Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage. http://researchspace.auckland.ac.nz/feedback

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.

Note: Masters Theses

The digital copy of a masters thesis is as submitted for examination and contains no corrections. The print copy, usually available in the University Library, may contain corrections made by hand, which have been requested by the supervisor.
XML Semantic Query Optimisation

Ke Geng
June 2011

Supervisor: Gillian C. Dobbie

A thesis submitted in partial fulfillment of the requirements of Doctor of Philosophy in Computer Science
XML Semantic Query Optimisation (XSQO) is a method that optimises execution of queries based on semantic constraints, which are extracted from XML documents. Currently most research into XSQO concentrates on optimisation based on structural constraints in the XML documents. Research, which optimises XML query execution based on semantic constraints, has been limited because of the flexibility of XML.

In this thesis, we introduce a method, which optimises XML query execution based on the constraints on the content of XML documents. In our method, elements are analysed and classified based on the distribution of values of subelements. Information about the classification is extracted and represented in OWL, which is stored in the database together with the XML document. The user input XML query is evaluated and transformed to a new query, which will execute faster and return exactly the same results, based on the element classification information. There are three kinds of transformation that may be carried out in our method: Elimination, which blocks the non-result queries, Reduction, which simplifies the query conditions by removing redundant conditions, and Introduction, which reduces the search area by introducing a new query condition.

Two engines are designed and built for the research. The data analysis engine is designed to analyse the XML documents and classify the specified elements. The query transformation engine evaluates the input XML queries and carries out the query transformation automatically based on the classification information. A case study has been carried out with the data analysis engine and we carried out a series of experiments with the query transformation engine. The results show that: a. XML documents can be analysed and elements can be classified using our method, and the classification results satisfy the requirement of XML query transformation. b. content based XML query transformation can improve XML query execution performance by about 20% to 30%.

In this thesis, we also introduce a data generator, which is designed and built to support the research. With this generator, users can build semantic information into the XML dataset with specified structure, size and selectivity. A case study with the generator shows that the generator satisfies the requirements of content-based XSQO research.
I spent more than 4 years on my doctoral thesis. This is a unforgettable experience in my life. Now it is the time to submit the thesis and it is also the time to thank all the people who have been involved in and supported my research.

First of all, I would like to thank my supervisor, Professor Gillian C. Dobbie. It is my fortune to meet her and work under her supervision. Not only her valuable guidance but also her great personality helped me greatly during my study for both my Masters and Ph.D degrees and changed the difficult days to a wonderful time. I would also like to thank the advisors for my doctoral study at the University of Auckland: Associate Professor Robert Amor and Dr Gerald Weber.

I wish to thank Yun Sing Koh. She did not join our research group very early but she gave me a lot of advice on my thesis and spent her precious time proofreading my thesis.

I am very lucky to be a part of a very active research group. Thanks to Scott Uk-Jin Lee, Daniel Bertinshaw, Muhammad Asif Naeem, Muhammad Shaban Jokhio and Shafiq Alam Burki for their advice and support to my research.

I wish to thank Dr Jing Sun, Dr John Hamer, Dr Hong Yul Yang and all the other people who have helped me.

I would like to express my gratitude towards my parents. They bore me, raised me, supported me, taught me and loved me.

Lastly, and most importantly, I wish to thank my wife and my son. Their support helped me overcome the difficulties to finish my thesis.
Parts of the content in this thesis have been previously published in journals and conferences.

We have previously published sections of Chapter 2 in:


We have published sections of Chapter 3 in:


Some sections of Chapter 6 have been published in:


We have published sections of Chapter 7 in:

## Contents

1 Introduction .......................................................... 1

2 Background ............................................................ 5
   2.1 General information about XML ................................. 6
      2.1.1 XML ......................................................... 6
      2.1.2 XML database system ...................................... 6
      2.1.3 XML query languages ...................................... 7
      2.1.4 Numbering coding schemes of XML ......................... 9
   2.2 XML semantic query optimisation (XSQO) and existing XSQO (including index) .................................................. 11
      2.2.1 Introduction .................................................. 11
      2.2.2 XML Semantic Query optimisation based on structure .... 11
      2.2.3 XML Semantic Query optimisation based on content ..... 14
      2.2.4 Discussion ................................................. 15
   2.3 Data analysis (data mining) and existing XML classification method ................................................................. 16
      2.3.1 Introduction .................................................. 17
      2.3.2 Existing data classification methods ....................... 17
      2.3.3 Discussion on data mining methods ......................... 20
   2.4 Summary ............................................................ 21

3 Our XSQO Method ......................................................... 23
   3.1 Our XSQO method .................................................. 24
   3.2 Data analysis and classification ................................. 26
      3.2.1 Manual data classification .................................. 27
      3.2.2 Automatic data analysis ..................................... 27
   3.3 Challenges in XML element classification ....................... 29
      3.3.1 Size of the XML document .................................. 29
      3.3.2 Same content expressed in different structure ............ 30
      3.3.3 Multiple-value elements .................................... 31
### 3 Solutions for challenges in XML classification

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.4 Missing-value elements</td>
<td>33</td>
</tr>
<tr>
<td>3.3.5 Multiple datatypes</td>
<td>35</td>
</tr>
<tr>
<td>3.3.6 Numbers of classifications</td>
<td>35</td>
</tr>
<tr>
<td><strong>3.4 Solutions for challenges in XML classification</strong></td>
<td></td>
</tr>
<tr>
<td>3.4.1 Solution for large XML documents</td>
<td>37</td>
</tr>
<tr>
<td>3.4.2 Solution for missing-value elements</td>
<td>39</td>
</tr>
<tr>
<td>3.4.3 Solution for multiple-value elements</td>
<td>40</td>
</tr>
<tr>
<td>3.4.4 Solution for multiple datatype values</td>
<td>41</td>
</tr>
<tr>
<td>3.4.5 Solution for same content expressed in different structure</td>
<td>41</td>
</tr>
<tr>
<td>3.4.6 Solution for number of classifications</td>
<td>41</td>
</tr>
<tr>
<td><strong>3.5 Classification information extraction</strong></td>
<td>43</td>
</tr>
<tr>
<td><strong>3.6 Classification information representation</strong></td>
<td>46</td>
</tr>
<tr>
<td>3.6.1 TXT</td>
<td>46</td>
</tr>
<tr>
<td>3.6.2 XML</td>
<td>47</td>
</tr>
<tr>
<td>3.6.3 RDF</td>
<td>48</td>
</tr>
<tr>
<td>3.6.4 OWL</td>
<td>48</td>
</tr>
<tr>
<td><strong>3.7 Query transformation</strong></td>
<td>50</td>
</tr>
<tr>
<td>3.7.1 Transformation categories</td>
<td>50</td>
</tr>
<tr>
<td>3.7.2 Elimination transformation</td>
<td>51</td>
</tr>
<tr>
<td>3.7.3 Reduction transformation</td>
<td>52</td>
</tr>
<tr>
<td>3.7.4 Introduction transformation</td>
<td>53</td>
</tr>
<tr>
<td><strong>3.8 Summary</strong></td>
<td>55</td>
</tr>
</tbody>
</table>

### 4 Influence on Query Transformation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Influence of data characteristics</td>
<td>58</td>
</tr>
<tr>
<td>4.1.1 Included</td>
<td>58</td>
</tr>
<tr>
<td>4.1.2 interSected</td>
<td>59</td>
</tr>
<tr>
<td>4.1.3 sOrted</td>
<td>61</td>
</tr>
<tr>
<td><strong>4.2 Influence of operators</strong></td>
<td>62</td>
</tr>
<tr>
<td><strong>4.3 Definitions for query transformation discussion</strong></td>
<td>65</td>
</tr>
<tr>
<td>4.3.1 Base element</td>
<td>65</td>
</tr>
<tr>
<td>4.3.2 Explanation about sOrted</td>
<td>67</td>
</tr>
<tr>
<td>4.3.3 Included of single value group</td>
<td>67</td>
</tr>
<tr>
<td>4.3.4 Query transformations discussed in this section</td>
<td>67</td>
</tr>
<tr>
<td>4.3.5 Explanations for the discussion of possible transformations</td>
<td>68</td>
</tr>
<tr>
<td><strong>4.4 UnIncluding, UnIntersect, UnSorted</strong></td>
<td>72</td>
</tr>
<tr>
<td>4.4.1 “= and =”</td>
<td>73</td>
</tr>
<tr>
<td>4.4.2 “= and &lt;”</td>
<td>73</td>
</tr>
</tbody>
</table>
5 Data Analysis Engine

5.1 Functionality ................................................. 92
5.2 The structure of the data analysis engine ....................... 93
5.3 Pre-analysis module ........................................ 95
  5.3.1 Receiving user parameters ............................... 95
  5.3.2 Assigning IDs for future operations ................... 96
  5.3.3 Multiple-value element splitting ......................... 99
  5.3.4 Loading values into the analysis engine ................. 100
  5.3.5 Missing-value element identification .................. 101
5.4 Data-analysis module ....................................... 102
5.5 Post-analysis module ...................................... 102
  5.5.1 Multiple-value element combining ....................... 102
  5.5.2 Missing-value element inserting ......................... 103
  5.5.3 Result analysis and OWL generation .................... 104
5.6 Summary .................................................. 105

6 Query Transformation Engine .................................. 107

6.1 Functionalities of the transformation engine .................... 108
  6.1.1 Information management ................................ 108
  6.1.2 Query transformation .................................. 109
6.2 Structure of the transformation engine .......................... 109
6.3 Information extraction module ................................ 110
6.4 Information management module ................................ 112
6.5 Blocking unsatisfied query module ................................... 113
6.6 Reducing condition module .................................. 116
6.7 Introducing condition module ................................ 118
6.8 Summary .................................................. 120
9 Conclusion and Future Work 163
  9.1 Conclusion ........................................ 163
  9.2 Future work ....................................... 165
  9.3 Summary ........................................... 167

References 169

A Query Transformations of Dataset with Different Characteristics 179
  A.1 UnIncluding-UnIntersect-UnSorted .......................... 179
    A.1.1 “= and =” ........................................ 180
    A.1.2 “= and <” ......................................... 182
    A.1.3 “= and >” ......................................... 183
    A.1.4 “< and =” ......................................... 183
    A.1.5 “> and =” ......................................... 186
    A.1.6 “> and >” ......................................... 188
    A.1.7 “< and <” ......................................... 191
    A.1.8 “> and <” ......................................... 191
    A.1.9 “< and >” ......................................... 191
  A.2 UnIncluding-UnIntersect_Sorted (increase) .................. 194
    A.2.1 “= and =” ........................................ 194
    A.2.2 “= and <” ......................................... 197
    A.2.3 “= and >” ......................................... 197
    A.2.4 “< and =” ......................................... 199
    A.2.5 “> and =” ......................................... 199
    A.2.6 “> and >” ......................................... 202
    A.2.7 “< and <” ......................................... 205
    A.2.8 “> and <” ......................................... 205
    A.2.9 “< and >” ......................................... 207
  A.3 UnIncluding-UnIntersect_Sorted (decrease) .................. 207
    A.3.1 “= and =” ........................................ 210
    A.3.2 “= and <” ......................................... 210
    A.3.3 “= and >” ......................................... 212
    A.3.4 “< and =” ......................................... 214
    A.3.5 “> and =” ......................................... 214
    A.3.6 “> and >” ......................................... 217
    A.3.7 “< and <” ......................................... 219
    A.3.8 “> and <” ......................................... 219
    A.3.9 “< and >” ......................................... 222
  A.4 UnIncluding_Intersect_UnSorted ............................ 225
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.4.1</td>
<td>“= and =”</td>
<td>225</td>
</tr>
<tr>
<td>A.4.2</td>
<td>“= and &lt;”</td>
<td>226</td>
</tr>
<tr>
<td>A.4.3</td>
<td>“= and &gt;”</td>
<td>229</td>
</tr>
<tr>
<td>A.4.4</td>
<td>“&lt; and =”</td>
<td>229</td>
</tr>
<tr>
<td>A.4.5</td>
<td>“&gt; and =”</td>
<td>231</td>
</tr>
<tr>
<td>A.4.6</td>
<td>“&gt; and &gt;”</td>
<td>233</td>
</tr>
<tr>
<td>A.4.7</td>
<td>“&lt; and &lt;”</td>
<td>236</td>
</tr>
<tr>
<td>A.4.8</td>
<td>“&gt; and &lt;”</td>
<td>236</td>
</tr>
<tr>
<td>A.4.9</td>
<td>“&lt; and &gt;”</td>
<td>236</td>
</tr>
<tr>
<td>A.5</td>
<td>UnIncluding Intersect Sorted (increase)</td>
<td>240</td>
</tr>
<tr>
<td>A.5.1</td>
<td>“= and =”</td>
<td>240</td>
</tr>
<tr>
<td>A.5.2</td>
<td>“= and &lt;”</td>
<td>241</td>
</tr>
<tr>
<td>A.5.3</td>
<td>“= and &gt;”</td>
<td>241</td>
</tr>
<tr>
<td>A.5.4</td>
<td>“&lt; and =”</td>
<td>244</td>
</tr>
<tr>
<td>A.5.5</td>
<td>“&gt; and =”</td>
<td>246</td>
</tr>
<tr>
<td>A.5.6</td>
<td>“&gt; and &gt;”</td>
<td>246</td>
</tr>
<tr>
<td>A.5.7</td>
<td>“&lt; and &lt;”</td>
<td>249</td>
</tr>
<tr>
<td>A.5.8</td>
<td>“&gt; and &lt;”</td>
<td>249</td>
</tr>
<tr>
<td>A.5.9</td>
<td>“&lt; and &gt;”</td>
<td>252</td>
</tr>
<tr>
<td>A.6</td>
<td>UnIncluding Intersect Sorted (decrease)</td>
<td>255</td>
</tr>
<tr>
<td>A.6.1</td>
<td>“= and =”</td>
<td>255</td>
</tr>
<tr>
<td>A.6.2</td>
<td>“= and &lt;”</td>
<td>256</td>
</tr>
<tr>
<td>A.6.3</td>
<td>“= and &gt;”</td>
<td>256</td>
</tr>
<tr>
<td>A.6.4</td>
<td>“&lt; and =”</td>
<td>259</td>
</tr>
<tr>
<td>A.6.5</td>
<td>“&gt; and =”</td>
<td>261</td>
</tr>
<tr>
<td>A.6.6</td>
<td>“&gt; and &gt;”</td>
<td>261</td>
</tr>
<tr>
<td>A.6.7</td>
<td>“&lt; and &lt;”</td>
<td>264</td>
</tr>
<tr>
<td>A.6.8</td>
<td>“&gt; and &lt;”</td>
<td>267</td>
</tr>
<tr>
<td>A.6.9</td>
<td>“&lt; and &gt;”</td>
<td>267</td>
</tr>
<tr>
<td>A.7</td>
<td>Including UnIntersect UnSorted</td>
<td>270</td>
</tr>
<tr>
<td>A.7.1</td>
<td>“= and =”</td>
<td>270</td>
</tr>
<tr>
<td>A.7.2</td>
<td>“= and &lt;”</td>
<td>271</td>
</tr>
<tr>
<td>A.7.3</td>
<td>“= and &gt;”</td>
<td>274</td>
</tr>
<tr>
<td>A.7.4</td>
<td>“&lt; and =”</td>
<td>274</td>
</tr>
<tr>
<td>A.7.5</td>
<td>“&gt; and =”</td>
<td>277</td>
</tr>
<tr>
<td>A.7.6</td>
<td>“&gt; and &gt;”</td>
<td>277</td>
</tr>
<tr>
<td>A.7.7</td>
<td>“&lt; and &lt;”</td>
<td>277</td>
</tr>
<tr>
<td>A.7.8</td>
<td>“&gt; and &lt;”</td>
<td>281</td>
</tr>
</tbody>
</table>
A.7.9 “< and >” ................................................. 281
A.8 Including UnIntersect_Sorted(increase) ............... 281
  A.8.1 “= and =” ........................................... 284
  A.8.2 “= and <” ........................................... 284
  A.8.3 “= and >” ........................................... 286
  A.8.4 “< and =” ........................................... 288
  A.8.5 “> and =” ........................................... 288
  A.8.6 “> and >” ........................................... 292
  A.8.7 “< and <” ........................................... 292
  A.8.8 “> and <” ........................................... 295
  A.8.9 “< and >” ........................................... 295
A.9 Including UnIntersect_Sorted(decrease) ............. 297
  A.9.1 “= and =” ........................................... 299
  A.9.2 “= and <” ........................................... 299
  A.9.3 “= and >” ........................................... 301
  A.9.4 “< and =” ........................................... 301
  A.9.5 “> and =” ........................................... 304
  A.9.6 “> and >” ........................................... 304
  A.9.7 “< and <” ........................................... 307
  A.9.8 “> and <” ........................................... 310
  A.9.9 “< and >” ........................................... 310
A.10 Including Intersect_UnSorted .......................... 313
  A.10.1 “= and =” ........................................... 313
  A.10.2 “= and <” ........................................... 314
  A.10.3 “= and >” ........................................... 317
  A.10.4 “< and =” ........................................... 317
  A.10.5 “> and =” ........................................... 320
  A.10.6 “> and >” ........................................... 320
  A.10.7 “< and <” ........................................... 320
  A.10.8 “> and <” ........................................... 324
  A.10.9 “< and >” ........................................... 324
A.11 Including Intersect_Sorted(increase) ............... 324
  A.11.1 “= and =” ........................................... 327
  A.11.2 “= and <” ........................................... 327
  A.11.3 “= and >” ........................................... 329
  A.11.4 “< and =” ........................................... 331
  A.11.5 “> and =” ........................................... 331
  A.11.6 “> and >” ........................................... 335
A.11.7 “< and <” .................................................. 335
A.11.8 “> and <” .................................................. 338
A.11.9 “< and >” .................................................. 338
A.12 Including_Intersection_Sorted(decrease) .................. 340
  A.12.1 “= and =” ............................................ 342
  A.12.2 “= and <” ............................................ 342
  A.12.3 “= and >” ............................................ 344
  A.12.4 “< and =” ............................................ 344
  A.12.5 “> and =” ............................................ 347
  A.12.6 “> and >” ............................................ 347
  A.12.7 “< and <” ............................................ 350
  A.12.8 “> and <” ............................................ 353
  A.12.9 “< and >” ............................................ 353
## List of Figures

2.1 The structure of University .......................... 13  
2.2 Example of Pattern Tree .......................... 15  
2.3 Characteristics of each optimisation method ............ 16  
2.4 Analysis procedure of method ......................... 18  
2.5 The Sequence transformation ......................... 19  
2.6 Characteristics of each data classification method ...... 20  
3.1 The staff.XML ........................................ 25  
3.2 The XML document with record that breaks the rules .... 26  
3.3 The XML document with disturbing value distribution .... 28  
3.4 Same content in different structure 1 .................. 30  
3.5 Same content in different structure 2 .................. 31  
3.6 Multiple-value elements example ....................... 31  
3.7 Multiple-value example with extra values ignored ....... 33  
3.8 Missing-value elements ................................ 33  
3.9 Two structures of missing-value elements ............... 34  
3.10 XML document of people records ....................... 36  
3.11 Example for data size reducing ....................... 38  
3.12 XML document of mail ................................ 40  
3.13 The treatment for multiple-value elements ............. 40  
3.14 Procedure of K-Means .............................. 42  
3.15 Various classifications on same XML document ......... 43  
3.16 The RDF description for the website ................. 49  
3.17 The OWL description for class “WhiteWine” ............ 50  
4.1 Possible transformations of data set 000 and operators “= and =” .... 74  
4.2 Possible transformations of data set 000 and operators “= and <” .... 76  
4.3 Possible transformations of data set 000 and operators “= and >” .... 78  
4.4 Possible transformations of data set 000 and operators “< and =” .... 80  
4.5 Possible transformations of data set 000 and operators “> and =” .... 82
4.6 Possible transformations of data set 000 and operators “> and >” .... 83
4.7 Possible transformations of data set 000 and operators “< and <” ... 85
4.8 Possible transformations of data set 000 and operators “> and <” ... 86
4.9 Possible transformations of data set 000 and operators “< and >” ... 87
5.1 Data analysis engine structure ............................................. 94
5.2 Parameters collected from the user .................................. 96
5.3 The flexible k-ary index system ...................................... 98
5.4 The new index system of eXist ...................................... 99
5.5 Data structure of “unit” ............................................. 100
5.6 Multiple-value element “people” and the split result .......... 103
5.7 A piece of OWL document for element “person” ............. 105
6.1 The structure for the query transformation engine ........... 110
6.2 Schema for element “age” ........................................... 111
6.3 The descriptions in OWL .......................................... 112
6.4 Algorithm for “Blocking unsatisfied query” .................... 116
6.5 Algorithm for “Finding query related groups” .................. 117
6.6 Algorithm for “Reducing query condition” ..................... 119
6.7 Algorithm for “Introducing query condition” ................... 120
7.1 The XBench datasets and examples ................................. 125
7.2 Possible modified structure .......................................... 140
7.3 The input interface of Insert multiple elements ................ 141
7.4 The structure of original XML document ......................... 142
7.5 The possible target elements ................................... 142
7.6 The positions for elements “geograph” .......................... 143
7.7 The positions for elements “maths” .............................. 143
7.8 The final structure .................................................. 144
8.1 The DTD of element “person” ........................................ 149
8.2 The generated XML document ..................................... 150
8.3 The selectivity values ............................................... 151
8.4 The values of inserted elements .................................. 152
8.5 A piece of analysis result of structural data set ................ 152
8.6 A piece of analysis result of data set with multiple-value elements ... 153
8.7 A piece of analysis result of data set with missing-value elements ... 154
8.8 The values of inserted elements .................................. 154
8.9 Queries for the experiments of Elimination .................... 155
8.10 Queries for the experiments of Reduction ...................... 155
8.11 Queries for the experiments of Introduction ........................................ 156
8.12 The results of Elimination ................................................................. 157
8.13 The performance of “Reduction selectivity=20%” ................................ 157
8.14 The performance of “Reduction selectivity=40%” ............................... 158
8.15 The performance of “Reduction selectivity=60%” ................................ 158
8.16 The performance of “Reduction selectivity=80%” ................................ 159
8.17 The performance of “Reduction selectivity=100%” ............................... 159
8.18 The performance of “Introduction selectivity=20%” ............................ 160
8.19 The performance of “Introduction selectivity=40%” ............................ 160
8.20 The performance of “Introduction selectivity=60%” ............................ 160
8.21 The performance of “Introduction selectivity=80%” ............................ 161
8.22 The performance of “Introduction selectivity=100%” .......................... 161
A.1 Possible transformations of data set 000 and operators “= and =” ............ 181
A.2 Possible transformations of data set 000 and operators “= and <” ............ 184
A.3 Possible transformations of data set 000 and operators “= and >” ............ 185
A.4 Possible transformations of data set 000 and operators “< and =” ............ 187
A.5 Possible transformations of data set 000 and operators “> and =” ............ 189
A.6 Possible transformations of data set 000 and operators “> and >” ............ 190
A.7 Possible transformations of data set 000 and operators “< and <” ............ 192
A.8 Possible transformations of data set 000 and operators “> and <” ............ 193
A.9 Possible transformations of data set 000 and operators “< and >” ............ 195
A.10 Possible transformations of data set 001 (increase) and operators “= and =” 196
A.11 Possible transformations of data set 001 (increase) and operators “= and <” 198
A.12 Possible transformations of data set 001 (increase) and operators “= and >” 200
A.13 Possible transformations of data set 001 (increase) and operators “< and =” 201
A.14 Possible transformations of data set 001 (increase) and operators “> and =” 203
A.15 Possible transformations of data set 001 (increase) and operators “> and >” 204
A.16 Possible transformations of data set 001 (increase) and operators “< and <” 206
A.17 Possible transformations of data set 001 (increase) and operators “> and <” 208
A.18 Possible transformations of data set 001 (increase) and operators “< and >” 209
A.19 Possible transformations of data set 001 (decrease) and operators “= and =” 211
A.20 Possible transformations of data set 001 (decrease) and operators “= and <” 213
A.21 Possible transformations of data set 001 (decrease) and operators “= and >” 215
A.22 Possible transformations of data set 001 (decrease) and operators “< and =” 216
A.23 Possible transformations of data set 001 (decrease) and operators “> and =” 218
A.24 Possible transformations of data set 001 (decrease) and operators “> and >” 220
A.25 Possible transformations of data set 001 (decrease) and operators “< and <” 221
A.26 Possible transformations of data set 001 (decrease) and operators “> and <” 223
A.27 Possible transformations of data set 001 (decrease) and operators “< and >” 224
A.28 Possible transformations of data set 010 and operators “= and =” . . . . . 227
A.29 Possible transformations of data set 010 and operators “= and <” . . . . . 228
A.30 Possible transformations of data set 010 and operators “= and >” . . . . . 230
A.31 Possible transformations of data set 010 and operators “< and =” . . . . . 232
A.32 Possible transformations of data set 010 and operators “> and =” . . . . . 234
A.33 Possible transformations of data set 010 and operators “> and >” . . . . . 235
A.34 Possible transformations of data set 010 and operators “< and <” . . . . . 237
A.35 Possible transformations of data set 010 and operators “> and <” . . . . . 238
A.36 Possible transformations of data set 010 and operators “< and >” . . . . . 239
A.37 Possible transformations of data set 011(increase) and operators “= and =” 242
A.38 Possible transformations of data set 011(increase) and operators “= and <” 243
A.39 Possible transformations of data set 011(increase) and operators “= and >” 245
A.40 Possible transformations of data set 011(increase) and operators “< and =” 247
A.41 Possible transformations of data set 011(increase) and operators “> and =” 248
A.42 Possible transformations of data set 011(increase) and operators “> and >” 250
A.43 Possible transformations of data set 011(increase) and operators “< and <” 251
A.44 Possible transformations of data set(increase) 011 and operators “> and <” 253
A.45 Possible transformations of data set 011(increase) and operators “< and >” 254
A.46 Possible transformations of data set 011(decrease) and operators “= and =” 257
A.47 Possible transformations of data set 011(decrease) and operators “= and <” 258
A.48 Possible transformations of data set 011(decrease) and operators “= and >” 260
A.49 Possible transformations of data set 011(decrease) and operators “< and =” 262
A.50 Possible transformations of data set 011(decrease) and operators “> and =” 263
A.51 Possible transformations of data set 011(decrease) and operators “> and >” 265
A.52 Possible transformations of data set 011(decrease) and operators “< and <” 266
A.53 Possible transformations of data set 011(decrease) and operators “> and <” 268
A.54 Possible transformations of data set 011(decrease)) and operators “< and >” 269
A.55 Possible transformations of data set 100 and operators “= and =” . . . . . 272
A.56 Possible transformations of data set 100 and operators “= and <” . . . . . 273
A.57 Possible transformations of data set 100 and operators “= and >” . . . . . 275
A.58 Possible transformations of data set 100 and operators “< and =” . . . . . 276
A.59 Possible transformations of data set 100 and operators “> and =” . . . . . 278
A.60 Possible transformations of data set 100 and operators “> and >” . . . . . 279
A.61 Possible transformations of data set 100 and operators “< and <” . . . . . 280
A.62 Possible transformations of data set 100 and operators “> and <” . . . . . 282
A.63 Possible transformations of data set 100 and operators “< and >” . . . . . 283
A.64 Possible transformations of data set 101 (increase) and operators “= and =” 285
A.65 Possible transformations of data set 101 (increase) and operators “= and <” 287
A.66 Possible transformations of data set 101 (increase) and operators “= and >” 289
A.67 Possible transformations of data set 101 (increase) and operators “< and =” 290
A.68 Possible transformations of data set 101 (increase) and operators “> and =” 291
A.69 Possible transformations of data set 101 (increase) and operators “> and >” 293
A.70 Possible transformations of data set 101 (increase) and operators “< and <” 294
A.71 Possible transformations of data set 101 (increase) and operators “> and <” 296
A.72 Possible transformations of data set 101 (increase) and operators “< and >” 298
A.73 Possible transformations of data set 101 (decrease) and operators “= and =” 300
A.74 Possible transformations of data set 101 (decrease) and operators “= and <” 302
A.75 Possible transformations of data set 101 (decrease) and operators “= and >” 303
A.76 Possible transformations of data set 101 (decrease) and operators “< and =” 305
A.77 Possible transformations of data set 101 (decrease) and operators “> and =” 306
A.78 Possible transformations of data set 101 (decrease) and operators “> and >” 308
A.79 Possible transformations of data set 101 (decrease) and operators “< and <” 309
A.80 Possible transformations of data set 101 (decrease) and operators “> and <” 311
A.81 Possible transformations of data set 101 (decrease) and operators “< and >” 312
A.82 Possible transformations of data set 110 and operators “= and =” . . . . . . 315
A.83 Possible transformations of data set 110 and operators “= and <” . . . . . . 316
A.84 Possible transformations of data set 110 and operators “= and >” . . . . . . 318
A.85 Possible transformations of data set 110 and operators “< and =” . . . . . . 319
A.86 Possible transformations of data set 110 and operators “> and =” . . . . . . 321
A.87 Possible transformations of data set 110 and operators “> and >” . . . . . . 322
A.88 Possible transformations of data set 110 and operators “< and <” . . . . . . 323
A.89 Possible transformations of data set 110 and operators “> and <” . . . . . . 325
A.90 Possible transformations of data set 110 and operators “< and >” . . . . . . 326
A.91 Possible transformations of data set 111(increase) and operators “= and =” 328
A.92 Possible transformations of data set 111(increase) and operators “= and <” 330
A.93 Possible transformations of data set 111(increase) and operators “= and >” 332
A.94 Possible transformations of data set 111(increase) and operators “< and =” 333
A.95 Possible transformations of data set 111(increase) and operators “> and =” 334
A.96 Possible transformations of data set 111(increase) and operators “> and >” 336
A.97 Possible transformations of data set 111(increase) and operators “< and <” 337
A.98 Possible transformations of data set 111(increase) and operators “> and <” 339
A.99 Possible transformations of data set 111(increase) and operators “< and >” 341
A.100 Possible transformations of data set 111(decrease) and operators “= and =” 343
A.101 Possible transformations of data set 111(decrease) and operators “= and <” 345
A.102 Possible transformations of data set 111(decrease) and operators “= and >” 346
A.103 Possible transformations of data set 111 (decrease) and operators “< and =” 348
A.104 Possible transformations of data set 111 (decrease) and operators “> and =” 349
A.105 Possible transformations of data set 111 (decrease) and operators “> and >” 351
A.106 Possible transformations of data set 111 (decrease) and operators “< and <” 352
A.107 Possible transformations of data set 111 (decrease) and operators “> and <” 354
A.108 Possible transformations of data set 111 (decrease) and operators “< and >” 355
# List of Tables

3.1 Unsatisfied dataset classification ........................................ 28
3.2 Ideal dataset classification ............................................. 29
3.3 Classification with missing-value elements ................................. 40

4.1 Dataset of Included .......................................................... 58
4.2 Dataset of unIncluded ......................................................... 59
4.3 Dataset of interSected ......................................................... 60
4.4 Dataset of uninterSected .................................................... 60
4.5 Dataset of sOrted ............................................................... 61
4.6 Dataset of unsOrted ........................................................... 61
4.7 Record of members ............................................................. 62
4.8 Record of members ............................................................. 64
4.9 Dataset of Unsorted (arranged based on “age”) .......................... 65
4.10 Dataset of Unsorted (no base element) .................................... 66
4.11 Dataset of sOrted ............................................................... 67
4.12 Dataset of Included .......................................................... 68
4.13 Classification characteristics and grouping codes ....................... 69
4.14 Operators combinations and examples ................................. 69
4.15 Classification with grouping code “001” ............................... 70
4.16 Classification with grouping code “101” ............................... 70
4.17 Query conditions and groups .............................................. 71
4.18 Dataset of 000 ................................................................. 73

7.1 Attributes in BaseType ........................................................ 128

A.1 Dataset of 000 ................................................................. 180
A.2 Dataset of 001_Increase ......................................................... 194
A.3 Dataset of 001_decrease ...................................................... 210
A.4 Dataset of 010 ................................................................. 225
A.5 Dataset of 011_increase ....................................................... 240
A.6  Dataset of 011_decrease .................................................. 255
A.7  Dataset of 100 ................................................................. 270
A.8  Dataset of 101_increase .................................................. 284
A.9  Dataset of 101_decrease .................................................. 297
A.10 Dataset of 110 ............................................................... 313
A.11 Dataset of 111_increase ................................................... 327
A.12 Dataset of 111_decrease ................................................... 340