

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

3.1.7.	Summary of Lessons Learned	82
3.2.	Corpus Exploration	84
3.2.1.	Requirements	85
3.2.2.	Identification and Evaluation of Topics	88
3.2.3.	Clustering of Time Periods	100
3.2.4.	Selection of Topics	105
3.2.5.	Term Co-Occurrences	108
3.2.6.	Keyness of Terms	112
3.2.7.	Sentiments of Key Terms	112
3.2.8.	Semantically Enriched Co-Occurrence Graphs	115
3.2.9.	Summary of Lessons Learned	122
3.3.	Classification for Qualitative Data Analysis	125
3.3.1.	Requirements	128
3.3.2.	Experimental Data	132
3.3.3.	Individual Classification	135
3.3.4.	Training Set Size and Semantic Smoothing	140
3.3.5.	Classification for Proportions and Trends	146
3.3.6.	Active Learning	155
3.3.7.	Summary of Lessons Learned	165
4.	Exemplary Study: Democratic Demarcation in Germany	167
4.1.	Democratic Demarcation	167
4.2.	Exploration	174
4.2.1.	Democratic Demarcation from 1950–1956	175
4.2.2.	Democratic Demarcation from 1957–1970	178
4.2.3.	Democratic Demarcation from 1971–1988	180
4.2.4.	Democratic Demarcation from 1989–2000	183
4.2.5.	Democratic Demarcation from 2001–2011	185
4.3.	Classification of Demarcation Statements	187
4.3.1.	Category System	188
4.3.2.	Supervised Active Learning of Categories	192
4.3.3.	Category Trends and Co-Occurrences	195
4.4.	Conclusions and Further Analyses	209

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