td>
</tr>
<tr>
<td>DMSWS</td>
<td>Data Mediation Semantic Web Service</td>
</tr>
<tr>
<td>FOL</td>
<td>First Order Logic</td>
</tr>
<tr>
<td>GRA</td>
<td>General Reachability Analysis</td>
</tr>
<tr>
<td>HTML</td>
<td>Hyper Text Markup Language</td>
</tr>
<tr>
<td>IOPE</td>
<td>Input, Output, Precondition and Effect</td>
</tr>
<tr>
<td>OWL</td>
<td>Web Ontology Language</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Definition Framework</td>
</tr>
<tr>
<td>RDFS</td>
<td>Resource Definition Framework (Scheme)</td>
</tr>
<tr>
<td>SAWSDL</td>
<td>Semantic Annotation for WSDL</td>
</tr>
<tr>
<td>SDMF</td>
<td>Semantic Data Mediator Framework</td>
</tr>
<tr>
<td>SOAP</td>
<td>Simple Object Access Protocol</td>
</tr>
<tr>
<td>SPA</td>
<td>Specific Protocol Analysis</td>
</tr>
<tr>
<td>SWRL</td>
<td>Semantic Web Rule Language</td>
</tr>
<tr>
<td>SWS</td>
<td>Semantic Web Service</td>
</tr>
<tr>
<td>UDDI</td>
<td>Universal Description, Discovery, and Integration</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>WS</td>
<td>Web Service</td>
</tr>
<tr>
<td>WSDL</td>
<td>Web Services Description Language</td>
</tr>
<tr>
<td>WSMF</td>
<td>Web Service Modelling Framework</td>
</tr>
<tr>
<td>WSML</td>
<td>Web Service Modelling Language</td>
</tr>
<tr>
<td>WSMO</td>
<td>Web Service Modelling Ontology</td>
</tr>
<tr>
<td>WSMX</td>
<td>Web Service Execution Environment</td>
</tr>
<tr>
<td>WWW</td>
<td>World Wide Web</td>
</tr>
</tbody>
</table>
## LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Source code to generate OWL file-generate_owl.php</td>
<td>231</td>
</tr>
<tr>
<td>B</td>
<td>Axioms that describe the taxonomy: For Gender X Religion</td>
<td>238</td>
</tr>
<tr>
<td>C</td>
<td>WSML codes for Web Service Component</td>
<td>242</td>
</tr>
<tr>
<td>D</td>
<td>List Of Publications</td>
<td>245</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

This chapter begins with a brief introduction on the subject of the research, i.e. using the Semantic Data Mediator Framework (SDMF) in supporting automation during the Web Service instance transformation. Various important aspects of the research as a whole are explained. Firstly, the background and statement of the problem are elaborated. This is then followed by the objectives and scope of this research. The final section contains the significance of this research and brief descriptions of some important terms that are used in this research.

1.1. Background of the Problem

It is important to have an overview of the background of the problem before investigating it. Therefore, this section begins with a brief introduction of Web Service (WS), Semantic Web Service (SWS) and interaction mismatches that support problem statements of this study. It is then followed by issues and challenges in instance transformation within the SWS data mediation.

1.1.1. Issues in Web Service Interactions

Web service is a growing technology that has been widely adopted by many organizations. It has provided a medium for communication between the service
provider and the service requestor in the business environment. Generally, the capability of an offered service will be matched with the goal of the requested service from various perspectives in order to discover and select the matched services. However, there is no assurance that these matched services can work well due to the various message mismatches which only are discovered during the actual invocation phase. These mismatches occur in the web services environment due to the significant increase in the number of web services and the distributed nature of the web services themselves. Both the service provider and the requestor are unable to achieve their business goals when there are message mismatches during their web service interactions.

There are five types of mismatches during message interactions namely Extra Messages (EM), Wrong Order Messages (WOM), One to Many Messages (OMM), Many to One Messages (MOM) and Missing Messages (MM). Data level mismatches which occur from the differences in structure, type and naming conventions of the data elements in interacting messages can lead to these mismatches. In order to solve these interaction mismatches, five types of mediation actions have been proposed and these are: stop, merge, split, generate, and reorder [1]. The original messages are split, merged, or reordered according to the required interaction. New messages may even be generated from the original messages in order to conform to the required interaction patterns. Therefore, there is a need for a data mediator to support the proposed mediation actions.

The role of the semantic in web service has played an important part in this communication by allowing automatic discovery, selection and composition between the service provider and the requestor. Data incompatibility can also occur between web services during composition, negation and invocation so therefore, there is a need for the data mediator to solve the data incompatibility problem between the service provider and the service requestor. The data mediator is an important application in problem solving, since it is an important component in automating the SWS discovery, composition and invocation processes. Most of the SWS frameworks like the Web Service Modelling Ontology (WSMO), the Semantic Markup for Web Service (OWL-S) and the Web Service Semantic (WSDL-S) are
using the ontology-based techniques like ontology mapping, ontology alignment and ontology merging to solve the data incompatibility problem [2-6].

1.1.2. Issues of Instance Transformation in Data Mediation

There are two phases in implementation of the data mediation in the SWS namely the design-time and the run-time. A developer’s involvement in the design-time process and the output from this phase will be executed automatically during the run-time. There are two further sub-components in the data mediation component in the SWS namely, the ontology mapping and the instance transformation. The existing approaches focus on creating the ontology mapping automatically; and only provide limited discussion on the instance transformation component.

![Figure 1.1: Ontology mapping that supports data mediation](image)

The current data mediation approach in SWS is illustrated by using a simple scenario as shown in Figure 1.1. It can be assumed that there is an OMM mismatch between the two Web services namely A and B which require a data mediator to split the message. Web service A contains the full name attributes of the Authors’ ontology, whereas Web service B is expecting three attributes termed as first name, middle name and last name which are linked to another ontology known as the Contributors as shown in Figure 1.1.
The existing approaches in the SWS only focus on generating the ontology mapping for one-to-many message mismatches. Various techniques to generate attributes mapping between full name to first name, middle name and last name to support web services interaction are used by the researcher. In this scenario, the data mediation must be able to split a single message into three messages.

Figure 1.2: Role of Developer in Assisting Instance Level Matching in Data Mediation

The machines that support the SWS interaction need to understand the actual instance to ensure that the message can be mediated correctly according to the content and the provided mapping as shown in Figure 1.2. It shows the role of the developer in the data mediation to understand the content of the messages and split them meaningfully for the provided data mediation scenario. It can be assumed that the author ontology contains an instance of full name, “Hazlifah Binti Mohd Rusli”. In this case the machines are unable to split the author’s full name by just basing on the ontology mapping.
1.2. Problem Statement

As explained in the problem background, ontology mapping itself is insufficient to handle the mediation actions. The existing automatic ontology mapping approaches in the SWS still requires the assistance of the developer to split the messages correctly according to the generated mapping. Thus, existing ontology mapping approaches in the SWS do not guarantee that the source messages can to be split into target messages correctly without the developer’s role at run time.

The existing approaches use handlers to store and send messages based on the interaction mismatch patterns. These approaches are only able to resolve the interaction mismatches due to the EM and WOM. They could stop the extra messages and reorder the incorrect messages. However, it requires further data mediation functionalities to support the OMM, MOM and MM mismatches. The existing semantics in the web service only focus on generating the ontology that describes each concept or class in a web service. The ontological descriptions of the concepts need to be extended to describe the content of the messages/instances. The ontological descriptions on the message content and the message manipulations are able to support mediation action such as splitting, merging and generating new messages during instance transformation.

The following are the challenges of the instance transformation to support meaningful data mediation for OMM, MOM and MM type interaction mismatches automatically and correctly at run-time:-

(a) Need to describe the content of the provided source message using ontology.

(b) Need to apply correct message manipulation to produce the required target message.

(c) Need to transform the content descriptions and message manipulation into the SWS environment.
1.3. Research Questions

The ultimate goal to of this research is to provide automated data mediation approach during instance transformation to support OMM, MOM and MM mismatches in web service interactions using the WSMO elements. The output of this research is expected to increase the correctness and automation in other tasks in the SWS such as composition, negation, testing and invocation. The general research question is as follows:

*How to enhance the existing data mediator to support interaction message mismatches automatically during the instance transformation?*

The following research questions are formulated to address the stated general research question and the discussed problems in this research area.

(a) **RQ1**: What are the elements that are required in supporting the instance level data mediation during a web service interaction?

(b) **RQ2**: How to represent a domain expert’s data mediation knowledge into a machine readable form?

   (i) How to describe a message content using ontology?

   (ii) How to describe message manipulations using ontology?

   (iii) How to query the ontology to extract the required messages according to the mediation actions?

(c) **RQ3**: What are the required components and processes involved in building the proposed SDMF?

(d) **RQ4**: How to build the SDMF to support instance level data mediation during web service interaction?

(e) **RQ5**: How to evaluate validity of the proposed approach to support instance level data mediation?
1.4. **Objectives of the Study**

The objectives of this study have been derived from the problem statement above. The objectives of this research are to:

(a) to study and investigate issues in Web services interactions associated to data mediation

(b) to develop a SDMF using the Data Mediation Rule Ontology (DMRO) and the Data Mediation Semantic Web Service (DMSWS).

(c) to evaluate the correctness of data retrieved from the DMRO using the Pellet Reasoner for splitting message and benchmark them with existing data mediation middleware applications.

1.5. **Scopes of the Study**

This section describes the limitations and the boundaries of this study. Below are the scopes of this research:-

(a) This research only focuses on three interaction mismatches in the web services namely OMM, MOM and MM mismatches and the respected mediation actions.

(b) This research describes all the activities and processes in developing ontology termed as the DMRO which describes the message content and the message manipulation to support mediation actions.

(c) The DMRO that describes the message content is modelled based on the knowledge extracted from a relational database and focuses on the string manipulations that support the mediation actions.

(d) This research also illustrates how the DMRO can be implemented as a SWS web service using the WSMO elements.

(e) This approach only evaluates based on two quality issues namely the automation and the correctness.
1.6. Significance of the Study

Related work on data mediation was analysed in the existing SWS Frameworks namely the WSMO, the OWL-S and the SAWSDL. Early data mediation effort in the Web services is found in [7] which introduces the mapping rules between the RDF scheme in the Triple Space Computing. This effort is extended by the WSMO data mediation initiatives which focus on the ontology alignment [5, 8, 9].

The WSMO initiatives generate mapping rules in the form of axioms based on the abstract mappings identified by the developers at the time of design. They demonstrate that the source instances that can be transformed into the target instances via posting query and retrieving answers from the mapping rules at the time of execution. However, the data mediation effort is found to focus only on the generating alignment between the attributes that are placed at different levels within the ontology. It describes implicitly the data mediation effort that involves splitting and merging messages during the mediation process.

Secondly, in the OWL-S, the data mediation that supports the mediation process is not explained in detail and is only mentioned as an external service [10]. The researchers however, have concluded on the need of better support for data mediation in order to allow real life Web service mediation [11]. Finally, data mediation efforts in the SAWSDL introduces the use of the context-based data type ranking algorithm to generate scheme mapping between the Web service messages [12]. Further discussion on data mediation to support process mediation actions is however not provided by the researchers [13].

Due to these limitations in the existing data mediation efforts in the SWS frameworks, this research has proposed the SDMF. This framework highlights how a semantically described message content using the OWL ontology and SWRL rule language can be useful in modelling mediation actions automatically and correctly. This research has also presented all the activities and processes in designing the SMDF which consists of the DMRO and the DMSWS. The DMRO is modelled
from the relational database with guidance from domain experts to express message content using the OWL and the SWRL at design-time. Then, the tested and evaluated DMRO is transformed into SWS using the WSML language to support the mediation action during the instance transformation at run-time. Below are some of the significant contributions of this research:-

(a) This research proposes the SDMF as it is able to overcome the existing data mediation problem in understanding the message content in order to mediate the web service message correctly. It extracts the required data mediation knowledge from the input resources such as database and developers/domain experts and uses the generated knowledge to mediate the messages correctly.

(b) Current SWS initiatives only focus on developing ontology that describes the concepts and classes. This research has proposed ways to describe message content using ontology. It also demonstrates how the message manipulation functionality can be embedded into the ontology and reasoning tools which can be used to retrieve the target messages.

(c) The role of the SDMF can replace the work of developers or data entry clerks to support automation in service discovery, selection, process mediation and composition in SWS.

(d) The DMRO that captures common data mediation knowledge can be reused for other web services interactions for a different service provider and requestor.

1.7. Glossary

This section explains some of the terms that have used in this research. The detailed explanation for each of these terms is provided in the Literature Review section.

- **Web Service** – a software system or technology which describes the services using the XML and these services can be accessed by other software systems using the XML based messages via web. It consists
of three important components which are the WSDL, the UDDI and the SOAP.

- **Semantic Web Service** – a new paradigm that brings semantic descriptions to data and behaviour of web services. It has evolved from the integration of the Semantic Web and Web Service technologies. The current research on SWS focuses on automation of discovery, selection, mediation and composition of the Web Services.

- **Interaction message mismatches** – interaction message mismatches are similar to solvable message mismatches [14] and unspecified reception in the web services interaction as mentioned by [15]. There are five types of interaction message mismatches namely the EM, WOM, OMM, MOM and MM.

- **Mediation Actions** – refers to the solutions or actions for each interaction message mismatches that are mentioned earlier. There are five mediation actions namely stop or hide, inverse order, split, merge and generate.

- **Data mediation** – a component of the SWS that resolves data mismatches in the WS interaction. It contains two important components which are the Ontology Mapping and the Instance Transformation.

- **Instance transformation** – a component that uses the mappings created by the ontology mapping component to assign correct target instance to the respected source instance during run-time.

- **Message manipulations** – refers to built-in functions in the database that manipulates data according to the required structure. There are three main categories in message manipulations that are required in the web services namely the aggregate, the string and the date/time manipulation functions.

- **Web Service Modelling Ontology (WSMO)** – a formal SWS framework that provides semantic descriptions to all the related aspects of the web service. The WSMO consists of four core elements namely the Goal, the Web Service, the Mediator and the Ontology.
• **Web Service Modelling Toolkit (WSMT)** – is an ontology engineering toolkit for the WSMO framework. It provides graphical interface to assist domain experts in creating ontology mapping between source and target ontology [16].

• **Web Service Modelling Language (WSML)** – a concrete formal language of the WSMO framework that is used to describe the Goal, the Web the Service, the Mediator and the Ontology elements.

• **Ontology** - Ontology refers to a formalization of the knowledge in the domain. It is able to interweave human and computer understanding of symbols. Basic building blocks of ontology design include: classes or concepts, properties of each concept describing various features and attributes of the concept such as restrictions or axioms, instances and relationships.

• **Protégé** – is a Java-based open source a stand- alone application that allows a user to load and save the OWL and the RDF based ontology [17]. It also allows the user to edit and visualize classes and properties of ontology and semantic rule languages such as the SWRL.

• **Semantic Web Rule Language (SWRL)** – helps to extend expressivity of the OWL by adding rules to the existing ontology. The SWRL rules contain unary predicates for describing classes and data types, binary predicates for properties, and some special built-in n-ary predicates [18].

### 1.8. Thesis Outline

This research discusses some specific issues of data mediation in the SWS approaches during message interactions. It also highlights the limitation of the existing approaches in resolving interaction mismatches at instance transformation. It describes a proposed SDMF that enhances automation and correctness of the existing approaches. The SDMF uses the OWL ontology and the SWRL rules to describe the message content and the required message manipulation and then transform them into the WSMO elements. This thesis is organised as follows:-
Chapter 2: It discusses the literature reviewed on the SWS and interaction mismatches. It begins with some preliminary studies that describe message mismatches in the web service interaction and the data mediation actions that support these interaction mismatches. This is followed by a discussion on the instance transformation component in data mediation. Also in this section the methods and elements that support instance transformation of data mediation in the WSMO framework are discussed.

Chapter 3: This chapter provides a survey on the-state-of-art data mediation approaches that support the interaction mismatches. A comparative evaluation on eight important elements of the message mediator is also presented in this section. The outcome of this survey highlights the need for further research in enhancing the semantic description of the web service messages to support data mediation. The final section presents the architecture and implementation of the data mediation to interaction mismatches in the WSMO framework. This section highlights on the elements of the WSMO that needs to be improved to overcome the existing limitation.

Chapter 4: It describes the research design, procedure and activities which are used in this research. It also discusses on the research instruments, the evaluation criteria, assumptions and limitations that have been adopted and observed in this research.

Chapter 5: presents a conceptual model of the proposed SDMF. It begins with the motivation of this research; summarises the limitations of the existing approaches; and analysis on the required elements to overcome the limitations. This is followed by a detailed discussion on the proposed SDMF. It explains the two main components in the SDMF, DMRO and DMSWS; and the six important procedures involved in modelling this framework.
**Chapter 6:** It explains the design and implementation of the SDMF. The design and implementation of the DMRO and DMSWS components are discussed in detail by elaborating the three important processes namely knowledge extraction; knowledge representation and evaluation; and knowledge modelling into the SWS. All the procedures that are related to these three design and implementation processes are also explained in detail.

**Chapter 7:** It explains the evaluation on the correctness and automation of the proposed SDMF in detail. Firstly a motivating scenario that requires data mediation at runtime between the Bibliographic Scholarly Database (BSD) organisations and the Higher Learning Institutions (HLI) in Malaysia are selected. The instance data mediation component that supports data mediation in this motivating scenario has been built using the proposed SDMF. Secondly, a middleware application that is currently used to support data mediation between the service provider and requestor is analysed and used as the benchmark. The results that are retrieved from the existing middleware application and the DMRO component of the SDMF are verified by domain experts and output from the database. Both the proposed SDMF and the existing middleware applications are measured using precision, recall and the F-measure measurements.

**Chapter 8:** It concludes this dissertation by describing the research achievements and contributions. This is followed by the research summary and suggestions for research future works.
REFERENCES


59. Scharffe, F., Correspondence Patterns Representation, in Faculty of Mathematics, Computer Science and Physics. 2009, University of Innsbruck: Austria. p. 237.
63. Euzenat, J.r.m. and P. Shvaiko, Basic techniques, in Ontology Matching. 2007, Springer Berlin Heidelberg. p. 73-116.
64. Euzenat, J.r.m. and P. Shvaiko, Matching strategies in Ontology Matching. 2007, Springer Berlin Heidelberg. p. 117-150.
66. Nitzsche, J. and B. Norton, Ontology-based data mediation in BPEL (For Semantic Web Services), in Lecture Notes in Business Information


84. Toma, I., B. Bishop, and F. Fischer, Defining the features of the WSMLRule v2.0 language. 2009. p. 32.


