# Contents

1. **Introduction: Qualitative Data Analysis in a Digital World** 1  
   1.1. The Emergence of “Digital Humanities” .... 3  
   1.2. Digital Text and Social Science Research .... 8  
   1.3. Example Study: Research Question and Data Set ... 11  
      1.3.1. Democratic Demarcation ............ 12  
      1.3.2. Data Set ...................... 12  
   1.4. Contributions and Structure of the Study .... 14

2. **Computer-Assisted Text Analysis in the Social Sciences** 17  
   2.1. Text as Data between Quality and Quantity .... 17  
   2.2. Text as Data for Natural Language Processing .... 22  
      2.2.1. Modeling Semantics ............... 22  
      2.2.2. Linguistic Preprocessing .......... 26  
      2.2.3. Text Mining Applications .......... 28  
   2.3. Types of Computational Qualitative Data Analysis ... 34  
      2.3.1. Computational Content Analysis .... 40  
      2.3.2. Computer-Assisted Qualitative Data Analysis . 43  
      2.3.3. Lexicometrics for Corpus Exploration .... 45  
      2.3.4. Machine Learning ................ 49

3. **Integrating Text Mining Applications for Complex Analysis** 55  
   3.1. Document Retrieval ........................ 56  
      3.1.1. Requirements ..................... 56  
      3.1.2. Key Term Extraction ................ 59  
      3.1.3. Retrieval with Dictionaries .......... 66  
      3.1.4. Contextualizing Dictionaries ........ 69  
      3.1.5. Scoring Co-Occurrences ............ 71  
      3.1.6. Evaluation ........................ 74
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.7. Summary of Lessons Learned</td>
<td>82</td>
</tr>
<tr>
<td>3.2. Corpus Exploration</td>
<td>84</td>
</tr>
<tr>
<td>3.2.1. Requirements</td>
<td>85</td>
</tr>
<tr>
<td>3.2.2. Identification and Evaluation of Topics</td>
<td>88</td>
</tr>
<tr>
<td>3.2.3. Clustering of Time Periods</td>
<td>100</td>
</tr>
<tr>
<td>3.2.4. Selection of Topics</td>
<td>105</td>
</tr>
<tr>
<td>3.2.5. Term Co-Occurrences</td>
<td>108</td>
</tr>
<tr>
<td>3.2.6. Keyness of Terms</td>
<td>112</td>
</tr>
<tr>
<td>3.2.7. Sentiments of Key Terms</td>
<td>112</td>
</tr>
<tr>
<td>3.2.8. Semantically Enriched Co-Occurrence Graphs</td>
<td>115</td>
</tr>
<tr>
<td>3.2.9. Summary of Lessons Learned</td>
<td>122</td>
</tr>
<tr>
<td>3.3. Classification for Qualitative Data Analysis</td>
<td>125</td>
</tr>
<tr>
<td>3.3.1. Requirements</td>
<td>128</td>
</tr>
<tr>
<td>3.3.2. Experimental Data</td>
<td>132</td>
</tr>
<tr>
<td>3.3.3. Individual Classification</td>
<td>135</td>
</tr>
<tr>
<td>3.3.4. Training Set Size and Semantic Smoothing</td>
<td>140</td>
</tr>
<tr>
<td>3.3.5. Classification for Proportions and Trends</td>
<td>146</td>
</tr>
<tr>
<td>3.3.6. Active Learning</td>
<td>155</td>
</tr>
<tr>
<td>3.3.7. Summary of Lessons Learned</td>
<td>165</td>
</tr>
<tr>
<td>4. Exemplary Study: Democratic Demarcation in Germany</td>
<td>167</td>
</tr>
<tr>
<td>4.1. Democratic Demarcation</td>
<td>167</td>
</tr>
<tr>
<td>4.2. Exploration</td>
<td>174</td>
</tr>
<tr>
<td>4.2.1. Democratic Demarcation from 1950–1956</td>
<td>175</td>
</tr>
<tr>
<td>4.2.2. Democratic Demarcation from 1957–1970</td>
<td>178</td>
</tr>
<tr>
<td>4.2.3. Democratic Demarcation from 1971–1988</td>
<td>180</td>
</tr>
<tr>
<td>4.2.4. Democratic Demarcation from 1989–2000</td>
<td>183</td>
</tr>
<tr>
<td>4.2.5. Democratic Demarcation from 2001–2011</td>
<td>185</td>
</tr>
<tr>
<td>4.3. Classification of Demarcation Statements</td>
<td>187</td>
</tr>
<tr>
<td>4.3.1. Category System</td>
<td>188</td>
</tr>
<tr>
<td>4.3.2. Supervised Active Learning of Categories</td>
<td>192</td>
</tr>
<tr>
<td>4.3.3. Category Trends and Co-Occurrences</td>
<td>195</td>
</tr>
<tr>
<td>4.4. Conclusions and Further Analyses</td>
<td>209</td>
</tr>
</tbody>
</table>
Contents

5. V-TM – A Methodological Framework for Social Science 213
   5.1. Requirements ........................................... 216
       5.1.1. Data Management ................................. 219
       5.1.2. Goals of Analysis ............................... 220
   5.2. Workflow Design ........................................ 223
       5.2.1. Overview ........................................... 224
       5.2.2. Workflows ......................................... 228
   5.3. Result Integration and Documentation .................. 238
       5.3.1. Integration ......................................... 239
       5.3.2. Documentation .................................... 241
   5.4. Methodological Integration .............................. 243

6. Summary: Qualitative and Computational Text Analysis 251
   6.1. Meeting Requirements ................................... 252
   6.2. Exemplary Study ....................................... 255
   6.3. Methodological Systematization ......................... 256
   6.4. Further Developments .................................. 257

A. Data Tables, Graphs and Algorithms 261

Bibliography 271
Text Mining for Qualitative Data Analysis in the Social Sciences
A Study on Democratic Discourse in Germany
Wiedemann, G.
2016, XVII, 294 p. 24 illus., Softcover
ISBN: 978-3-658-15308-3