

CONTENTS

1	Introduction	1
1.1	<i>The Coronavirus Pandemic on Social Media</i>	3
1.2	<i>Related Research</i>	6
	<i>References</i>	14
2	COVID-19 Corpora	19
2.1	<i>CORD-19: The COVID-19 Open Research Dataset</i>	19
2.2	<i>COVID-19 Twitter Chatter Dataset for Open Scientific Research</i>	20
2.3	<i>The Coronavirus Corpus</i>	22
2.4	<i>Parallel Corpora</i>	23
2.5	<i>GeoCoV19</i>	23
2.6	<i>Chen et al.'s Coronavirus Twitter Corpus (CCTC)</i>	24
	<i>References</i>	27
3	Managing Large Twitter Datasets	31
3.1	<i>Twitter Content</i>	33
3.2	<i>Downloading and Managing a Large Twitter Corpus</i>	35
3.2.1	<i>Anatomy of a Tweet</i>	35
3.2.2	<i>Downloading and Extracting Data</i>	37
3.2.3	<i>Data Organization and File Format Selection</i>	40
3.3	<i>Data Sampling</i>	41
3.4	<i>Extracting Geotagged Tweets</i>	47

3.5	<i>Subcorpora. Using Metadata with XML-Aware Corpus Tools</i>	52
	<i>References</i>	55
4	Keywords	59
4.1	<i>The Concept of “Keyword” in Corpus Linguistics</i>	61
4.1.1	<i>Experiment: The Keywords of Keywords</i>	67
4.2	<i>Keyword Extraction Methods in Natural Language Processing</i>	73
4.2.1	<i>Machine Learning Approaches</i>	73
4.2.2	<i>Unsupervised Approaches</i>	74
4.2.3	<i>Graph-Based Approaches</i>	83
4.3	<i>Comparing Keyword Sets</i>	86
4.4	<i>Keyword Extraction Using Word Embeddings</i>	95
4.4.1	<i>Experiment: Comparing Keywords from Two Countries Using KeyBERT</i>	96
	<i>References</i>	99
5	Topics	103
5.1	<i>“Traditional” Topic Modelling Methods</i>	106
5.1.1	<i>Experiment: LDA vs NMF for Topic Modelling</i>	108
5.2	<i>Embeddings-Based Topic Modelling</i>	115
5.2.1	<i>Experiment: Extracting COVID-19 Topics Using BERTopic</i>	118
5.3	<i>Dynamic Topic Modelling</i>	132
	<i>References</i>	136
6	Sentiment	141
6.1	<i>Sentiment Analysis Methods</i>	142
6.1.1	<i>Deterministic Methods</i>	144
6.1.2	<i>Probabilistic Methods</i>	145
6.2	<i>Experiment: Sentiment Analysis of the CCTC by Country</i>	146
6.2.1	<i>Tweet Classification and Sentiment Over Time</i>	147
6.2.2	<i>The Sentiment Lexicon of the Pandemic on Twitter</i>	152
6.3	<i>The Role of Emojis in the Expression of Sentiment</i>	160
	<i>References</i>	164

7	Hashtags	169
	7.1 <i>Hashtags in the CCTC</i>	171
	<i>References</i>	179
8	Lessons Learned and Key Takeaways	181
	Index	187