Contents

Foreword ix
Acknowledgements xi
List of Figures xiii

1 Introduction 1
  1.1 Finding a Way Out of the Dilemma 3
  1.2 The Background to this Book 5
  1.3 The Structure of the Book 5
    1.3.1 Data modelling and ontologies 6
    1.3.2 Information integration with relational databases and XML 6
    1.3.3 The show case 6
    1.3.4 Semantic information integration 7
    1.3.5 Data source queries 7
    1.3.6 Generating transformations 7
    1.3.7 Best Practices and Methodologies 8
## CONTENTS

2 Data Modelling and Ontologies 9
   2.1 The Information Integration Problem 10
      2.1.1 How databases view the world 11
      2.1.2 How ontologies view the world 16
      2.1.3 Comparison 20
   2.2 Semantic Information Management 21
      2.2.1 Principles 21
      2.2.2 The methodology 24
   2.3 Conclusions 26

3 Information Integration with Relational Databases and XML 27
   3.1 Introduction 27
      3.1.1 Areas of data integration 28
      3.1.2 Business drivers of data integration 28
      3.1.3 Scope of this chapter 31
   3.2 Relational Database Integration 31
      3.2.1 Integration considerations 31
      3.2.2 Integration approaches/degrees 32
      3.2.3 Data centralization, sharing and federation 33
   3.3 XML-based Integration 42
      3.3.1 XML tools 42
      3.3.2 XML and objects 43
      3.3.3 XML and databases 46
      3.3.4 XML transformations 51
      3.3.5 XML, eCommerce and Web services 53
   3.4 Conclusions 53
      3.4.1 Summary 54
      3.4.2 Variety in data integration 54

4 The Show Case 55
   4.1 Data Sources 56
   4.2 Identifying Overlaps between the Data Sources 58
   4.3 Current Ways of Dealing with Heterogeneity 59
5 Semantic Information Integration 63
  5.1 Approaches in Information Integration 64
  5.2 Mapping Heterogeneous Data Sources 65
    5.2.1 The Unicorn Workbench 67
    5.2.2 Ontology construction and rationalization in the COG project 75
  5.3 Other Methods and Tools 81
    5.3.1 The MOMIS approach 82
    5.3.2 InfoSleuth 83
    5.3.3 OBSERVER 86
    5.3.4 Ontology mapping in the KRAFT project 89
    5.3.5 PROMPT 91
    5.3.6 Chimæra 94
    5.3.7 ONION 96
    5.3.8 Other ontology merging methods 99
  5.4 Comparison of the Methods 100
    5.4.1 Comparison criteria 100
    5.4.2 Comparing the methodologies for semantic schema integration 104
  5.5 Conclusions and Future Work 105
    5.5.1 Limitations of the Unicorn Workbench and future work 106

6 Data Source Queries 107
  6.1 Querying Disparate Data Sources Using the Unicorn Workbench 107
    6.1.1 Queries in the Unicorn Workbench 108
    6.1.2 Transforming conceptual queries into database queries 111
    6.1.3 Limitations of the current approach 113
  6.2 Querying Disparate Data Sources 113
    6.2.1 The querying architecture in the COG project 114
    6.2.2 Querying in the COG showcase 116
    6.2.3 Overcoming the limitations of the Unicorn Workbench 120
CONTENTS

6.3 Related Work 121
  6.3.1 Ontology query languages 121
  6.4 Conclusions 126

7 Generating Transformations 127
  7.1 Information Transformation in the COG Project 128
    7.1.1 Generating transformations with the Unicorn Workbench 129
    7.1.2 Automatic generation of transformations in the COG project 133
  7.2 Other Information Transformation Approaches 135
    7.2.1 Approaches that perform instance transformation 138
    7.2.2 Approaches that do not perform instance transformation 143
  7.3 Conclusions, Limitations and Extensions 146

8 Best Practices and Methodologies Employed 149
  8.1 Best Practices 149
    8.1.1 Selective mapping 150
    8.1.2 Domain vs application modelling 151
    8.1.3 Global-as-view vs local-as-view 154
  8.2 Lessons Learned 155
    8.2.1 Quality of global model depends on local models 155
    8.2.2 Refinement of ontological concepts 156
    8.2.3 Automation is hard to achieve in real-life situations 157
    8.2.4 Queries vs transformations 157
  8.3 Conclusions 158

9 Conclusion 159

References 163

Glossary 173

Index 177