POSTPRINT

Kostiantyn Kucher

Sentiment and Stance Visualization of Textual Data for Social Media

Doctoral Dissertation

Computer Science

2019

Linnæus University

Postprint version with active hyperlinks

The officially published version of this dissertation is available at
http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-81081

To family and friends, both present and absent ones

Abstract

Rapid progress in digital technologies has transformed the world in many ways during the past few decades, in particular, with the new means of communication such as social media. Social media platforms typically rely on textual data produced or shared by the users in multiple timestamped posts. Analyses of such data are challenging for traditional manual methods that are unable to scale up to the volume and the variety of the data. While computational methods can partially address these challenges, they have to be used together with the methods developed within information visualization and visual analytics to gain knowledge from the text data by using interactive visual representations.

One of the most interesting aspects of text data is related to expressions of sentiments and opinions. The corresponding task of sentiment analysis has been studied within computational linguistics, and sentiment visualization techniques exist as well. However, there are gaps in research on the related task of stance analysis, dedicated to subjectivity that is not expressible only in terms of sentiment. Research on stance is an area of interest in linguistics, but support by computational and visual methods has been limited so far. The challenges related to definition, analysis, and visualization of stance in textual data call for an interdisciplinary research effort. The StaViCTA project addressed these challenges with a focus on written text in English. The corresponding results in the area of visualization are reported in this work, based on multiple publications.

The main goal of this dissertation is to define, categorize, and implement means for visual analysis of sentiment and stance in textual data, in particular, for social media. Our work is based on the theoretical framework and automatic classifier of stance developed by our project collaborators, involving multiple non-exclusive stance categories such as certainty and prediction. We define a design space for sentiment and stance visualization techniques based on literature surveys. We discuss multiple visualization and visual analytics approaches developed by us to facilitate the underlying research on stance analysis, data collection and annotation, and visual analysis of sentiment and stance in real-world text data from several social media sources. The work described in this dissertation was carried out in cooperation with domain experts in linguistics and computational linguistics, and our approaches were validated with case studies, expert user reviews, and critical discussion. The results of this work open up further opportunities for research in text visualization and visual text analytics. The potential application areas are academic research, business intelligence, social media monitoring, and journalism.

Keywords: stance visualization, sentiment visualization, text visualization, stance analysis, sentiment analysis, opinion mining, visualization, interaction, visual analytics, NLP, text mining, text analytics, social media

Acknowledgments

I have been lucky and privileged to be surrounded by a lot of amazing individuals, so I have a long list[*] of people to be grateful to (and for!).

First of all, I would like to thank my main doctoral supervisor Prof. Dr. Andreas Kerren for igniting my interest for information visualization and visual analytics, accepting me as a PhD student in 2013, and supporting me on my way to becoming a researcher. I could not wish for a better supervisor, and I can only hope that I will be able to reach the same level as a scholar in the future.

Of course, I should thank my co-supervisors Prof. Dr. Carita Paradis and Dr. Magnus Sahlgren for teaching me a lot about their respective disciplines and being great collaborators. I am also grateful to all the participants of our interdisciplinary StaViCTA project, especially Dr. Maria Skeppstedt, Dr. Vasiliki Simaki, and Teri Schamp-Bjerede. StaViCTA was funded by the Swedish Research Council (Vetenskapsrådet), who also have my sincere thanks.

I would like to thank my friends and colleagues from the ISOVIS group for helping to make my PhD studies such a great experience, including Angelos Chatzimparmpas, Dr. Rafael M. Martins, Dr. Björn Zimmer, Dr. Ilir Jusufi, and Dr. Daniel Cernea. The same goes for my other friends who are either in the process or have already completed their PhD studies, including Yevhen Ivanenko, Alisa Lincke, Dr. Antonina Danylenko, Oleg Danylenko, Nico Reski, Dr. Muhammad Usman Iftikhar, Dr. Nadeem Abbas, and all the others.

I am also grateful to the rest of my colleagues at Linnæus University in Växjö, especially Dr. Jonas Lundberg, Dr. Anna Wingkvist, Dr. Morgan Ericsson, Dr. Aris Alissandrakis, Dr. Jesper Andersson, Prof. Dr. Welf Löwe, and Prof. Dr. Danny Weyns.

Taking a step back in time, I would like to thank the faculty at Odessa National Polytechnic University in Ukraine, especially Prof. Dr. Svetlana Antoshchuk, Dr. Anatoliy Nikolenko, and Dr. Viktor Boltenkov. I should also thank Galina Bokareva who was my teacher of English during the middle and high school years—without mastering this language, I would have never made it that far.

Next, I'd like to give a shout-out to the rest of my friends in Sweden, Ukraine, and elsewhere, whom I have not already mentioned above. You guys rock!

And finally, special thanks go to my dear *sambo* Kateryna for her love and support. I wish to express my deepest gratitude and appreciation for all of my family and relatives for always being there and believing in me, especially my parents Sergiy and Iryna, my late grandparents Galyna, Vitaliy, Olena, and Mykhailo, and my late cousin First Lieutenant Yevhen Makarian—I hope I would make them all proud.

Once again, thank you all, *дякую*, *спасибо*, *tack så mycket*, and *vielen Dank*!

[*] I considered including a diagram instead of the text here, but there will be plenty of those in the rest of this dissertation, I promise!

Table of Contents

Abstract **ix**

Acknowledgments **xi**

Table of Contents **xiii**

List of Figures **xv**

List of Tables **xvii**

List of Publications **xix**

1 Introduction **1**
- 1.1 Motivation for Sentiment and Stance Visualization 3
- 1.2 Research Problem, Goal, Objectives, and Scope 6
- 1.3 Overview of the StaViCTA Project 7
- 1.4 Dissertation Outline and Contributions 9

2 Background **13**
- 2.1 Sentiment in Linguistics 13
- 2.2 Computational Sentiment Analysis 14
- 2.3 Stance in Linguistics 15
- 2.4 Computational Stance Analysis 17
- 2.5 Information Visualization and Visual Analytics 19
- 2.6 Summary and Challenges 22

3 Related Work and Design Space for Sentiment and Stance Visualization **25**
- 3.1 Text Visualization 26
- 3.2 Sentiment Visualization 36
- 3.3 Stance Visualization 69
- 3.4 Summary 72

4 Visualization of Sentiment and Stance Markers with uVSAT **75**
- 4.1 Background 78

4.2 Related Work . . . 78
4.3 Architecture . . . 81
4.4 Requirements Analysis . . . 83
4.5 Visualization Approach . . . 85
4.6 Case Study: Linguistics Research . . . 96
4.7 Expert Reviews and Discussion . . . 102
4.8 Summary . . . 105

5 Active Learning and Visual Analytics for Stance Classification with ALVA 107
5.1 Background . . . 109
5.2 Related Work . . . 110
5.3 System Overview and Visualization Tasks . . . 115
5.4 Visualization Methodology . . . 119
5.5 Case Study: Annotation Data Exploration . . . 130
5.6 Discussion . . . 135
5.7 Summary . . . 139

6 Visual Analysis of Sentiment and Stance in Social Media Texts with StanceVis Prime 141
6.1 Background and Related Work . . . 143
6.2 Requirements Analysis . . . 147
6.3 Architecture . . . 149
6.4 Visualization Methodology . . . 150
6.5 Case Studies . . . 159
6.6 Discussion . . . 165
6.7 Summary . . . 168

7 Further Applications of Stance Visualization 169
7.1 Interactive Exploration of Temporal, Thematic, and Geographic Aspects of Stance in Social Media Texts with StanceXplore . . . 170
7.2 Visualization of Stance Categories in Longer Individual Texts with DoSVis . . . 179
7.3 Visualization of Sentiment and Stance for Supporting Argument Extraction with Topics2Themes . . . 189
7.4 Summary . . . 192

8 Conclusions and Future Work 193
8.1 Research Findings . . . 193
8.2 Discussion . . . 199
8.3 Future Work . . . 204

Bibliography 207

List of Figures

1.1 High-level overview of the relevant disciplines and domains . . . 3
1.2 Overview of the StaViCTA project activities and results 8
1.3 Overview of the main results of this dissertation 10

3.1 Text visualization survey histogram 28
3.2 TextVis Browser user interface 33
3.3 TextVis Browser entry details 34
3.4 Co-authorship network for text visualization techniques 35
3.5 Sentiment visualization survey methodology 39
3.6 Sentiment visualization survey histogram 40
3.7 Sentiment visualization pipeline 42
3.8 Sentiment visualization category correlation matrix 63
3.9 Sentiment visualization category sparkline plots 65
3.10 SentimentVis Browser user interface 67
3.11 Co-authorship network for sentiment visualization techniques . . 68

4.1 uVSAT approach overview 77
4.2 uVSAT architecture 81
4.3 Timeline view in uVSAT 86
4.4 Document view in uVSAT 87
4.5 Visualization of temporal trends in uVSAT 90
4.6 History diagram in uVSAT 91
4.7 Aggregation charts in uVSAT organized by observer 94
4.8 Aggregation charts in uVSAT organized by marker 95
4.9 Data configuration for the uVSAT case study 97
4.10 Timeline overview for the uVSAT case study 97
4.11 Detailed timeline view for the uVSAT case study 98
4.12 Aggregation chart for “anger” for the uVSAT case study 99
4.13 Document view for the uVSAT case study 100
4.14 Selected document overview for the uVSAT case study 101

5.1 ALVA approach overview 109
5.2 Overview of user tasks supported by ALVA 115

5.3 Architecture of ALVA 116
5.4 Annotation-related user interfaces of ALVA 117
5.5 Entity–relationship diagram for text annotations in ALVA 120
5.6 CatCombos representation design 121
5.7 CatCombos visualization for the complete data set 122
5.8 Annotated data visualization in ALVA 123
5.9 Category co-occurrence matrix in ALVA 125
5.10 Utterances table in ALVA 126
5.11 Annotation process graph in ALVA 126
5.12 Classifier performance plot in ALVA 127
5.13 VSM visualization in ALVA 128
5.14 CatCombos visualization for the ALVA case study 132
5.15 Comparison of VSM visualizations in ALVA 137

6.1 StanceVis Prime approach overview 144
6.2 Visual analysis workflow in StanceVis Prime 149
6.3 Data processing and visualization pipelines in StanceVis Prime . . 151
6.4 Data loading dialog in StanceVis Prime 152
6.5 Visualization interface in StanceVis Prime 153
6.6 Visualization of data series similarity in StanceVis Prime 154
6.7 Visualization of data series values in StanceVis Prime 156
6.8 Document list representations in StanceVis Prime 158
6.9 Results of case study A for StanceVis Prime 160
6.10 Results of case study B for StanceVis Prime 163

7.1 Overview of StanceXplore 174
7.2 DoSVis approach overview 180
7.3 Visualization interface in DoSVis 183
7.4 Overview visualizations for the DoSVis use cases 186
7.5 Stance category combinations for the DoSVis use case 186
7.6 Motivation for Topics2Themes 190
7.7 Topics2Themes approach overview 190
7.8 User interface of the customized version of Topics2Themes 191

List of Tables

2.1 Stance categories defined in StaViCTA 17

3.1 Comparison of text visualization categorizations 27
3.2 Text visualization techniques categorization 30
3.3 Authorship statistics for text visualization techniques 35
3.4 Publication statistics for sentiment visualization techniques 41
3.5 Sentiment visualization techniques categorization 43
3.6 Authorship statistics for sentiment visualization techniques 68
3.7 Comparison of the existing stance visualization approaches 70

4.1 Stance markers in a document set in the uVSAT case study 100
4.2 Stance markers in the selected document in the uVSAT case study 101

5.1 Stance categories annotated with ALVA 110
5.2 Category combination statistics for the ALVA case study 131
5.3 Interesting combinations discovered in the ALVA case study . . . 133
5.4 Interesting utterances discovered in the ALVA case study 134
5.5 Comparison of ALVA with previous approaches 138

6.1 Data categories supported by StanceVis Prime 148
6.2 Examples of utterances in case study A for StanceVis Prime 161
6.3 Examples of utterances in case study B for StanceVis Prime 164

7.1 Classification categories supported by Topics2Themes 190

8.1 Our approaches in the text visualization design space 194
8.2 Our approaches in the sentiment visualization design space . . . 196

List of Publications

This dissertation is based on the following refereed publications in chronological order (I have contributed to all stages of work as the lead author):

1. Kostiantyn Kucher and Andreas Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *Proceedings of the IEEE Pacific Visualization Symposium*, short paper track, PacificVis '15, pages 117–121. IEEE, 2015. `doi:10.1109/PACIFICVIS.2015.7156366`. Materials appear in Chapter 3.

2. Kostiantyn Kucher, Teri Schamp-Bjerede, Andreas Kerren, Carita Paradis, and Magnus Sahlgren. Visual analysis of online social media to open up the investigation of stance phenomena. *Information Visualization*, 15(2):93–116, April 2016. `doi:10.1177/1473871615575079`. Materials appear in Chapter 4.

3. Kostiantyn Kucher, Carita Paradis, Magnus Sahlgren, and Andreas Kerren. Active learning and visual analytics for stance classification with ALVA. *ACM Transactions on Interactive Intelligent Systems*, 7(3):14:1–14:31, October 2017. `doi:10.1145/3132169`. Materials appear in Chapter 5.

4. Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. The state of the art in sentiment visualization. *Computer Graphics Forum*, 37(1):71–96, February 2018. `doi:10.1111/cgf.13217`. Materials appear in Chapter 3.

5. Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. DoSVis: Document stance visualization. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP '18) — Volume 3: IVAPP*, short paper track, IVAPP '18, pages 168–175. SciTePress, 2018. `doi:10.5220/0006539101680175`. Materials appear in Chapter 7.

Additionally, the materials of Chapter 6 have been used to prepare the following full paper manuscript:

- Kostiantyn Kucher, Rafael M. Martins, Carita Paradis, and Andreas Kerren. StanceVis Prime: Visual analysis of sentiment and stance in social media texts. 2019.

This dissertation is also based on the following refereed poster papers (I have contributed to all stages of work as the lead author):

1. Kostiantyn Kucher and Andreas Kerren. Text visualization browser: A visual survey of text visualization techniques. In *Poster Abstracts of the IEEE Conference on Information Visualization*, InfoVis '14, 2014. Related materials are discussed in Chapter 3.

2. Kostiantyn Kucher, Andreas Kerren, Carita Paradis, and Magnus Sahlgren. Visual analysis of stance markers in online social media. In *Poster Abstracts of the IEEE Conference on Visual Analytics Science and Technology*, VAST '14, pages 259–260. IEEE, 2014. `doi:10.1109/VAST.2014.7042519`. Related materials are discussed in Chapter 4.

3. Kostiantyn Kucher, Andreas Kerren, Carita Paradis, and Magnus Sahlgren. Visual analysis of text annotations for stance classification with ALVA. In *Poster Abstracts of the EG/VGTC Conference on Visualization*, EuroVis '16, pages 49–51. The Eurographics Association, 2016. `doi:10.2312/eurp.20161139`. Related materials are discussed in Chapter 5.

4. Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. Visual analysis of sentiment and stance in social media texts. In *Poster Abstracts of the EG/VGTC Conference on Visualization*, EuroVis '18, pages 49–51. The Eurographics Association, 2018. `doi:10.2312/eurp.20181127`. **Best poster award**. Related materials are discussed in Chapter 6.

5. Kostiantyn Kucher, Maria Skeppstedt, and Andreas Kerren. Application of interactive computer-assisted argument extraction to opinionated social media texts. In *Poster Abstracts of the 11th International Symposium on Visual Information Communication and Interaction*, VINCI '18, pages 102–103. ACM, 2018. `doi:10.1145/3231622.3232505`. Materials appear in Chapter 7.

This dissertation partially uses the materials of the following refereed publications (I have contributed to some stages of work in conceptualization, implementation, or writing):

1. Rafael M. Martins, Vasiliki Simaki, Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. StanceXplore: Visualization for the interactive exploration of stance in social media. In *Proceedings of the 2nd Workshop on Visualization for the Digital Humanities*, VIS4DH '17, 2017. URL: `http://vis4dh.dbvis.de/papers/2017/`. Materials appear in Chapter 7.

2. Vasiliki Simaki, Carita Paradis, Maria Skeppstedt, Magnus Sahlgren, Kostiantyn Kucher, and Andreas Kerren. Annotating speaker stance in discourse: The Brexit Blog Corpus. *Corpus Linguistics and Linguistic Theory*,

2017. doi:10.1515/cllt-2016-0060. Related materials are discussed in Chapter 2.

3. Maria Skeppstedt, Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. Language processing components of the StaViCTA project. In *Proceedings of the Workshop on Logic and Algorithms in Computational Linguistics*, abstract paper track, LACompLing '17, pages 137–138, 2017. Related materials are discussed in Chapter 2.

4. Maria Skeppstedt, Kostiantyn Kucher, Manfred Stede, and Andreas Kerren. Topics2Themes: Computer–assisted argument extraction by visual analysis of important topics. In *Proceedings of the 3rd Workshop on Visualization as Added Value in the Development, Use and Evaluation of Language Resources at LREC '18*, VisLR III, ELRA. URL: http://lrec-conf.org/workshops/lrec2018/W16/summaries/2_W16.html. Related materials are discussed in Chapter 7.

Further publications not related to this dissertation (I have contributed to all or some stages of work in conceptualization, implementation, or writing):

1. Kostiantyn Kucher, Daniel Cernea, and Andreas Kerren. Visualizing excitement of individuals and groups. In *Proceedings of the ACM IUI 2016 Workshop on Emotion and Visualization*, EmoVis 2016, pages 15–22. Linköping University Electronic Press, 2016. doi:10.3384/ecp10303.

2. Andreas Kerren, Kostiantyn Kucher, Yuan-Fang Li, and Falk Schreiber. MDS-based visual survey of biological data visualization techniques. In *Poster Abstracts of the EG/VGTC Conference on Visualization*, EuroVis '17, pages 85–87. The Eurographics Association, 2017. doi:10.2312/eurp.20171175.

3. Andreas Kerren, Kostiantyn Kucher, Yuan-Fang Li, and Falk Schreiber. BioVis Explorer: A visual guide for biological data visualization techniques. *PLOS ONE*, 12(11), 11 2017. doi:10.1371/journal.pone.0187341.

4. Kostiantyn Kucher, Rafael M. Martins, and Andreas Kerren. Analysis of VINCI 2009–2017 proceedings. In *Proceedings of the 11th International Symposium on Visual Information Communication and Interaction*, short paper track, VINCI '18, pages 97–101. ACM, 2018. doi:10.1145/3231622.3231641.

Chapter 1

Introduction

Contents

1.1 Motivation for Sentiment and Stance Visualization 3
1.2 Research Problem, Goal, Objectives, and Scope 6
1.3 Overview of the StaViCTA Project 7
1.4 Dissertation Outline and Contributions 9

During the past decade we have witnessed how massively available digital communication channels, such as online social media (forums, blogs, Facebook, Twitter, etc.), affect the world politics and shape the agenda in multiple areas of life. Textual data in particular has been playing an increasingly important role for various analytical tasks in academic research, business intelligence, social media monitoring, journalism, and other areas. The understanding of phenomena occurring in such data is therefore interesting and important for decision makers, researchers, and the general public. For example, text data generated within social media makes it possible for researchers in the discipline of linguistics to employ a bottom-up approach to understand various aspects of language: while the traditional way of manual text investigation involved static corpora, linguists nowadays are able to analyze text data that reflects global events and ongoing language evolution.

Textual data has traditionally been studied within the humanities using methods such as *close reading* [251,299,300] in studies of literature. Researchers in linguistics similarly had to rely on manual methods for collecting and analyzing text corpora before the introduction of computer-assisted methods in the second half of the 20th century [409]. Such manual analyses, however, do not scale up to the volume and variety of digital text data produced in the modern world. Researchers who would like to explore the data beyond the previously compiled corpora can easily find themselves overwhelmed. Thus, the need to provide them with the means to carry out the following tasks becomes apparent: (1) get an *overview* of the available data, (2) *identify* interesting data subsets, (3) *navigate* to them, (4) investigate them *in detail*, and then (5) continue iteratively by switching to the *related* data subsets. In fact, this scenario is applicable not only to researchers or expert analysts, but most kinds of users interacting with large amounts of digital data.

One of the possible solutions to this problem lies within the area of computational disciplines such as artificial intelligence (AI), data mining (DM), and machine learning (ML). Research in ML has demonstrated a lot of promising results during the past decade as soon as truly large training data sets and efficient hardware components became available to at least some institutions and companies. Applications of computational approaches to textual data are studied within the discipline of computational linguistics (CL), and, more specifically, the field of natural language processing (NLP) [279]. However, many linguists (as well as users from other domains) face difficulties when trying to interpret the output of NLP algorithms. For NLP experts, it is equally challenging to gain insight into the underlying text data and to provide useful feedback in order to refine their automatic analyses. In fact, NLP researchers would benefit from techniques that could improve their understanding of computational processes associated with the state-of-the-art NLP algorithms: for example, it is difficult to interpret the state of a large artificial neural network just by weight matrices. Another major problem related to the typical ML-based approaches is the need for large amounts of reliably labeled data in order to train the corresponding computational model.

These predicaments can be resolved by involving the approaches studied within the disciplines of information visualization and visual analytics. Information visualization (InfoVis) focuses on using interactive, computer-aided visual representations of data to support human cognition [64]. Visual analytics (VA) uses interactive visual interfaces to facilitate the analytical reasoning process [212, 417]. A typical VA approach combines techniques from interactive visualizations, as studied in InfoVis, with computational analyses, as studied in DM or ML [219]. The human analyst is then using the VA system interactively to close the sensemaking loop and extract knowledge from the source data [15, 350]. InfoVis and VA techniques take advantage of human perception and cognition, thus being able to help the analysts make sense of the data in cases when completely automatic computational methods either fail or are not feasible at all.

InfoVis and VA methods have been successfully applied to various kinds of textual data. The corresponding field of study, called text visualization (TextVis), has been attracting steadily rising interest from the research community during the past decade [58, 235, 236]. The majority of text visualization techniques rely on the methods originating from CL/NLP that analyze specific aspects of texts, such as (1) syntactical structure of sentences, (2) presence of named entities and relations between them, and (3) topic structure of individual documents, corpora, or text streams. The last of this tasks, topic extraction and visualization, has been especially popular in the research community, as it aims at summarization of the

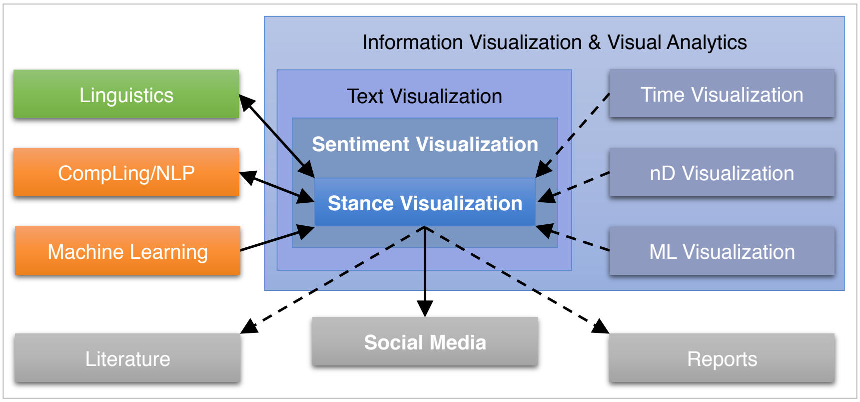


Figure 1.1: High-level overview of the disciplines, fields, and data domains relevant to this dissertation. Shades of blue correspond to visualization-related fields, green—linguistics, orange—computational disciplines, and gray—data domains.

contents of large text collections, which often have additional attributes such as timestamps attached in case of social media [76,105].

Another group of language aspects present in textual data has also been of interest to researchers in linguistics and CL/NLP: expressions of subjectivity such as valence, emotions, and opinions, usually referred to by the umbrella term of *sentiment*. There is also a related (and overlapping) concept of *stance*, which is of interest to researchers in these disciplines. Research on sentiment and stance phenomena can benefit from text data collected from web sources such as social media. Those texts are typically created by multiple authors who engage in discussions and refer to each other's messages, in which they express their thoughts and opinions. However, the difficulties discussed above also apply to attempts of direct manual or computational analyses of such data. This dissertation investigates how to address these issues by the means of *sentiment and stance visualization*.

1.1 Motivation for Sentiment and Stance Visualization

To understand the motivation for pursuing research on sentiment and stance visualization, we must look at these topics in the context of (1) the related disciplines providing the methods for sentiment and stance analysis, (2) the encompassing and related subfields of information visualization and visual analytics, and (3) the relevant data domains and application scenarios (see Figure 1.1).

As discussed above, the notion of sentiment has been studied in linguistics [133, 305] and extensively used in computational linguistics for the task of *sentiment analysis* (the term being often interchangeable with *opinion mining* and *affect analysis*), which is generally concerned with detecting attitudinal content in text at various levels of granularity [260, 295, 317]. Usually, textual data is classified into POSITIVE[1], NEGATIVE, or NEUTRAL at the level of words, utterances, or complete documents. In some cases, even more fine-grained categories related to affect and emotions such as ANGER or JOY are supported. The models used for sentiment analysis range from lexical matching to complex ML classifiers. Historically, sentiment analysis and opinion mining have been researched and applied in CL/NLP and information retrieval (IR) for tasks such as analysis of customer reviews [317], literary analysis [294], and, more recently, for social media monitoring [295]. The interest in practical applications of this approach has also been demonstrated outside of academia. For example, the emotion detection and recognition technologies market (which involves, among other methods, *emotion analysis* of text data with NLP methods) is predicted to grow from USD 6.72 billion in 2016 to USD 36.07 billion by 2021 [119]. Sentiment analysis has even been applied for making automatic stock trading decisions based on the sentiments in tweets by a famous politician [428].

Sentiment visualization is an established subfield within the field of text visualization. The applications and tasks of sentiment visualization include, for instance, monitoring of public opinion in social media, literature analysis for digital humanities, or support for research of sentiment in linguistics and NLP. Some of the earliest papers mentioning visualization of sentiment actually originate in DM and NLP and use basic visual representations in most cases. Recent state-of-the-art techniques often reflect the advances in InfoVis and VA, incorporating sentiment in complex settings involving heterogeneous data. Despite the existence of multiple scientific publications describing individual visualization techniques, the problem with the state of the sentiment visualization field in general has been related to the fact that the existing work had not been properly covered and categorized by any comprehensive survey before the contribution described in this dissertation.

In contrast to sentiment, discussion of *stance* is virtually absent from the visualization literature, and discussed only in a few CL publications. Stance is an area of topical interest in linguistics. The terms "stance" and "stancetaking" are associated with (inter-)subjectivity, evaluation, and appraisal [121]. While the concept of stance is related to the concept of sentiment (more familiar to the visualization and NLP communities), the following example illustrates how an utterance (or sentence) might express subjectivity that is not possible to characterize just in terms of POSITIVE or NEGATIVE polarity (we can argue

[1] Here and below, small caps font is used to indicate concrete text classification categories relevant to sentiment and stance.

that it conveys such stance categories as CONCESSION AND CONTRARINESS and UNCERTAINTY):

> *"Well, I have to admit that I am not completely sure about this . . . "*

Research on stance and stance-taking in linguistics has often used written corpus data of argumentative texts and transcripts of conversations, public speeches, and debates as the source of stance-annotated data with a particular focus on close reading analyses. Social media presents another promising data source for stance analysis. However, the use of social media data for research in linguistics is associated with practical challenges such as the need for identification of potential messages or utterances with stance expressions and further annotation of such source data. With regard to computational analysis, research in CL/NLP has so far mainly operationalized (and often restricted) stance in terms of speakers' attitude towards a given topic and AGREEMENT/DISAGREEMENT between speakers, based only on word form. Developing computational models for automatic detection/classification of stance in a wider sense is, thus, a challenge for CL/NLP.

Finally, the problem of *stance visualization* has so far not been addressed by the visualization community. Stance visualization can, for both theoretical and practical reasons, be treated as a subfield within sentiment visualization, with which it shares both similarities and challenges. In order to support visualization and visual analysis of various aspects of stance outlined above, stance visualization techniques have to address the challenges of representing *multidimensional* data, which usually involves the *temporal* attribute in case of social media data. An additional challenge is to convey the intermediate data provided by the underlying *machine learning* methods, e.g., the confidence or uncertainty of the classification decision. Stance visualization techniques can then be used in (1) actual research on stance in linguistics and CL, (2) social media monitoring, (3) other applications and data domains such as visual analysis of business reports or literature.

The challenges related to definition, analysis, and visualization of stance in textual data require a certain level of support for theoretical research, computational analysis, text annotation tools, and visual exploratory analysis, thus calling for an interdisciplinary research effort. The project "Advances in the Description and Explanation of Stance in Discourse Using Visual and Computational Text Analytics" (StaViCTA)[2] had been launched to answer precisely these challenges. This dissertation builds upon the results of this project in the area of visualization.

[2]`http://cs.lnu.se/stavicta/` (last accessed in February 2019)

1.2 Research Problem, Goal, Objectives, and Scope

As stated above, this dissertation is dedicated to the topic of sentiment and stance visualization. The **research problem** associated with this topic is the absence of a framework describing and instantiating the design space for sentiment and stance visualization in relation to the encompassing and related fields/disciplines: the existing sentiment visualization efforts are typically driven by the specific application needs without a big picture in mind, and stance visualization is barely supported by any existing work at all. Therefore, the main **goal** of this work is to define, categorize, and implement means for visual representation and visual analysis of sentiment and stance in textual data, in particular, for the data originating in social media.

In order to accomplish this goal, the following **objectives** have to be attained:

O1 Position the existing sentiment and stance visualization techniques in the wider context of text visualization;

O2 Define a design space for sentiment and stance visualization techniques;

O3 Enable the design and development of stance visualization techniques by facilitating the underlying research on stance analysis; and

O4 Instantiate the sentiment and stance visualization design space by implementing techniques for textual data from social media as well as other data domains and application scenarios.

The first objective is related to the analysis of the state of the art in the broader field of text visualization. We achieve this objective by establishing a categorization of the existing text visualization techniques, including the tasks of sentiment and stance visualization, surveying the techniques described in the peer-reviewed literature, and analyzing the corresponding results using an interactive browser. This step allows us to relate the tasks and techniques in the focus of our research to other existing work, for instance, we could check if sentiment visualization is often used together with topic visualization.

Consequently, the second objective is concerned with a more detailed investigation of the state of the art in sentiment and stance visualization. We define a design space for these areas with a detailed categorization, survey the related work, and conduct several analyses on the collected categorized data set. The results allow us to identify the most prominent approaches for sentiment and stance visualization and detect gaps in the existing research, which provide opportunities for future work. Additionally, we contribute an interactive survey browser similar to the first step.

In contrast to the previous objectives, the third one focuses more specifically on stance rather than sentiment. It is related to the problems of stance analysis studied in the disciplines of linguistics and computational linguistics. To support the corresponding research tasks, we contribute visual analytics approaches facilitating (1) identification and collection of textual data suitable for stance analysis and (2) data annotation, exploration, and support for training a machine learning classifier. Reaching this objective in cooperation with the domain experts in the respective disciplines leads us to (1) a complete theoretical framework of stance analysis and (2) a stance classifier implementation ready for further usage.

Finally, the fourth objective is concerned with design and development of visualization and visual analytics approaches involving sentiment and stance analysis. We contribute several approaches developed in conjunction with the previous objective as well as several approaches that make use of the automatic classifier supporting multiple stance categories. Besides supporting the visual analysis of temporal text data from social media, we address other parts of our design space and demonstrate applications of sentiment and stance visualization for additional domains, data types, and user tasks.

The goal and the objectives reflect the **scope** of the work described in this dissertation, which is limited to information visualization and visual analytics approaches involving textual data. Therefore, this work is not addressing such problems and approaches as visualization of emotional states extracted with hardware sensors [67, 234] or novel machine learning algorithms for text classification [389, 391]. Furthermore, the work described in this dissertation was carried out as part of the interdisciplinary StaViCTA project described below, thus providing specific constraints for the research scope.

1.3 Overview of the StaViCTA Project

The StaViCTA project was funded by the Swedish Research Council (Vetenskapsrådet) [grant no. 2012–5659][3] and was running during 2013–2017. The main aim of this interdisciplinary project was to develop a both theoretical and practical framework for stance analysis of written text in English with the focus on social media. The project members belonged to three groups representing different research disciplines and fields. A domain expert group in linguistics at the Centre for Languages and Literature, Lund University, was in charge of task identification, stance theory construction, data annotation, and subsequent analyses. A computational linguistics group at the company Gavagai AB was responsible for developing automatic analysis techniques and tools for the project using natural language processing, machine learning, and data mining methods. Gavagai AB also provided the data used for the early stages of the project based on the text documents from online discussion forums and blogs. Finally, the ISOVIS group at the Department of Computer Science and Media Technology, Linnæus University, was responsible for research in information visualization

[3] `https://swecris.se/betasearch/details/project/201205659VR` (last accessed in February 2019)

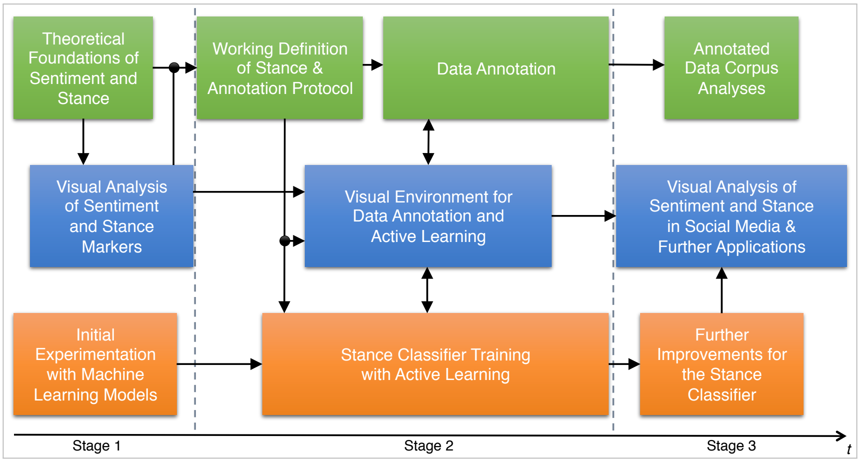


Figure 1.2: Overview of the StaViCTA project activities and results. Horizontal order of blocks approximates the temporal order of work on the corresponding tasks, techniques, and systems. Block width approximates the duration of the activities. Green color corresponds to the tasks mainly carried out by the linguistics group at Lund University, blue—the visualization group at Linnæus University, and orange—the computational linguistics group at Gavagai AB.

and visual analytics and the development of the visual approaches needed in the project and presented in this work.

Figure 1.2 provides an overview of the main activities and results of the project. While the computational linguistics group started by experimenting with various machine learning models [333, 392] using available textual data sets (not specifically labeled for stance), the linguistics group analyzed the existing work on sentiment and stance analysis in the literature [357–359]. In order to investigate the existing sentiment and stance phenomena before the implementation of a proper stance classifier, we implemented a visual analytics approach called uVSAT [237, 243] that consumed data from Gavagai AB, which was processed with a straightforward lexical matching method using lists of seed/marker words. At this stage, both the analysis and visualization were mainly focused on sentiment rather than stance, however, uVSAT supported several other categories, namely, CERTAINTY and UNCERTAINTY, which went beyond standard sentiment classification.

In order to proceed with the implementation of an automatic stance classifier, the next stage of the project had to be dedicated to the data annotation process since there were no available labeled data sets suitable for our purposes. Using the collected data and the results of analytical sessions with uVSAT, the experts in linguistics formulated the working definition of stance that focused on multiple non-exclusive aspects/categories such as HYPOTHETICALS and CONCESSION AND CONTRARINESS, which were then described in detail in the annotation protocol [383]. During the annotation process, the experts in computational linguistics decided to follow the active learning approach with a standard support vector machine (SVM) model in order to train the stance classifier with a very limited amount of training data [389, 390]. To facilitate this stage of the project, we developed an integrated visual environment, called ALVA [238, 242], which supported the data annotation, management of the active learning process, and several visual analyses of the collected data. From the point of view of information visualization, we had to address the challenge of representing multiple non-exclusive category labels visually, which resulted in a novel representation called *CatCombos* ("category combinations").

The final stage of StaViCTA allowed the project members to focus on analyzing and applying the artefacts and findings in their respective disciplines. The experts in linguistics engaged in extensive analyses of the annotated data corpus [381, 382]. The computational linguistics experts worked on further improvements of the stance classifier, eventually switching from the SVM model to the logistic regression (LR) model [386, 388, 393]. Finally, the visualization group applied the stance classifier alongside a standard sentiment classifier for visual analysis of social media data as well as additional applications [239, 241, 244, 286].

1.4 Dissertation Outline and Contributions

After this brief general description of the StaViCTA project, we can now focus on the contents of the remainder of this dissertation. An additional overview of the main results of Chapters 3–7 is also provided in Figure 1.3.

Chapter 2 provides the background information on the definition of sentiment and stance, as studied in linguistics, as well as the existing approaches for automatic sentiment and stance analysis, as studied in computational linguistics. The background information allows us to formulate the corresponding challenges for sentiment and stance visualization. This chapter also provides a brief introduction to the information visualization and visual analytics disciplines, including the discussion of main definitions, tasks, subdisciplines/fields, and application domains.

After discussing the background information, in Chapter 3 we narrow down the scope of our inquiry to more specific research on text visualization and then focus on its subfield dedicated to sentiment visualization. Afterwards, we discuss the existing stance visualization techniques. The contributions of this chapter include the following:

- an overview and a fine-grained categorization of text visualization techniques;

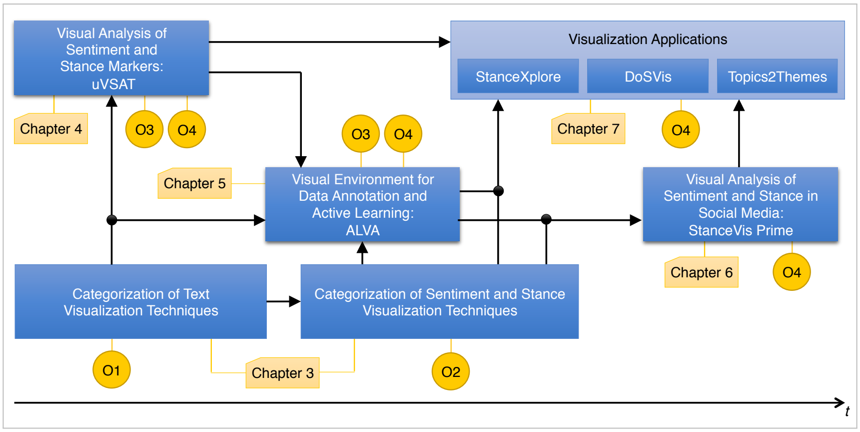


Figure 1.3: Overview of the main results of this dissertation. Horizontal order of blue rectangular blocks approximates the temporal order of work on the corresponding tasks, techniques, and systems. Block width approximates the duration of the activities. Yellow nodes refer to the research objectives and dissertation chapters associated with the related results.

- a survey and a fine-grained categorization of sentiment visualization techniques that provides evidence about supported data, tasks, and visualization aspects;
- an investigation of existing trends and correlations between sentiment visualization categories based on temporal and correlation analyses;
- an analysis of the existing stance visualization techniques in the design space of sentiment visualization;
- interactive survey browsers for both text visualization and sentiment visualization; and
- additional analyses of the research communities for these fields.

Validation of the ideas discussed in this chapter is based on literature survey, critical discussion, and feedback from the visualization research community.

Chapter 4 is dedicated to the discussion of our early efforts in the StaViCTA project, which were based on the previous work in sentiment analysis & visualization and resulted in the development of a visual analytics approach called uVSAT. The contributions of this chapter include the following:

- an analysis of user tasks related to visual analysis of stance phenomena in temporal text data;

- a visual analytics solution for investigating stance phenomena based on sentiment analyses of document texts and time series;
- an *interactive history diagram* for document set queries that facilitates the analysis provenance; and
- interactive *aggregation charts* that provide document set overview, navigation, and comparison functionality with regard to stance categories or specific stance markers.

Validation of this approach includes a case study on visual analysis of temporal and textual data, expert user reviews, and critical discussion.

After the first stage dedicated to collecting the initial data, analyzing it, and conducting experiments with various machine learning models, the next stage of the StaViCTA project included data annotation and training of a machine learning classifier for stance. Chapter 5 introduces a visual analytics environment called ALVA which was designed to facilitate this stage of the project. The contributions of this chapter include the following:

- an analysis of user tasks related to visual analysis of annotated data for stance classification;
- an integrated solution for supporting data annotation, visual analysis, and classifier training using active learning for a complex multi-label text classification task (stance classification); and
- a novel visual representation of multidimensional annotation data, called *CatCombos*, which focuses on the combinations of labels (stance categories) assigned by annotators.

Validation of this approach includes a case study on visual analysis of annotated data, expert user reviews, and critical discussion.

The stance classifier implemented as part of the StaViCTA project could then be applied to the real-world textual data from social media. Chapter 6 describes our visual analytics approach called StanceVis Prime, which is designed to support both sentiment and stance classification results. The contributions of this chapter include the following:

- an analysis of the workflow and user tasks related to visual analysis of sentiment and stance in social media texts;
- a design study involving temporal and textual data analysis methods in order to represent sentiment and stance; and
- a visual analytics solution supporting exploratory data analysis of sentiment and stance classification results for temporal text data from multiple sources.

Validation of this approach includes two case studies on visual analysis of temporal and textual data, expert user review, and critical discussion.

Besides our main activities and tools developed in StaViCTA, we have developed several additional tools presented in Chapter 7. These tools illustrate further applications of sentiment and stance visualization to various tasks, data types, and data domains defined in our design space. The contributions of this chapter include the following:

- a visualization tool, called StanceXplore, designed for exploration of stance in social media data from Twitter with the coordinated multiple views approach and support for geospatial data (illustrated with a case study);
- a visualization tool, called DoSVis, designed for visual stance analysis of longer individual text documents (illustrated with several use cases in different data domains); and
- a version of a visualization tool, called Topics2Themes (designed for computer-aided argument extraction involving topic modeling and visualization), customized for application to social media data and multiple sentiment and stance categories (illustrated with a use case).

Finally, in Chapter 8 we summarize and discuss the results of the work presented in this dissertation and describe directions for the future work related to sentiment and stance visualization.

Chapter 2

Background

Contents

2.1 Sentiment in Linguistics 13
2.2 Computational Sentiment Analysis 14
2.3 Stance in Linguistics 15
2.4 Computational Stance Analysis 17
2.5 Information Visualization and Visual Analytics 19
2.6 Summary and Challenges 22

In order to arrive at the discussion of the design space and the state of the art in sentiment and stance visualization, we firstly have to provide the necessary background information on the concepts of sentiment and stance as studied in linguistics. The discussion of computational analysis methods for both sentiment and stance that are developed by the computational linguistics / natural language processing community then takes place. Finally, in this chapter we provide a brief overview of the disciplines of information visualization and visual analytics in general, which establishes a foundation for a more specialized discussion in the next chapter.

2.1 Sentiment in Linguistics

With the advent of machine-readable corpora, research in linguistics has shown that language is not primarily a means of providing information about facts, but rather to evaluate what we are talking about, to take a stance, and to express opinions and emotions. Language use in different contexts is highly view-pointed, interactive, and interpersonal. Human communication has a purpose. It is in a constant flux and so is the use of language itself [121,319].

Evaluative meanings are not easy to specify in advance because they are not confined to traditional areas of grammar or specific words, but may be expressed by parts of words, words, or longer chunks. Such meanings have been studied under a range of different names in various research traditions in linguistics such as *attitude* [3], *evaluation*, *appraisal*, and *stance taking* [188,284], *epistemic modality*, *subjectivity*, and *intersubjectivity* [282,318,443], and *emotions* and *affect* [21,133].

Research on attitude and emotions in linguistics overlaps with other disciplines such as psychology [138] and computational linguistics / NLP [295]. The relation between emotions and sentiment is analyzed by Munezero et al. [305], who discuss the definitions of sentiment, opinion, emotion, and affect in detail from the standpoint of linguistics and psychology. They point to subtle differences between these concepts that are often overlooked in work originating from more technical fields such as InfoVis, VA, or DM. The emotion space is typically represented by a categorical, dimensional, or hybrid model. One of the most well-known categorical models is Ekman's "Big Six" basic emotions: ANGER, FEAR, HAPPINESS, SURPRISE, DISGUST, and SADNESS [114]. Dimensional models describe emotions in terms of continuous spaces along axes such as *valence/pleasure*, *arousal*, and *dominance* [349]. Finally, hybrid models such as the Plutchik's wheel of emotions [327] define a set of basic emotions, their relative similarity, and possible combinations resulting in numerous derivative emotions.

2.2 Computational Sentiment Analysis

The term *sentiment analysis* as it is used in CL/NLP is usually defined as the task of automatically detecting and classifying affective content in texts at various levels of granularity (from individual words to complete documents, or with regard to specific aspects and entities discovered in text) into a small number of classes representing different kinds of sentiments. Sentiment analysis has become a staple in CL/NLP, both in research and in commercial applications, with a large number of vendors offering solutions for social media monitoring where sentiment analysis is an important part of the analytics suite.

In the simplest formulation, sentiment analysis is considered a binary problem, where we are interested either in detecting the presence of emotionally loaded content, or in distinguishing positively from negatively loaded content. The former of these tasks is closely related to what has been referred to as *subjectivity detection* [317], while the latter is sometimes referred to as *polarity detection* [434].

Arguably, the most common approach to sentiment analysis is to formulate the problem as a three-way categorization task over the categories NEGATIVE, NEUTRAL, and POSITIVE. More complex formulations of the sentiment analysis task involve a broader range of possible sentiment classes, either in terms of a graded scale (e.g., WEAKLY to STRONGLY NEGATIVE and POSITIVE) [397], or in terms of a broader palette of sentiment types [169,396]. One example of a more complex sentiment palette is the RepTrak model used in the RepLab evaluation campaign that includes eight different categories designed specifically for reputation classification [13]. Another closely related task is *emotion detection*, which typically employs a categorical, dimensional, or hybrid model of emotions described above; for instance, Ekman's six basic emotions could be used [25].

Sentiment analysis has also been combined and integrated with other CL/NLP and ML techniques such as topic detection and tracking (TDT), in which case we are interested not only in sentiment expressed in the data, but also the target of the sentiment. As an example, a sentiment analysis system might detect that customers are predominantly negative to the release of a novel product. However, it would be valuable for the product manufacturer to know if there are specific aspects of the product that are more negatively perceived—it might be the case that the negativity only concerns one specific aspect of the product, in which case it might be reasonably easy for the company to make the necessary adjustments. Such analysis is commonly referred to as *opinion mining* or *aspect-based sentiment analysis* [51, 203, 317].

As with any research area that gains popularity in a research community, there has been a wide variety of approaches suggested in the literature. Classification can be based on (1) lexical matching of keywords from a previously constructed dictionary/lexicon [124] (such as WordNet-Affect [404], MPQA Subjectivity Lexicon [472], SentiWordNet [23], LIWC [414], or SenticNet [56]), (2) knowledge about word/concept similarity (e.g., using a distributional semantics model or an ontology such as WordNet [290]), (3) the use of topic modeling algorithms and latent variable models [257, 274, 450], or (4) a variety of machine learning (ML) classification models, both standard [317,451] and deep learning architectures [396, 411]. State-of-the-art approaches to sentiment analysis now approach, and in some cases even exceed, 90% accuracy on standardized benchmark test suites [208, 249, 451]. For practical purposes of sentiment classification of short texts from social media, rather simple models with standard openly available implementations continue to be widely used, for instance, the rule-based classifier called VADER presented by Hutto and Gilbert [191].

In general, variation in terminology and computational models of sentiment analysis are covered by several survey articles—the arguably most comprehensive one is that of Pang and Lee [317], who discuss work done in linguistics, CL/NLP, DM, and ML. More recent surveys include the works by Tsytsarau and Palpanas [431], Cambria et al. [57], Ravi and Ravi [335], and a comprehensive survey by Mohammad [295].

2.3 Stance in Linguistics

Stance is a topical area of interest in linguistics associated with *subjectivity/intersubjectivity*, *evaluation*, and *appraisal* [121]. It is closely related to and overlapping with *evaluation*, *positioning*, and *alignment* in discourse, as well as *sentiment* discussed above in Section 2.1. Stance provides an interesting research problem to linguists because the interactive nature of communication between individuals is considered vital. The function of taking stance in the communicative situation is to convey the speaker's viewpoint of what is talked about and to regulate the exchange between the dialog partners. Communication here works on more than pure understanding of words. Words are always understood in the light of the contexts and the situations where they are used [158,196]. In doing so, language is used to recontextualize human experiences into written and spoken forms. Its social role is to affect the state of mind of other people and to negotiate meanings in order to bring about cognitive changes [319,463]. Language users construe their expressions to communicate their particular perspective and viewpoint of what is talked about.

Stance has been studied under different headings and scope, such as *evaluation* [189,418], *sentiment* [293], *appraisal* [284], and of course under the title *stance* itself [36,37,121,149,319]. Du Bois defines stance-taking as follows [110, p. 163]:

> *Stance is a public act by a social actor, achieved dialogically through overt communicative means, of simultaneously evaluating objects, positioning subjects (self and others), and aligning with other subjects, with respect to any salient dimension of the sociocultural field.*

According to his model, the process of taking stance comprises three parts: (1) speaker evaluation of what is talked about, (2) speaker positioning (epistemicity), and (3) alignment in communication, i.e., establishment of agreement or disagreement.

Our work in the StaViCTA project was geared toward identifying multiple aspects and categories of stance in written language, such as CONTRAST and CONDITIONALS [392], CERTAINTY/UNCERTAINTY [243], and the previously mentioned AGREEMENT/DISAGREEMENT [391]. While the initial efforts relied on preconceived lists of seed/marker works associated with emotions and CERTAINTY/UNCERTAINTY for practical reasons, later the researchers of StaViCTA took an utterance-based approach to the analysis of speaker stance in communication based on the identification of constructions that actually express stance on the occasion of use [383]. An *utterance* here refers to a chunk of text between two delimiters such as full stops. The researchers in linguistics defined a cognitive-functional framework consisting of ten notional stance categories (presented in Table 2.1) and prepared an annotation manual for labeling the textual data. The set of stance categories was selected through a process involving analysis of previous work in linguistics and CL, multiple discussions, test annotations based on small data sets, and elimination/merging of several categories [390,393]. Note that the categories are not mutually exclusive, i.e., the same utterance could be labeled with several categories by the annotator. Further details about our annotation process and the resulting data set used for the practical purpose of training a stance classifier are given in Chapter 5.

Table 2.1: The stance categories defined in the StaViCTA project

Category	Description	Examples
AGREEMENT AND DISAGREEMENT	Expression of a similar or a different opinion by the speaker	*You got it, it is so right*; *I beg to differ*
CERTAINTY	Expression of confidence in the truth of the utterance	*I know that the bus is late*; *Without a doubt, today was the hottest ever*
CONCESSION AND CONTRARINESS	Expression of a compromising or a contrastive/comparative opinion	*It's quite good, though it could be better*; *It'll work for now, yet in due course it needs to be replaced*
HYPOTHETICALS	Expression of a possible consequence of a condition	*If it's nice tomorrow, we will go*; *And if you were in the mood we could at least go*
NEED/ REQUIREMENT	Expression of a request, recommendation, instruction, or obligation	*This ought to be done before noon*; *Place the chicken in a 9x13 dish*
PREDICTION	Expression of a guess/conjecture about a future event	*I guess it'll rain*; *I knew it would be a good experience*
SOURCE OF KNOWLEDGE	Expression of the origin of what is said in the utterance	*I saw her talking to Bill yesterday*; *We perceived the problem at once*
TACT AND RUDENESS	Expression of pleasantries or unpleasantries	*Please, do come in*; *Bloody hell, are you mad!*
UNCERTAINTY	Expression of clear doubt about the likelihood of the utterance	*There might be a few problems*; *I was under the impression that there wasn't one*
VOLITION	Expression of wishes and refusals (inclinations and disinclinations)	*We wanted the table by the window*; *I wouldn't do that even if you paid me*

2.4 Computational Stance Analysis

The analysis of stance in written language reveals the feelings and attitude of speakers (utterers) towards their own and other people's utterances. From the computational point of view, it is rather natural to draw parallels with the related task of sentiment analysis discussed above in Section 2.2. Sentiment analysis is normally considered as a classification problem over two or three classes, where POSITIVE and NEGATIVE define the basic polarity, and NEUTRAL is used to describe a lack of attitudinal content. From the perspective of stance analysis, this is a very simplistic ontology of subjectivity that is likely to be too restricted. While the phenomena such as sentiment and subjectivity have enjoyed considerable attention in the CL/NLP community [258, 260, 317], other phenomena related to stance like belief, trust, and uncertainty have remained comparatively peripheral (but there is a number of efforts to analyze uncertainty and speculation [144, 442], respectively).

With regard to computational analysis, research in CL/NLP has so far mainly operationalized (and often restricted) stance in terms of speakers' attitude towards a given topic and agreement/disagreement between speakers. The SemEval-2016 contest included a shared task on automatic stance analysis in Twitter texts, focusing on the speaker's FOR/AGAINST position with regard to a certain target, such as climate change or certain US election candidates [296]. Mohammad et al. [297] describe a state-of-the-art classifier that achieves an average F-score of 0.703 by using the features based on the particular targets present in this data set. However, their experiments with a single model for all targets and experiments with a data subset where opinions were expressed towards entities/topics other than the given target have showed much lower results, which opens up for future work.

The CL/NLP researchers in the StaViCTA project contributed to the body of work on computational stance analysis by developing a classifier for the notional stance categories presented in Table 2.1. Earlier investigations of NLP and ML methods in the project indicated that a standard *support vector machine* (SVM) [326, 422] was suitable for text classification tasks for related categories such as REPUTATION [333] and SPECULATION [392]. Therefore, SVM was chosen as the initial model for stance classification at the level of individual utterances/sentences. The process of developing and training the classifier took place in parallel with the annotation process, which was facilitated by the visual environment described below in Chapter 5. Since the approach taken by the researchers in linguistics meant that the stance categories were not mutually exclusive and the same utterance could be potentially labeled with all the categories simultaneously, the task of classification was a *multi-label* [430] problem rather than a more usual *multi-class* one. To address this task, the classifier was implemented as a collection of technically separate, independent binary SVM classifiers. Each binary classifier detects whether a specific category of stance such as NEED/REQUIREMENT is present or not in the input utterance. The classifier is using *n-gram* features [279] weighted with the *TF-IDF* method [354], and it is implemented with the Scikit-learn library [321] for Python.

Since no suitable labeled data sets existed for our stance classification tasks, had to pursue the annotation process as part of the project. It was evident early that our training data set would be rather small, thus providing additional constraints for the computational analyses. The CL/NLP researchers thus decided to follow the *active learning* approach [369, 370, 422] for training the classifier. After some initial experiments [390], the early annotation rounds took place for the set of stance categories presented above in Table 2.1. Afterwards, the initial version of the classifier was trained, and based on the state of the model, a batch of yet unlabeled utterances was selected as the most promising candidates for further annotation. This process was repeated in a loop of annotating text data, re-training the classifier, and retrieving a set of utterances for the next annotation round. Just as the data annotation process itself, the active learning process was facilitated by the visual environment discussed in Chapter 5.

In order to investigate ways to further improve the quality of stance classification, the researchers in CL/NLP also conducted additional token-level annotations of stance markers within the previously labeled utterances, arguing that local cue words/chunks might signal the presence of stance categories and thus be useful for classification [393]. A tool titled PAL was developed for the automatic pre-annotation for this task [388]. The researchers then incorporated additional annotations into the training process for a new version of the stance classifier that was based on the *logistic regression* (LR) model [183]. At this stage, the set of stance categories was also changed for practical reasons [386]: (1) VOLITION was excluded due to extreme scarcity of labeled data, (2) AGREEMENT and DISAGREEMENT as well as TACT and RUDENESS were split into separate categories for better differentiation, and (3) a category titled CONTRAST [22] was introduced.

The output of the SVM- and LR-based stance classifiers for any given input utterance, thus, comprises a set of up to 10 or 12 stance categories. The lack of any detected categories means that the utterance is NEUTRAL. Besides these categorical values, both classifiers provide a normalized measure of *confidence* of the classification decision: the SVM-based classifier uses Platt scaling [326], and the LR-based classifier provides a probability estimate [183] directly. These values can be used for the analysis of classification results, for instance, by using the methods developed within information visualization and visual analytics. A brief overview of these disciplines is introduced in the next section.

2.5 Information Visualization and Visual Analytics

As mentioned in Chapter 1, one of the potential solutions for the issues of data analysis involves the methods and techniques studied within the discipline of *information visualization* (InfoVis). The standard definition of InfoVis is, arguably, the one coined by Card et al. [64, p. 7]:

> *The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.*

While the related discipline of *scientific visualization* (SciVis) [50,100] maintains close ties to the computer graphics community, focuses on spatial data, and addresses tasks such as volume and flow visualization, InfoVis has developed into a separate discipline during 1990s with a focus mainly on non-spatial, abstract data. InfoVis draws inspiration from earlier work on visual design, infographics, and statistical charts by designers such as Bertin [35] and Tufte [432]. In contrast to SciVis, InfoVis also has closer association with the discipline of human-computer interaction (HCI) [331,373]. InfoVis makes use of research on human perception and cognition [461] in order to facilitate the design of effective and efficient visualization and interaction techniques. The main goal of InfoVis approaches is usually to support either exploratory data analysis [433] or visual storytelling [368].

A typical InfoVis technique can be illustrated by the reference model proposed by Card et al. [64], which defines several steps of data preprocessing and visual encoding, at the same time supporting user interactions at various stages of this pipeline. In addition to the traditional visual representations such as tables, line charts, and node-link diagrams, over the years researchers in InfoVis have proposed a large number of novel and complex visual metaphors for representation of the corresponding data [427, 487]. Visualizations can also follow the strategy of coordinated multiple views [343, 455], combining a number of interlinked representations rather than using a single integrated view that might be too complex and confusing for the users.

Following some earlier attempts to categorize the existing visualization techniques and define a design space for the new ones (e.g., see the work of Wehrend and Lewis [468] in SciVis), researchers in InfoVis have created a number of categorizations (taxonomies, typologies, etc.) of *user tasks* over the years. Shneiderman has proposed a taxonomy that includes seven general data types and seven user tasks, famous for its "Visual Information Seeking Mantra" [374, p. 337]:

> *Overview first, zoom and filter, then details-on-demand.*

The complete list of tasks in this taxonomy includes "Overview", "Zoom", "Filter", "Details-on-demand", "Relate", "History", and "Extract". Brehmer and Munzner [48] discuss a detailed multi-level typology of tasks addressing the *why*, *how*, and *what* aspects. This typology could be used for task analysis at a deeper level, if the ambiguity between the ends and means of tasks, i.e., the *why* and *how* aspects, presents a problem.

The traditional InfoVis techniques mainly focused on representing the existing data available from databases by introducing novel visual metaphors and interactions. However, the level of support for various computational methods as part of the complete analytical pipeline was rather low at that stage. Joint efforts from the researchers and practitioners in data mining and InfoVis led to emergence of *visual data mining* techniques [98, 341] and, eventually, the discipline of *visual analytics* (VA), which was described by Keim et al. [212, p. 157] as follows:

> *Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets.*

The typical user tasks in VA include previously defined InfoVis tasks such as "Overview" and "Filter", but also the tasks related to computational methods and the overall analytical process, e.g., "Adjust *[the model parameters]*" and "Guide *[the analyst through the workflow]*", as discussed by Kerren and Schreiber [219]. Interactions between the human analyst, the data, and both the computational and visual methods are crucial for the complete analytical process. One model widely used in VA research is the sensemaking process model for intelligence analysis by Pirolli and Card [325]. It comprises a sequence of abstract artifacts and steps (e.g., external data, searching & filtering, and formulating a hypothesis) connected with a number of local feedback loops and two global loops, one for data foraging and one for sensemaking. The products of such a process are the data schema/model, the hypotheses supported by the data, the hypothesis testing results, and the accompanying insights that can be used by the analysts and/or presented to other decision makers. Based on this more general model, Sacha et al. [350] introduce a knowledge generation model for VA, where hypotheses and insights are related to *knowledge* as the final product desirable by human analysts. In a recent article, Andrienko et al. [15] propose to interpret the VA process as building appropriate models for the respective problem domains rather than analyzing the data for its own sake.

There are multiple fields of research within InfoVis and VA dedicated to specific data types and properties, e.g., networks or geospatial data; we introduce several areas most relevant to this dissertation below. *Temporal data visualization* is one of the most important fields since time-varying data is present in a wide spectrum of problems and domains, and time cannot be interpreted just as another regular data dimension in most cases. Aigner et al. [4] provide a comprehensive survey of time visualization techniques and complement it with an interactive survey browser [421], which has itself been very influential for the visualization community. *Dynamic & streaming data visualization* is concerned with the problem of representing the data that is continuously arriving and, in the latter case, eventually becoming unavailable; the visualization is usually expected to occur in *(near-)real time* in this problem. Cottam et al. [89] discuss the taxonomy and the design space for dynamic data visualization, and Dasgupta et al. [97] provide a comprehensive survey of the streaming data visualization techniques. The survey by Wanner et al. [458] focuses specifically on event detection in text streams, a data type relevant to the subject of this dissertation.

Another important field of study is *multidimensional data visualization*, which is concerned with visual representations of non-trivial data with a large number of abstract dimensions/attributes. To address this problem, visualization techniques often have to rely on clever metaphors, for example, *glyphs* [40, 460] and *parallel coordinates* [173]. A comprehensive survey of the existing visualizations for high-dimensional data is provided by Liu et al. [267]. Besides encoding as many data dimensions as possible with the visual representation, another possible strategy is to select the most informative dimensions or to synthesize new ones. This approach of *dimensionality reduction* (DR) was originally popularized as part of the feature engineering process in machine learning (ML) and other computational disciplines before its introduction into InfoVis and VA. The classic DR methods include, for instance, *principal component analysis* (PCA) [204], which originates from the discipline of statistics. PCA essentially projects the original high-dimensional data into a low-dimensional space defined by its *principal components* as the dimensions or axes. The target space is typically two- or three-dimensional, and it can be visualized with a standard scatterplot representation. While PCA uses the source data in form of observations/items (rows) and attributes (columns), another classic technique called *multidimensional scaling* (MDS) [39] takes input in form of a distance matrix. Among the more recent DR techniques, *t-distributed stochastic neighbor embedding* (t-SNE) [275] is arguably the most popular one. This non-linear DR method attempts to position projections of the original high-dimensional items in a low-dimensional space so that the neighborhood between the pairs of items is preserved. DR methods have been widely used in the visualization community, with some general guidelines described by Sedlmair et al. [367] and Sacha et al. [351], among others.

Finally, the recent trend in InfoVis and especially VA is related to representation and interaction with AI/ML models. Earlier work on this issue includes interactive visualization for supporting decision tree construction [462], training text [172] or video data [178] classifiers, and feature selection [271], for instance. Recent work is covered by several survey articles, including the surveys on predictive visual analytics [269], integration of ML models into VA [120], and interactive ML [112].

Interdisciplinary collaboration is a crucial aspect of InfoVis and VA. While the example above indicated the rising interest for collaborations with AI/ML communities, there are some fields within visualization that were established quite a long time ago to address the needs of domains experts in the corresponding disciplines. Some examples here include *biological data visualization* (BioVis) [216, 217] and *software visualization* (SoftVis) [113,437], for instance. The disciplines and data domains most relevant to this dissertation include linguistics [218], digital humanities [197], and social media [76]. One feature common for these three areas is the usage of *textual* data. Collaboration between the domain experts and researchers in InfoVis and VA would then involve visualization and visual analysis of such data in order to make sense of it, possibly including sentiment and stance concepts and analyses introduced above. An in-depth look into *text visualization* and then *sentiment* and *stance visualization* in relation to more general InfoVis and VA approaches described above will be discussed in the next chapter.

2.6 Summary and Challenges

This chapter has provided a brief introduction to the research topics in linguistics and computational linguistics / natural language processing relevant to this dissertation. We have also described the origins, main tasks, and research fields within the disciplines of information visualization and visual analytics, which provide us with a background for a detailed discussion of more specific fields in the scope of these disciplines.

Based on the discussion of the well-established research on sentiment analysis in linguistics and CL/NLP, we can summarize the corresponding general properties and challenges of *sentiment visualization* as follows:

- there is a need to support visualization of a variety of tasks related to sentiment analysis, ranging from subjectivity detection to emotion analysis and stance analysis;
- there is a need to support the data specific to the sentiment analysis model (e.g., lexicon-based or ML-based) and scope (word-level, utterance-level, etc.); and
- there is a need to support a variety of data domains and user tasks existing in research and applications of sentiment analysis, which range from theoretical research in linguistics and NLP to social media and news monitoring, thus implying the usage of various visual channels and representations.

Based on the discussion of the previous research on stance in linguistics and the rather scarce work on computational stance analysis in CL/NLP, the general research challenges of stance analysis and *stance visualization* can be summarized as follows:

- there is a need for large-scale studies of stance in linguistics based on notional functional-semantic categories rather than single words;
- there is a need for stance-annotated text data sets for analysis and ML purposes from multiple domains and genres, including social media, which are labeled not only with regard to the speakers' FOR/AGAINST positions, but also other fine-grained categories of stance;
- there is a need for corresponding annotation tools (which could use visual methods, among others) that would explicitly support the data types and categories discussed above and decrease the overall complexity, duration, and cost of the annotation process;
- there is a need for effective stance classification models and methods that could be used for research and applied purposes, for instance, identification of stance in Twitter texts for a social media monitoring application; and
- there is a need for effective visualization and interaction techniques to facilitate the annotation, exploration, and presentation/dissemination stages of stance analysis.

In order to arrive at such stance visualization techniques, in the next chapter we will firstly discuss the design space of text visualization as part of InfoVis and VA, and then we will use it to derive a more specialized design space for sentiment and stance visualization.

Chapter 3

Related Work and Design Space for Sentiment and Stance Visualization

Contents

3.1 Text Visualization 26
3.1.1 Text Visualization Categorization 26
3.1.2 Interactive Browser 32
3.1.3 Discussion and Analysis 34
3.2 Sentiment Visualization 36
3.2.1 Related Surveys 37
3.2.2 Survey Methodology 38
3.2.3 Data Aspects 44
3.2.4 Tasks 51
3.2.5 Visualization Aspects 56
3.2.6 Discussion and Analysis 62
3.3 Stance Visualization 69
3.4 Summary 72

In this chapter, we focus on the analysis of related work in sentiment and stance visualization and synthesis of a design space based on a fine-grained categorization. To identify and categorize the relevant existing work, firstly we have to start with a more general field of text visualization. We briefly describe our categorization, an online survey browser, a manually curated data set of visualization techniques, and additional analyses carried out with this data. Then we introduce our detailed categorization of sentiment and stance visualization techniques, discuss the existing works, and analyze the general trends and patterns discovered in this field. This chapter provides us with the understanding of both best practices to follow and research gaps to address with the sentiment and stance visualization approaches introduced in the rest of this dissertation.

3.1 Text Visualization

The interest for text visualization and visual text analytics has been increasing for the past 10–15 years. The reasons for this development are manifold, but for sure the availability of large amounts of heterogeneous text data (caused by the popularity of online social media) and the adoption of text processing algorithms (e.g., for topic modeling) by the InfoVis and visual analytics communities are two possible explanations. Inspired by the TreeVis.net [364] and TimeVis [4, 421] projects, in this section [236][1] we describe an interactive visual survey of text visualization techniques that can be used for getting an overview of the field, teaching purposes, and finding related work based on various categories defined in a survey categorization, which was introduced in our previous poster paper [235]. As of February 2019, our web-based survey browser is available at

`http://textvis.lnu.se/`

The term "text visualization" is typically used for information visualization techniques that in some cases focus on raw textual data, in other cases on results of text mining algorithms. In the same way, they can be rather general or very specialized and dedicated to specific analytic tasks or application domains. This is the reason why we have decided to construct a categorization with numerous categories and category groups that is exploited by the survey browser in order to facilitate the interactive exploration of the current set of entries. Our visual survey has been implemented as an interactive web page and includes 430 techniques[2] at present originating from peer-reviewed work in InfoVis, VA, and other relevant research fields. After a short discussion on relevant surveys in the following subsection, we highlight the categorization used by our survey browser, some implementation details, and the results of analyses conducted on the collected entries data in the remainder of this section.

3.1.1 Text Visualization Categorization

There are a number of survey papers in the literature that focus on text visualization or its specific subproblems. Šilić and Bašić [379] classify about 30 text visualization methods with regard to data source, underlying text representation & processing method, temporal aspects, and supported user interactions. Alencar et al. [7] describe roughly 30 techniques by means of data source, underlying text representation, visual metaphor, layout, and supported user tasks. Gan et al. [142] discuss approximately 40 techniques with regard to data source, user

[1]This section is based on the following publication: Kostiantyn Kucher and Andreas Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In Proceedings of the IEEE Pacific Visualization Symposium, short paper track, PacificVis '15, pages 117–121. IEEE, 2015. doi:10.1109/PACIFICVIS.2015.7156366 © 2015 IEEE.

[2]Please note that our text and sentiment visualization surveys focus on *techniques* rather than *publications*, see Section 3.2.2.

Table 3.1: The comparison of text visualization categorizations

Category Group vs **Categorization**	Šilić and Bašić [379]	Alencar et al. [7]	Gan et al. [142]	Nualart-Vilaplana et al. [311]	Wanner et al. [458]	**Our catego-rization**
Data Domain				●		●
Data Source	●	●	●	●	●	●
Data Properties (temporal, etc.)	●			●	●	●
Underlying Data Representation	●	●		●		
Data Processing Methods	●			●	●	○
Analytic Tasks		●	●		●	●
Visualization Tasks	●	●	●		●	●
Visual Dimensionality		●	●	●	●	●
Visual Representation (metaphor)		●	●	●	●	●
Visual Alignment (layout)		●	●	●	●	●

Note: Supported categories are marked by ●, partial support denoted by ○. *Reprinted from [236] © 2015 IEEE.*

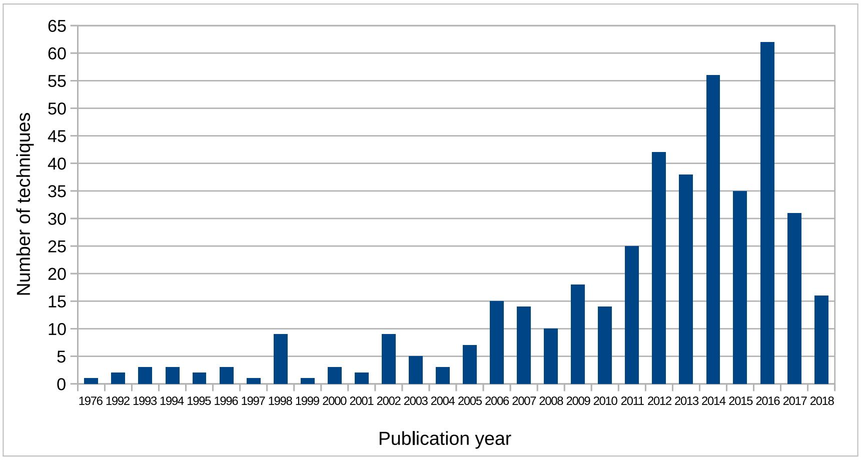


Figure 3.1: Histogram of the collected text visualization techniques set (**430** techniques in total as of February 6, 2019) with regard to the publication year.

tasks, visual representation, and supported interactions. Nualart-Vilaplana et al. [311] categorize about 50 techniques on the basis of data source, underlying text structure & corresponding processing method, support for temporal aspect (as well other special data properties), data domain, and visual metaphor. The work of Wanner et al. [458] on event detection in texts classifies approximately 50 visualization approaches with regard to data source, text processing methods, event detection methods, visualization representations, and tasks. Table 3.1 provides an overview of all these surveys and the categorizations they used.

Finally, the aforementioned visual survey projects use dimensionality, visualization metaphor, and visual alignment to classify tree-oriented techniques [364]; and data properties, temporal properties, and visual representation to classify time-oriented techniques [4,421].

We have arranged a categorization with multiple categories and category groups in order to classify the techniques with fine granularity. The categorization presented in Table 3.2 is the result of refinements occurring while categorizing entries for the survey, i.e., the choice of concrete categories is motivated by the underlying data. The collected set of text visualization techniques includes 430 entries as of February 6, 2019, and it is mostly based upon publications from 2006–2018 (see Figure 3.1) with some earlier examples such as the original description of the tag cloud technique [289] going as far as 1976. While we cannot claim that our classification is absolutely definite (numerous techniques have been ambiguous, especially in case of hybrid approaches), we have tried to base the choice of categories for particular entries on the description and claims of the original author(s). For example, certain techniques could be easily applied to domains other than originally described, but we do not reflect that in our choice of categories for those techniques. On the other hand, some papers mentioned specific domains only for the sake of giving examples, though the corresponding techniques were not tailored for those domains. In such cases, we have not assigned entries to the domains. In the remainder of this subsection, we briefly introduce the categories comprising our categorization and provide some references to prominent examples, with a much more detailed discussion of most categories below in Sections 3.2 and 3.3.

3.1.1.1 Data Aspects

First of all, the *domain* category group describes the dedicated data domains a technique was developed for.

- **Online Social Media** Twitter, Facebook, blogs, forums, etc. [103].
- **Communication** We include email, instant messaging logs, or snail mail letters into this category [413].
- **Patents** Official patents for detailed disclosure of inventions [99].
- **Reviews / (Medical) Reports** This category denotes user reviews, medical report data, and reviews & reports from other sources [453].
- **Literature/Poems** Various artistic, historical, and documentary texts [403].
- **Scientific Articles/Papers** Scientific texts of various genres and fields [394].
- **Editorial Media** Text data from organizations (newspapers, etc.) as well as pre-moderated websites (e.g., Wikipedia) [412].

We have decided to even categorize the techniques with regard to both data source and special data properties (if any supported). Data *sources* include the following self items: **Document** [465], **Corpora** [337], and **Streams** [231]. The special data *properties* include **Geospatial** [106], **Time series** [134], and **Networks** [62].

3.1.1.2 Analytic Tasks

This category group describes high-level analytic tasks that are facilitated by corresponding techniques: these categories are critical to the main analysis goals that users expect to achieve when employing a text visualization technique.

- **Text Summarization / Topic Analysis / Entity Extraction** We have decided to combine entity extraction/recognition with topic analysis/modeling in a single category item, since visualization techniques treat entity names simply as topics in many cases encountered by us [116].
- **Discourse Analysis** This category concerns the linguistic analysis of the flow of text or conversation transcript [16].

Table 3.2: The complete categorization of text visualization techniques used in our survey

Data Domain	338
Online Social Media	146
Communication	33
Patents	5
Reviews / (Medical) Reports	44
Literature/Poems	52
Scientific Articles/Papers	42
Editorial Media	89
Data Source	**425**
Document	89
Corpora	342
Streams	54
Data Properties	**266**
Geospatial	61
Time Series	203
Networks	102

Analytic Tasks	417
Text Summarization / Topic Analysis / Entity Extraction	284
Discourse Analysis	25
Stance Analysis	16
Sentiment Analysis	162
Event Analysis	48
Trend Analysis / Pattern Analysis	184
Lexical/Syntactical Analysis	48
Relation/Connection Analysis	208
Translation / Text Alignment Analysis	19
Visualization Tasks	**430**
Region of Interest	91
Clustering/Classification/Categorization	320
Comparison	352
Overview	403
Monitoring	45
Navigation/Exploration	282
Uncertainty Tackling	27

Visual Dimensionality	430
2D	413
3D	27
Visual Representation	**430**
Line Plot / River	139
Pixel/Area/Matrix	173
Node-Link	147
Clouds/Galaxies	169
Maps	93
Text	214
Glyph/Icon	121
Visual Alignment	**423**
Radial	87
Linear/Parallel	259
Metric	224

Note: Each row contains the number of corresponding visualization techniques in our survey as of February 6, 2019. The percentage relative to the current total of **430** techniques is also illustrated by heatmap-style icons.

- **Stance Analysis** This category is associated with the techniques facilitating analysis of stance, as discussed in more detail in Section 3.3.
- **Sentiment Analysis** We have used this category for techniques related to the analysis of sentiment, opinion, and affect—see Section 3.2 for more details.
- **Event Analysis** While event analysis and visualization is in fact a separate subfield [458], some of the corresponding techniques deal with the extraction of events from the text data or involve visualization of text in some different manner [272].
- **Trend Analysis / Pattern Analysis** This category denotes the tasks of both automated trend analysis and manual investigation directed at discovering patterns in the textual data [492].
- **Lexical/Syntactical Analysis** We have included this category to represent various linguistic tasks, for instance, analysis of lexemes and sentences in poems [465].
- **Relation/Connection Analysis** This category is dedicated to comparison of data items, including the analysis of explicit relationships exposed by visualizations [483].
- **Translation / Text Alignment Analysis** We use this category for corpus linguistics tasks, for instance [147].

3.1.1.3 Visualization Tasks

This category group describes lower-level representation and interaction tasks that are supported by the text visualization techniques. In comparison to analytic tasks, we have included more instrumental items here, for example, clustering could be used in various visualizations as merely an auxiliary feature.

- **Region of Interest** This task denotes the automatic highlighting/suggestion of data items/regions that could be of interest to the user for more detailed investigation, including peaks and outliers [345].
- **Clustering/Classification/Categorization** Here, we combine several tasks related to (semi-)automatic tagging or grouping of data elements [61].
- **Comparison** This category denotes the comparison of several entities facilitated by the visualization technique, e.g., laying out several objects side by side or marking discrepancies [261] (also see the survey by Gleicher et al. [148]).
- **Overview** We use a very general notion of "overview" for this item, including both techniques that provide "the big picture" by displaying a significant portion of the data set as well as techniques which use special aggregated representations to provide overview while reducing the visual complexity [190].

Monitoring This task is related to visualization techniques that are designed to alert users to the changes in the dynamically updated data [59].

Navigation/Exploration We use this category for techniques that facilitate the process of navigating around the data set, while possibly switching the visual representations or underlying data types [262].

Uncertainty Tackling This category—which is currently not very prominent in the present techniques—is generally related to techniques that handle and/or visualize uncertainty in source or processed data as well as uncertainty in computations [481].

3.1.1.4 Visualization Approach

Finally, to categorize the techniques with regard to the used visualization approach, we use three groups of categories. While visual *dimensionality* does not require additional description, we list the others. *Representation* includes the following items: **Line Plot / River** [93,167], **Pixel/Area/Matrix** [16,83,113], **Node-Link** [464], **Clouds/Galaxies** [5,14], **Maps** [473], **Text** [344], and **Glyph/Icon** [99,406]. *Alignment*, i.e., layout, includes **Radial** [481], **Linear/Parallel** [84], and **Metric-dependent** [259].

3.1.2 Interactive Browser

We have implemented our visual survey as an interactive HTML/JavaScript page that merely requires a modern web browser for access, see Figure 3.2 for a screenshot. The survey browser has a main view with a collection of thumbnails (ordered by time) that represent the individual visualization techniques as well as filter controls that include a text search field, a publication year range slider, and category radio buttons. Since the included technique entries may be assigned with arbitrary sets of category tags, and the filtering is based on logical "OR" operation, the interface contains additional category filters for "Other" entries to support precise filtering, e.g., to display only entries that are not associated with any domain.

After clicking on an entry's thumbnail image, the corresponding details are displayed in a dialog box. Here, a slightly larger thumbnail, a complete list of assigned category tags, a bibliographical reference, a URL (optional), and a link to a BibTeX file (if available) are displayed, as seen in Figure 3.3.

We have also provided an additional form for authors who wish to add a new entry to our survey. The form generates a JSON entry [206] that can be sent to us via email to prevent direct manipulation of the survey browser content. Finally, we visualize some basic statistics about the current entry set in the "About" dialog. Since the techniques can be assigned with multiple category tags, the sets of corresponding techniques overlap for sibling categories—therefore, we currently use simple bar charts for showing the statistics.

Figure 3.2: The web-based user interface of our visual survey called *TextVis Browser* ("Text Visualization Browser"). By using the interaction panel on the left hand side, researchers can look for specific visualization techniques and filter out entries with respect to a set of categories (cf. the categorization given in Section 3.1.1). Details for a selected entry are shown by clicking on a thumbnail image in the main view. The survey contains **430** categorized visualization techniques as of February 6, 2019.

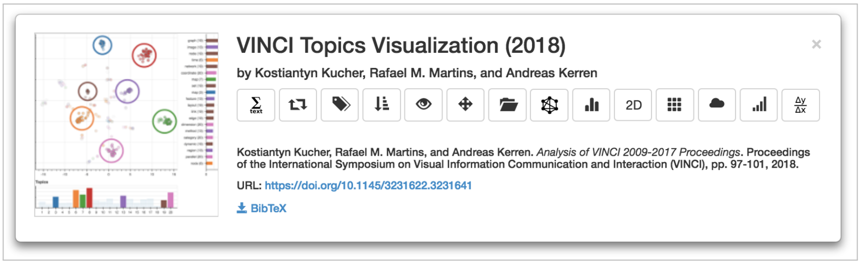


Figure 3.3: Details of a survey entry in TextVis Browser.

3.1.3 Discussion and Analysis

Our decision to design a rather extensive categorization was motivated by the need for fine-grained technique search or filtering as well as for the comparison of entries. We have compared our resulting categorization to the ones described in the related text visualization surveys (cf. Table 3.1). In order to match the categorizations, we have mapped the categories used by the other surveys into several fine-grained categories. We have not included the category “event detection methods” into the comparison, since it is used only by a single, more specialized survey. As displayed in the table, our categorization includes most of the categories except for two: we believe that the underlying data representation (e.g., bag-of-words vs language model [379] or whole text vs partial text [311]) is more relevant to the underlying computational methods than to observable visualization techniques. And the same naturally holds for data processing methods (e.g., the specification of involved MDS methods [7]) that are partially covered by other categories in our categorization, for instance, the analytic task of topic analysis implies the usage of corresponding computational methods. However, we do not negate the possibility of extending our categorization as part of the future work.

Using the data collected for the survey, we have been able to analyze the general state of the text visualization field, to compare the usage of various analysis and visualization techniques (with regard to our categorization), and to analyze the information about researchers in this field. According to our current set of entries, the trend for rapid increase of text visualization techniques started around 2006 (cf. Figure 3.1). With regard to category statistics presented in Table 3.2, there is an obvious interest for tasks related to summarization and topics (66% of all entries). The majority of the techniques support corpora as data sources (80% of all entries), and a lot of them support time-dependent data (47% of all entries). Another result—which is probably expected—is that only 6% of all entries use 3D 3-dimensional visual representations.

Table 3.3: Authorship count distribution for text visualization techniques

#techniques	1	2	3	4	5	6	7	8	9	10	12	13	17	28
#authors	874	135	52	20	15	13	5	3	3	1	2	1	1	1

Note: the current data set includes **430** techniques and **1,126** authors in total.

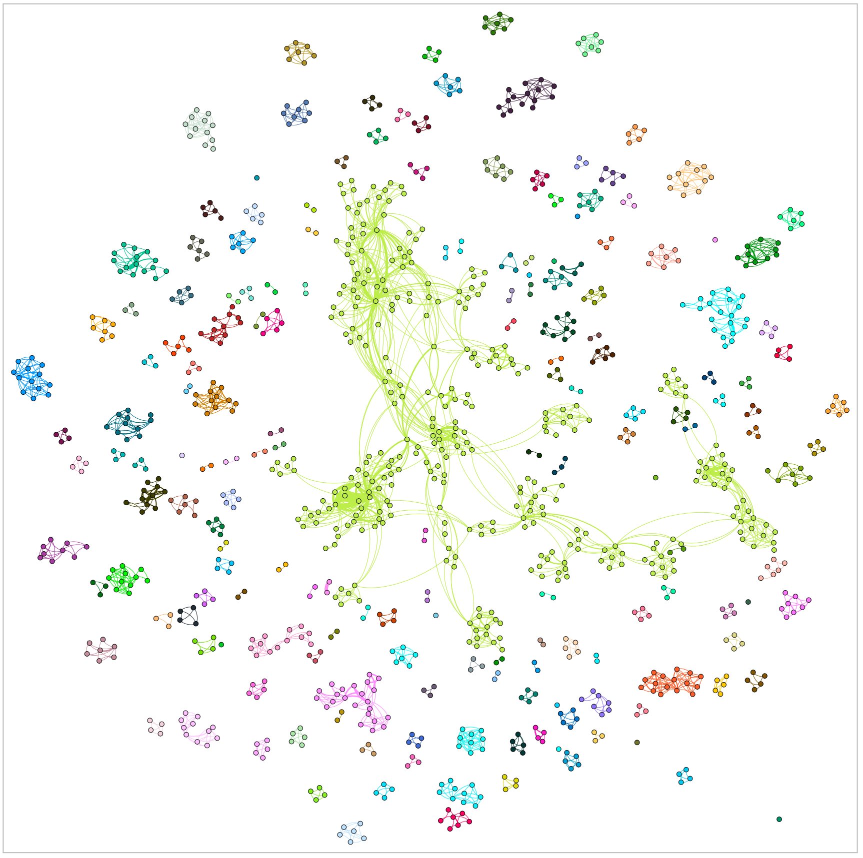


Figure 3.4: Co-authorship network for the text visualization survey entries (as of February 6, 2019) visualized in Gephi with force-directed layout algorithms. Note the giant connected component in the center containing **315** author nodes.

We have also taken a look at the authorship statistics for the current data set. The top five authors with regard to number of techniques are Daniel A. Keim (28 entries), Shixia Liu (17 entries), Christian Rohrdantz (13 entries), Christopher Collins (12 entries), and Huamin Qu (12 entries). As seen from Table 3.3, besides a large number of authors with one or two contributions, the research community includes a core group of researchers with three or more techniques.

After extracting the co-authorship network (1,126 nodes, 2,672 edges), we have analyzed it with Gephi [26]. As seen in the drawing in Figure 3.4 (laid out with ForceAtlas and ForceAtlas2 algorithms), the majority of author nodes are included into isolated connected components of small sizes (less than 15 nodes) while there is a giant connected component with 315 nodes and 1,095 edges present in the graph. The major clusters in that component include the aforementioned authors as cluster center nodes. The fact that research groups from Germany, China, and the United States are so active in the field of text visualization is quite interesting when we set this in relation to the geolocation statistics of visits to our interactive survey browser (according to Google Analytics by mid-February 2019) that list the United States, Russia, China, Germany, and the United Kingdom as the top 5 user locations.

We have also analyzed several network centralities [308] in this co-authorship network. We were mostly interested in the *betweenness* centrality, because it has the largest effect on the research impact in comparison to other centralities in co-authorship networks, as shown by Li et al. [252]. The largest betweenness values in the current network are associated with Jaegul Choo, Niklas Elmqvist, Shixia Liu, Daniel A. Keim, and Ross Maciejewski, who are all active contributors in the text visualization community.

A GMap [143] was also generated to facilitate the exploration of the current co-authorship network. The resulting map is available online[3].

3.2 Sentiment Visualization

The analyses of more than 400 techniques and the corresponding metadata in the previous section provide evidence about text visualization being an active field of research with multiple data domains and tasks involved. Our categorization of text visualization techniques introduced in Section 3.1.1 was influenced by the existing survey articles. However, if we focus more specifically on the existing work in sentiment[4] visualization, we can discover that this subfield has not been covered by any comprehensive survey yet, as discussed below in Section 3.2.1. Therefore, our description of the sentiment visualization design space and a survey of existing techniques can be beneficial for visualization researchers working on this problem—there is already a significant body of work involving sentiment visualization which can be difficult to explore since the publications are scattered across a large number of outlets and disciplines. Our survey can

[3] `http://gmap.cs.arizona.edu/map/9143/` (last accessed in February 2019)

[4] The term *sentiment* is used here and below synonymously with terms like *emotion*, *affect*, *attitude*, and so on, unless a more specific term is required.

also be useful for researchers from other fields as well as practitioners interested in visualization / visual analysis methods for sentiment data.

In this section [240][5], we discuss the design space for sentiment visualization and provide a survey based on the collection and analysis of a substantial number of sentiment visualization techniques described in peer-reviewed papers in InfoVis, VA, and other disciplines (NLP, DM, etc.). To refine our categorization, discover interesting patterns, and facilitate data exploration for the readers, we have developed an interactive survey browser available (as of February 2019) at

`http://sentimentvis.lnu.se`

In this survey, we have limited ourselves only to visualization techniques based on the analysis of text data, and have *not* included visualization techniques related to emotion measurement with the help of brain-computer interfaces (as opposed to emotions discovered in text) or similar approaches. We refer the interested readers to the survey by Cernea and Kerren [67] that covers the corresponding research area. Another related field that concerns itself with analysis and, in some cases, visualization of opinions is social network analysis. For example, Du et al. [109] discuss OpinionRings, a visualization technique for networks with explicit user opinion values. Since such approaches do not involve text data, we have considered them to be beyond the scope of this survey.

The rest of this section is organized as follows. Section 3.2.1 provides a discussion of existing visualization surveys relevant to our work. Afterwards, we discuss our methodology, categorization, and initial statistical results for the collected data in Section 3.2.2. Sentiment visualization techniques are discussed according to the categorization in Sections 3.2.3–3.2.5. We discuss our interactive survey browser, findings, and perspectives for the research field in Section 3.2.6.

3.2.1 Related Surveys

Sentiment visualization has not enjoyed the same level of interest in systematic/comprehensive reviews compared to other visualization areas that are also related to data extracted from text, such as the visualization of topic models or events. Only a few text visualization surveys include analysis and visualization of sentiment/opinion as one of their categorization aspects. An example is the survey on the visual analysis of events in text data streams written by Wanner et al. [458]. They select *polarity extraction* as one of the text processing methods used in visual analytics systems: it was utilized by 14 out of 51 papers included in their report. Several surveys mention *sentiment and affect analysis* as a potential feature extraction method for text visualization, for instance, the paper by Risch et al. [341] in the context of visual analytics (without any examples) or the paper

[5]Sections 3.2 and 3.3 are based on the following publication: Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. The state of the art in sentiment visualization. Computer Graphics Forum, 37(1):71–96, February 2018. John Wiley & Sons Ltd. doi:10.1111/cgf.13217 © 2017 The Authors.

by Šilić and Bašić [379] in connection with text stream visualization (with a single example). Given the total amount of work on sentiment visualization, there is clearly a gap in this area in the visualization survey literature.

To the best of our knowledge, there are only several existing surveys focusing on sentiment visualization techniques. Boumaiza [42] aims to provide a survey of sentiment and opinion visualization techniques with the focus on social media texts. The author does not provide clear criteria and a clear categorization/taxonomy, though. In consequence, numerous techniques are included that are not directly related to sentiment visualization (or even visualization per se); this makes it difficult to navigate the survey and use it for reference. By our estimate, that work mentions around 35 peer-reviewed sentiment visualization techniques. Shamim et al. [371] provide an overview of 11 techniques and classify them according to visual metaphor. However, the focus of that work is not on categorization, but rather on evaluation: the authors conduct a study to compare the techniques with regard to metrics such as user-friendliness or usefulness. In contrast, our survey focuses on the categorization of a much larger number of techniques with regard to multiple aspects related to computational model, data, user tasks, and visual representation.

3.2.2 Survey Methodology

The steps that we took while working on this survey are summarized in Figure 3.5. The overall methodology can be compared to the model described by Pirolli and Card [325], which was adapted to scientific literature analysis by Beck et al. [29].

Based on our previous work on text visualization [236] discussed in Section 3.1, we started with an initial set of text visualization techniques related to sentiment as well as an initial categorization applicable to such techniques. We should state that we use a *technique* as a unit for this survey as opposed to a publication—therefore, we describe several cases below where multiple techniques originate from the same publication. Since our survey includes work not only from InfoVis, but also from VA and even non-visualization disciplines, a single technique does not necessarily mean a novel metaphor/representation, but also an approach or a system relevant to sentiment visualization.

In addition to the initial set of techniques, we have conducted a search in several visualization outlets: IEEE TVCG, Information Visualization, Computer Graphics Forum, IEEE CG&A, and Journal of Visualization as well as proceedings of IEEE InfoVis, IEEE VAST, EuroVis, IEEE PacificVis, TextVis workshop, ACM CHI, ACM IUI, IV, IVAPP, and VINCI. We have also conducted a search in IEEE Xplore, ACM DL, and Google Scholar using such key phrases as "sentiment visualization", "emotion visualization", and "opinion visualization" (considering only literature in English). Finally, we have investigated references from related surveys as well as already detected research publications. We have also con-

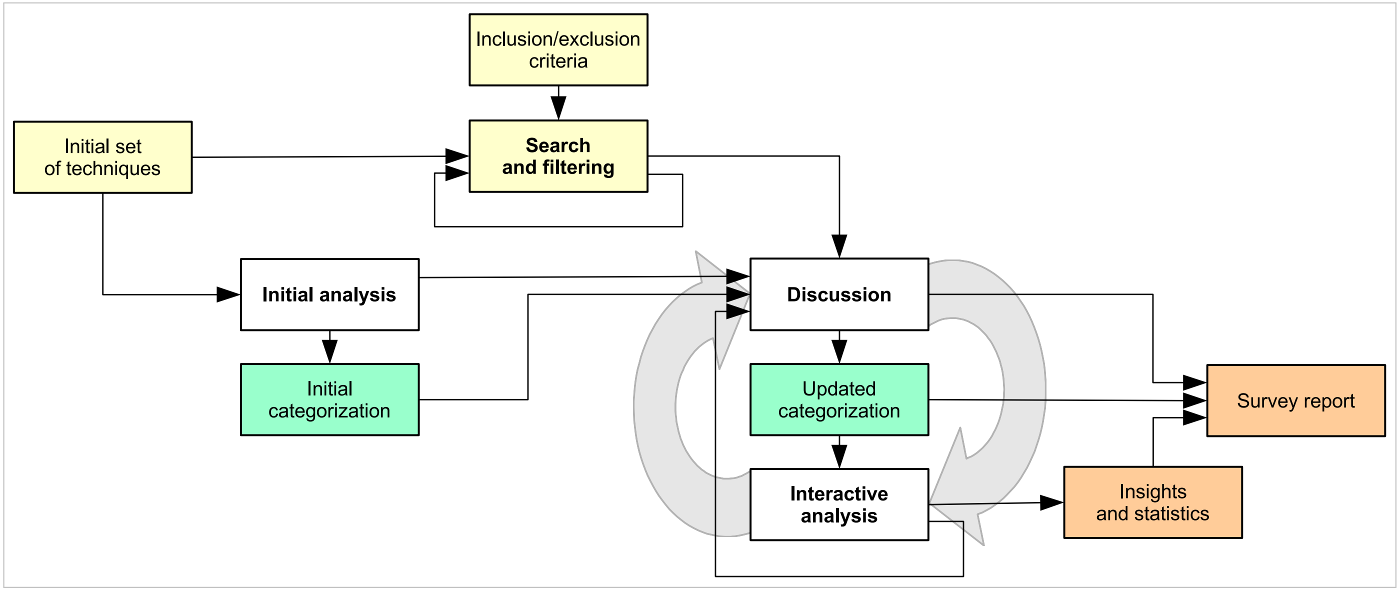


Figure 3.5: Flowchart describing our methodology for the survey of sentiment visualization techniques. Color coding is used to denote the corresponding concept/artifact/activity (yellow for visualization techniques, green for categorization, orange for the survey report, and white for miscellaneous), and bold font is used to denote activities. Interactive analysis of the techniques data, subsequent discussions, and corresponding categorization changes together form a sensemaking loop. *Reprinted from [240] © 2017 The Authors.*

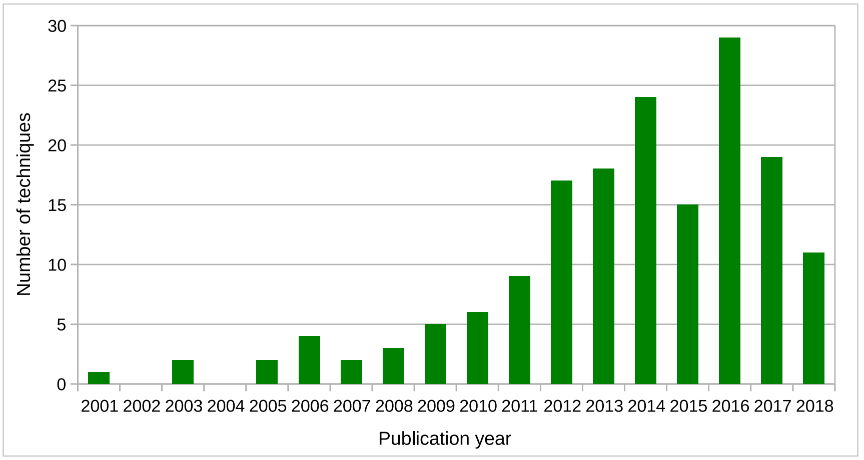


Figure 3.6: Histogram of the collected techniques set (**167** techniques in total as of February 6, 2019) with regard to the publication year.

tinued updating our survey with additional entries after the publication of the corresponding article [240] in 2017.

Selection Criteria We have used the following criteria for including/excluding techniques in our survey:

- a technique must be related to visualization of sentiment associated with text data (either extracted automatically or annotated manually);
- a technique must be illustrated by at least a single figure in the corresponding publication;
- a technique must be described in a peer-reviewed publication (including poster papers and extended abstracts);
- since we are focusing on techniques as opposed to publications, incremental work in several papers by the same authors is not considered as separate techniques; and
- a technique must actually involve an implementation used for visual representation or analysis (possibly even without interactive features) as opposed to figures generated with third-party tools solely for illustrative purposes in the respective publication.

Some of the candidate techniques had to be excluded with regard to the criteria above. Brath and Banissi [46] discuss the usage of text layout and font attributes for the purposes of text-related data visualization and mention sentiment analysis

Table 3.4: Publication statistics for sentiment visualization techniques

Outlet	Number
IEEE TVCG	13
IEEE VAST	10
ACM CHI	9
ACM IUI	6
CGF	5
Inf Vis	5
TextVis	5
VINCI	5
EuroVis	3
IVAPP	3
AVI	3
IEEE VIS	2
IEEE PacificVis	2
IEEE VISSOFT	2
ACM TiiS	2
J Vis	2
VIS4DH	2
ESIDA	2
CDVE	2
CLIHC	2
Others	*10*
Total	**95**

Outlet	Number
AAAI ICWSM	4
IEEE THMS	3
ACM TIST	3
WWW	3
HICSS	3
IEEE NLP-KE	2
IEEE DSAA	2
IEEE ICDMW	2
Inf Process Manag	2
Decis Support Syst	2
WIMS	2
iiWAS	2
Others	*42*
Total	**72**

Note: Statistics are provided with regard to the respective publication outlets in visualization (left) and other (right) disciplines. Information about outlets with a single technique is combined in the penultimate row.

as one of the possible applications, however, the sentiment values are not directly visualized by their technique. OpinionRings by Du et al. [109] use network data explicitly labeled with user opinions as opposed to extracting the data from text—we have also not included Opinion Space by Faridani et al. [126] for the same reason. Li et al. [256] mention a visualization module used in their opinion mining system, but provide no further details or figures, so it is impossible to analyze it. The same applies for the work by Oliveira et al. [316] which lacks any figures of their visualization technique SentiBubbles. Saif et al. [353] discuss SentiCircles, a vector space representation useful for sentiment analysis which can directly be interpreted visually, however, the authors demonstrate this only for illustrative purposes.

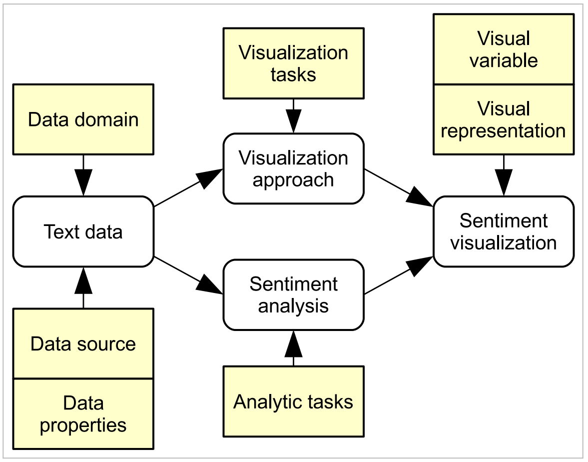


Figure 3.7: The mapping of our categorization aspects (highlighted in yellow) to the sentiment visualization pipeline. *Reprinted from [240] © 2017 The Authors.*

Chosen Publications/Techniques The resulting set of sentiment visualization techniques comprises 167 entries discussed in publications from a wide range of journals and conferences. Statistics for publication outlets in Table 3.4 provide us with an insight that researchers from multiple non-visualization disciplines have demonstrated interest for sentiment visualization (based on the variety of outlets), thus reinforcing our claim for the importance of this research problem. The analysis of temporal distribution (see Figure 3.6) shows that a stable interest for the problem emerged in mid-2000s and strongly increased in the beginning of 2010s.

Categorization The initial version of the categorization was based on our previous work related to text visualization [236]. Inspired by the VA pipeline model by Keim et al. [212], the model presented in Figure 3.7 treats the resulting sentiment visualization as a combination of the general InfoVis approach and computational methods, both applied to text data. The design space aspects used in our categorization (highlighted in yellow) vary from general to specific (left to right). These 7 aspects, or groups, include the total of 35 categories listed in Table 3.5. The categorization facilitates the search for visualization techniques and corresponding publications for interested readers based on their data, required analytic and visualization tasks, and even specific encodings used for sentiment. In the next several subsections, we discuss the individual categories and corresponding prominent examples; for full details on each technique's categorization, see the interactive online browser discussed in Section 3.2.6.

Table 3.5: The complete categorization of sentiment visualization techniques used in our survey

Data Domain	**159**
Online Social Media	100
Communication	17
Reviews / (Medical) Reports	39
Literature/Poems	9
Scientific Articles/Papers	3
Editorial Media	23

Data Source	**167**
Document	26
Corpora	143
Streams	29

Data Properties	**124**
Geospatial	35
Time Series	103
Networks	37

Analytic Tasks	**167**
Polarity Analysis / Subjectivity Detection	128
Opinion Mining / Aspect-based Sentiment Analysis	103
Emotion/Affect Analysis	42
Stance Analysis	16

Visualization Tasks	**167**
Region of Interest	12
Clustering/Classification/Categorization	104
Comparison	155
Overview	160
Monitoring	27
Navigation/Exploration	112
Uncertainty Tackling	15

Visual Variable	**167**
Color	146
Position/Orientation	84
Size/Area	73
Shape	18
Texture/Pattern	5

Visual Representation	**167**
Line Plot / River	66
Pixel/Area/Matrix	70
Node-Link	34
Clouds/Galaxies	45
Maps	38
Text	46
Glyph/Icon	58

Note: Each row contains the number of corresponding visualization techniques in our survey as of February 6, 2019. The percentage relative to the current total of **167** techniques is also illustrated by heatmap-style icons.

3.2.3 Data Aspects

Every sentiment visualization technique relies on certain *data*, and both higher-level (for instance, the original domain) and lower-level (e.g., source representation) data aspects affect the later stages of the pipeline.

3.2.3.1 Data Domain

Most sentiment visualization techniques are designed with a specific data domain in mind, some of which have been historically or currently associated with sentiment analysis and opinion mining tasks in visualization, NLP, and DM communities.

A prominent example of such a domain that provides a lot of text data suitable for sentiment analysis is **Online Social Media** including forums, blogs, microblogs, and social networks. It is no surprise that the majority of techniques in our data (around 60%) support this category. Some of the early examples include MoodViews by Mishne and de Rijke [291], an NLP system for the analysis of affect ("mood") in blogs which uses explicit mood tags provided by users as well as predicted mood levels for at least 132 mood types proposed by the blog platform. The system applies basic line plots for temporal visualization, and allows the users to interactively investigate salient terms and phrases for selected time intervals. Ink Blots by Abbasi and Chen [1] is a technique for exploration of documents and corpora that uses a simple bubble metaphor to mark regions of interest in communication and social media (i.e., forums) texts. With the development of microblogs, more and more techniques started to focus on this data. Diakopoulos et al. [103] describe Vox Civitas, a visual analysis tool for investigation of sentiment, relevance, and salient keywords in social media posts related to public events, which uses a pixel-based stacked bar timeline to represent sentiment values. TwitInfo by Marcus et al. [281] is a visual event analysis system for Twitter streams which supports sentiment polarity detection. The aggregated polarity value is visualized with a pie chart, and the polarity for individual documents is encoded as color of the corresponding markers in a geographical map view. Other examples for this category include BLEWS by Gamon et al. [141], ForAVis by Wanner et al. [456], webLyzard by Scharl et al. [361], MoodLens by Zhao et al. [493], Whisper by Cao et al. [59], OpinionFlow by Wu et al. [480], PEARL by Zhao et al. [492], SentiCompass by Wang et al. [448], and many more.

Communication such as emails and chats can also be subject to sentiment analysis and visualization (used by 10% of techniques). The earliest example in our data is CrystalChat by Tat and Carpendale [413], a visualization technique for personal chat history that represents individual messages as circles and organizes them in a 3D layout with regard to temporal order and chat contact. The technique encodes the emotional content of conversations based on detected emoticons as the background plane color. Another technique for communication texts is described by Gobron et al. [150], and it is a rather unusual emotion visualization technique bordering on computer graphics rather than InfoVis. The authors use the Facial Action Coding System (FACS) to create animated 3D avatars whose faces convey the emotions related to the corresponding text data. Chen et al. [75] conduct visual analysis of chat logs using emoticons classified by valence/polarity. Mail data is supported by Ink Blots by Abbasi and Chen [1], the work of Mohammad [294], 5W Summarization by Das et al. [96], the technique by Guzman [160], and TargetVue by Cao et al. [61].

Before the rise of social media, sentiment analysis was almost solely used to analyze product reviews and customer feedback. The category **Reviews / (Medical) Reports** is supported by 23% of sentiment visualization techniques, including the ones such as Affect Inspector by Subasic and Huettner [407] which uses star plots to visualize affect profiles of text documents, including movie reviews. Pulse by Gamon et al. [140] uses a tree map to represent a clustering of sentence-level sentiment classification of car reviews. Opinion Observer by Liu et al. [261] provides a visualization of customer opinions using modified bar charts. AMAZING by Miao et al. [288] is an opinion mining system for product reviews that visualizes NLP processing results with a line chart (using the review timestamps) and a pie chart (a simple summary for the proportion of positive/negative reviews). In general, a lot of techniques in this category originate from NLP and DM rather than the visualization community, focus mostly on the analytical part, and use rather simple visual representations and interactions. In contrast, Chen et al. [73] use multiple analytical and visualization techniques for investigation of conflicting opinions in customer reviews. Hao et al. [164] present four novel visualization techniques for customer feedback analysis: pixel sentiment geo map, key term geo map, pixel-based sentiment calendar, and self-organizing term association map. Some other examples for this category include the map of opinion clusters introduced in the work by Oelke et al. [312], OpinionSeer by Wu et al. [481], and comparative relation maps by Xu et al. [483]. While this category is not attracting as much interest from the authors of sentiment visualization techniques in the past years as it used to, there are still some works supporting it that were published recently. For instance, Xu et al. [482] propose a visual analytics approach for analyzing controversy and diverging sentiments in reviews. Their implementation visualizes the detected aspect-level sentiments and derived controversy indexes with multiple visual representations including bubble charts, pie chart glyphs, and sunburst charts. Bader et al. [24] introduce Plutchik Radar, a star plot based on the emotions from the Plutchik's wheel model detected in movie reviews. Similar data is also supported by the Emotion Map technique by Topal and Ozsoyoglu [423] that uses a heat map to represent the detected occurrences of four categories of affect: sensitivity, pleasantness, aptitude, and attention.

Literature/Poems is a surprisingly underrepresented domain with regard to sentiment visualization techniques, which includes only nine entries or approximately 5% of our current data set. Subasic and Huettner [407] demonstrate how their Affect Inspector can be applied for affect analysis of a poem by T. S. Eliot. Affect Color Bar by Liu et al. [262] uses a pixel-based metaphor to represent the emotions associated with text document segments. The resulting visualization provides an overview of affective structure as well as navigation over the text. The work of Mohammad [294] focuses on analysis of emotions in mail and books that is based on lexical matching of terms associated with eight emotions as well as POSITIVE and NEGATIVE categories. The author uses multiple basic representations such as line plots, bar charts, and word clouds to analyze the distribution and temporal trends of emotion-bearing word usage in individual documents and corpora. Finally, Weiler et al. [469] apply their text stream analysis and visualization system called Stor-e-Motion for a combined text of the whole "Harry Potter" series. The resulting visualization represents sentiment as a river-like stream graph with an overlay of salient topic terms list for each time interval (in this particular case, position in text is treated as timestamp).

We did not expect to find a lot of work focusing on sentiment visualization in **Scientific Articles/Papers** due to the style standard in this genre. The few existing techniques (2%) detect polarity or stance in text that surrounds citations and use this information for analysis and visualization of citation networks. Schäfer and Spurk [356] classify polarity and reuse of citations in scientific articles and use the classification results for color coding of the citation graph edges. The work of Wang et al. [454] involves polarity classification and visualization of citations to represent a citation graph for paper review purposes.

The last category of data domains is **Editorial Media** such as news or pre-moderated websites (e.g., Wikipedia), and this category is supported by 14% of techniques. Some of the early examples here include SATISFI, "Sentiment and Time Series: Financial Analysis System" by Taskaya and Ahmad [412], which uses lexical matching with specific markers to analyze the polarity of financial news documents. The polarity values are aggregated and treated as time series which are visualized with simple line plots. Fukuhara et al. [137] use line charts to visualize topic data associated with eight affect categories in news and blogs, and vice versa, affect data associated with a specific topic. Gamon et al. [141] describe BLEWS, a system dedicated to the analysis of relations between news articles and political blog posts that refer to such articles. The number of detected subjective posts is encoded visually as the amount of glow around the corresponding bars representing the number of liberal and conservative blog posts. Zhang et al. [490] introduce Sentiment Map, a lightweight visualization based on a geographical map. The tool detects eight emotion categories in news articles over time and visualizes the resulting time series with line plots for corresponding geographical regions based on the zoom level. In contrast to some of these techniques with simple and standard representations, TextWheel by Cui et al. [93] introduces a combination of several complex and novel representations for monitoring of sentiment associated with specific entities in the news streams. Some of the representations used in the system include a radial node-link diagram representing relations between entities (a *keyword wheel*), a U-shaped *transportation belt* that acts as a substrate for moving document glyphs, and a *significance trend chart* that is represented by a line plot.

We have also found several techniques (around 5%) that are not designed for any particular data domain. Duan et al. [111] describe VISA, a system for temporal aspect-based visual sentiment analysis which extends the more general system TIARA with sentiment analysis capabilities. VISA uses a river-like stream graph with embedded tag clouds as well as several auxiliary views (pie charts, bar charts, text views) with multiple interactions to support a number of user tasks. The authors demonstrate VISA with a use case involving hotel reviews data, but their approach is not specific to this data domain. Semantize by Wecker et al. [467] is a lightweight web-based visualization technique that uses font style and background color to encode the word-level polarity, sentence-level subjectivity, and paragraph-level polarity directly in the HTML document. Gold et al. [154] list sentiment polarity as one of the possible annotation types for their Lexical Episode Plots technique. In this case, particular words as well as bar segments representing text document regions can be highlighted to facilitate the understanding of affective structure of the text. Typographic Set Graph by Brath and Banissi [47] provides a map representation of a word-emotion association lexicon. This technique manages to simultaneously represent membership of individual words in ten sets (eight emotion and two polarity categories) by using font style and color. Huang et al. [184] describe a visual interface for facilitating the active learning process for a machine learning classifier. They demonstrate their approach with a data set involving customer reviews and the task of polarity labeling and classification; the provided visual representations (a DR projection scatterplot, a force-directed graph diagram, and a chord diagram) support color coding for the NEGATIVE and POSITIVE categories, however, the approach in general is not specific to these. Finally, ConceptVector by Park et al. [320] is a somewhat related approach focusing on interactive construction of lexicon-based concepts, which involves specification of relevant and irrelevant data instances, computation of word embeddings, and several visual representations such as scatterplots and bar charts. One of the scenarios supported by ConceptVector includes the interactive construction of the concept of HAPPINESS, which is related to emotion/affect analysis.

3.2.3.2 Data Source

The type of text data source has implications for the design of sentiment visualization techniques in most cases. However, sometimes the categorization can become somewhat fuzzy, since a single large document (e.g., a novel or a transcript) can be treated as a collection or sequence of its sections; and vice versa, multiple separate texts can be concatenated for analysis and visualization.

First of all, we have identified techniques that support data from an individual **Document** (16% of the total set). In most of such cases, visualization of a single document is used either to support details on demand, or to represent a subset of data from another type of data source. For instance, Affect Inspector by Subasic and Huettner [407] can visualize affect profiles for a single document or several documents at a time. Ruppert et al. [347] provide visual summaries of subjectivity and polarity analysis for selected individual documents in their system PolicyLine. Techniques by Gobron et al. [150] and Krcadinac et al. [229] support either a real-time conversation text stream, or its transcript as a single document.

In some cases, though, the focus of the visualization technique is set on a single document. Besides Affect Color Bar [262], Semantize [467], and Lexical Episode Plots [154] discussed above, Li and Ren [254] visualize 2D/3D heatmap visualizations of emotion matrices for individual text documents. AmbiguityMatrix by Stoffel et al. [403] focuses on the task of interactive refinement of named entity recognition (NER) results in works of literature. It supports, among others, analysis and visualization of Plutchik's basic emotions for the text-based context information associated with fictional characters (using a simple pixel-based representation), which facilitates the task of comparing the emotional profiles of characters to decide whether their entries should be merged. Gomez-Zara et al. [155] propose an approach for analysis of individual news articles in order to detect the main story actors/entities fitting the roles of the hero, the villain, and the victim. Their approach combines NER and polarity detection methods; the detected entities and top relevant terms are displayed alongside simple glyphs color-coded based on the role. Finally, Story Explorer by Kim et al. [222] supports visual analysis of individual movie script documents combined with additional metadata. The main focus of this approach is on visualization of non-linear narratives using *story curves*, a representation similar to the storyline metaphor. The polarity of each movie character's dialog lines in each scene is analyzed, and it can be used for color coding of the story curve segments, among other options.

The absolute majority of techniques in our data set (86%) support text data from a collection of documents, a corpus, or corpora. In contrast to the visualization of individual documents, techniques that support such data sources typically have to address challenges related to larger data set sizes, varying text lengths, relationships between documents, and additional data properties discussed below. A typical example here would be a technique oriented at a collection of customer reviews [483] or previously fetched tweets [480]. The corresponding category **Corpora** includes techniques spanning the complete time range of our survey, from Affect Inspector [407], Opinion Observer [261], and Pulse [140] to OpinionSeer [481] and Vox Civitas [103] to ConVis [180], SocialHelix [60], and SemEval-2016 data set visualization [296].

During the past decade, research in streaming data visualization has also produced a number of sentiment visualization techniques which support **Streams** as data sources (17% of our collected set). Most of these techniques consume data from microblogs such as Twitter or Weibo, for instance, TwitInfo by Marcus et al. [281] or MoodLens by Zhao et al. [493]. In some cases, the major focus of a technique is set on event detection, and sentiment analysis/visualization plays an auxiliary role. For example, Krstajić et al. [230] describe a visual analysis system for event detection in Twitter data which uses the aspect-based sentiment analysis method introduced by another technique [345] to calculate polarity scores for individual documents among other features. StreamExplorer by Wu et al. [479] focuses on event detection and topic summarization of Twitter data streams with multiple representations and views for visual analysis, while the polarity of individual tweets is classified and represented as part of glyphs in one of the views.

Other streaming data visualization techniques treat sentiment as the first class data. Liu et al. [263] use a river representation for sentiment polarity values extracted from a Twitter stream with several co-trained classifiers. Kranjc et al. [228] describe the usage of a general-purpose cloud-based platform ClowdFlows for sentiment analysis and visualization. It detects the polarity of streaming Twitter data by using an *active learning* classifier and visualizes it with standard line plots, stream graphs, and word clouds. Matisse by Steed et al. [401] is a VA system for text stream data that focuses specifically on temporal sentiment analysis of Twitter data. Polarity of individual tweets is visualized with stream graphs and geospatial heat maps, and a fine-grained analysis of emotions in the text data is available with a scatterplot representation. PaloPro by Tsirakis et al. [429] is a brand monitoring platform which conducts opinion mining of data streams from multiple social media and news sources. Its dashboard visualization includes line plots and bar charts representing polarity for specific topics or named entities.

3.2.3.3 Data Properties

Besides the data source type, we have also analyzed special properties of the data used by sentiment visualization techniques. The results of our analyses discussed in Section 3.2.6 confirm that these properties are correlated with the later stages of the sentiment visualization pipeline.

Some of the relatively recent techniques make use of **Geospatial** information (21% of techniques), starting with Sentiment Map by Zhang et al. [490]. This tool identifies affective content (namely, eight emotion categories) in news articles. The resulting temporal emotion values are visualized with line plots for corresponding geographical regions based on the zoom level. Scharl et al. [361] present webLyzard, a platform for monitoring and visual analysis of social media, news, and other text documents from the Web. Among other analyses, it supports polarity detection for specific topics and uses these polarity values with line charts, a map, and a tag cloud. Nguyen et al. [309] conduct both dictionary- and ML-base polarity detection analyses on a geotagged Twitter data set and visualize the results using choropleth maps, treemaps, and line charts. Mitchell et al. [292] describe somewhat similar analyses, but their approach focuses on the range of emotions between HAPPINESS and SADNESS detected in tweets from various regions of the United States; maps as well as bar charts are used to represent the results. One of the techniques described by Zhang et al. [489] provides a geospatial visualization of sentiment in social media. The technique combines a regular geographical map with the representations of kernel density estimation (KDE) analysis for sentiment polarity in specific cities and edges between cities representing the retweeting network. Caragea et al. [63] combine geospatial information present in tweets with polarity analysis results to study social media patterns emerging during disaster scenarios. Finally, Hundt et al. [187] provide a visual analytics approach for Twitter data that uses the geospatial information to position glyphs representing individual tweets on the map view; polarity-based color coding is applied to these glyphs as well as other visual representations used in the tool.

The majority of techniques in our survey (62%) support temporal data, denominated by the category **Time Series**. Most of such techniques support overview of sentiment value series over time with representations such as standard line plot [81,96,177,291] or river/stream graph [111,263,480,492]. Other popular approaches to visualize temporal sentiment data involve glyphs aligned along a timeline [271,457,459] and animation-based methods [115,139,489].

Finally, we have identified the techniques that use **Networks** present in input or processed data (22% of the total set). The examples of such networks include citation graphs [356,394,454] and graphs of relationships between terms [45], comments [78], users [447], communities [60], and named entities [360]. Hoque and Carenini [180] discuss ConVis, a visualization system for discourse analysis of social media discussions that uses a network of users, topics/concepts, and opinions. The resulting visualization combines a radial node-link diagram with an indented sequence of stacked bars that represents the conversation tree. The subsequent work of the same authors on ConVisIT [181] integrates support for interactive topic modeling, and MultiConVis [182] extends the analysis to multiple discussions and temporal data. A similar network of users, entities/concepts, and opinions is also visualized with ORCAESTRA by Prasojo et al. [329]. The authors propose a planetary metaphor for their main node-link representation which represents nodes as stars, planets, and asteroids, and uses a heliocentric radial layout. Asteroid nodes represent individual comments and use color to encode the corresponding sentiment polarity.

3.2.4 Tasks

The process of designing a sentiment visualization technique includes the analysis of intended data, intended audience, and intended tasks. We focus on the latter aspect in this subsection and analyze the tasks related to both the computational and visual/exploratory methods.

3.2.4.1 Analytic Tasks

High-level analytic tasks supported by sentiment visualization techniques are based on the respective sentiment analysis models introduced in Chapter 2.

Polarity Analysis / Subjectivity Detection is the most common analytic task in our survey associated with 77% of techniques. The techniques which support only this analytic task provide a summary about the overall polarity of text data, for example, SATISFI by Taskaya and Ahmad [412] or Vox Civitas by Diakopoulos et al. [103]. Annett and Kondrak [17] extend a blog visualization tool eNulog with sentiment analysis by classifying movie blog posts into positive, negative, and neutral/uncertain and using these labels for color coding of the blog map nodes. Pupi et al. [330] highlight the polarity of individual sentiment-bearing terms in their system Ent-it-UP. Agave by Brooks et al. [53] is one of the few sentiment visualization systems that focus on collaborative visual analysis of social media data. The system visualizes the aggregated polarity of temporal text data using representations such as line plots and stream graphs. The final example in this category is BLEWS by Gamon et al. [141], a system dedicated to the analysis of relations between news articles and political blog posts that refer to such articles. The characteristic detail of BLEWS is that it focuses only on detecting subjectivity in the blog posts without more detailed polarity analysis. The number of detected subjective posts is visually encoded as the amount of glow around the corresponding bars representing the number of liberal and conservative blog posts.

Another category with high support in our data set (62%) is **Opinion Mining / Aspect-based Sentiment Analysis**. We have used this item for techniques supporting sentiment analysis at the level of particular aspects/features, topics, named entities, or clusters detected in text. This task has often been supported historically with customer reviews data. For instance, Review Spotlight by Yatani et al. [485] is a simple visualization tool for summarizing customer reviews with a tag cloud of salient adjective/noun word pairs. The sentiment polarity for each word is calculated with lexical matching over its counterparts and used for color coding. OpinionBlocks by Alper et al. [11] provides an aspect-based sentiment overview for customer reviews. The visualization combines multiple coordinated bar charts and text tags that can be explored interactively to investigate the polarity and salient keywords associated with specific product features. SentiVis by Di Caro and Grella [102] visualizes results of aspect-based sentiment analysis of customer reviews. After selecting a specific aspect, the users are provided with a visual representation of polarity scores for review objects (in this case, restaurants) that combines a scatterplot and a line plot. Görg et al. [156] discuss the text analysis and visualization features added to Jigsaw, a general-purpose VA system. The support for sentiment analysis includes document-level polarity detection, which can also be used for analysis of specific aspects when combined with topic analysis, as demonstrated with car reviews data. The visual representations involving the polarity data include word clouds and pixel-based Document Grid Views that can encode document-level polarity. The recent E-Comp approach by Wang et al. [453], which is designed for comparison of customer reviews, includes an augmented word cloud with clustered adjective/noun word pairs inspired by the work by Yatani et al. [485] mentioned above; the polarity of such word pairs is encoded with color.

Another prominent application of techniques in this category is the analysis of social media texts. Wensel and Sood [470] describe several basic visualizations of sentiment with regard to the specific topics discovered in personal blog posts with their system VIBES. The authors estimate valence of the texts (hence their claims about analyzing the emotional content of the text data—nevertheless, valence detection on its own is similar, if not equivalent, to polarity detection) and use this data in line plots, glyphs based on the gauge metaphor, and tag clouds. Sentiment Card by Cervantes et al. [69] provides a real-time summary about a topic of interest in Twitter by detecting the polarity of corresponding tweets, and the follow-up work by Valdiviezo et al. [438] also represents the relations between topics with regard to the polarity in the system called SCWorld. Mahmud et al. [277] focus on specific aspects of the Twitter users' attitude useful for brand analysis such as favorability or resistance and provide the corresponding visual summary.

Emotion/Affect Analysis is related to analysis of affective content in text beyond the positive/negative categories, usually involving a categorical or dimensional emotion model. Approximately 25% of techniques in our survey support this task. Affect Inspector by Subasic and Huettner [407] uses star plots to visualize affect profiles of text documents. The authors combine lexical matching with fuzzy tagging to label documents with 83 affect categories (including basic emotions) and support visual exploration and annotation with their tool. Gregory et al. [157] conduct analyses of affect bearing words in customer reviews using lexical methods, an annotation tool, and a general-purpose visualization system IN-SPIRE. They propose a novel metaphor based on a rose plot to visualize statistics for eight affect categories. The users can investigate the rose plots for the corpus in general as well as particular clusters of reviews. Kang and Ren [209] conduct a joint emotion/topic analysis in blog posts. They analyze the output of their method by examining node-link diagrams generated for topic networks that use color coding based on one of the eight emotions dominating for the respective topic. Zhao et al. [492] support visual analysis of emotions detected in tweets over time for an individual user with their system PEARL. They use two emotion models: the categorical Plutchik's model with eight basic emotions and the dimensional VAD model. PEARL uses river-like stream graphs as the main visual representation for temporal emotion data alongside multiple auxiliary representations (line and area charts, glyphs, scatterplots, and tag clouds) to support overview and detailed exploratory analysis. Wang et al. [448] discuss a technique for emotion analysis of Twitter data called SentiCompass. Its visual representation uses star plots that naturally correspond to the polar valence-arousal space and organizes them in a nested fashion resembling a spiral to support the simultaneous visual analysis of multiple temporal intervals. Additionally, the same data is represented by linearly ordered line plots as an auxiliary representation. Emotion analysis is also supported by the recent work of Cuenca et al. [92] on MultiStream, a technique which facilitates exploration of hierarchical time series with a stream graph by using *focus+context*. While the approach is not specific to sentiment visualization, the authors demonstrate a successful use case involving multiple time series based on emotion classification results for tweets.

The final category in this task-related group is **Stance Analysis**, which is discussed below in Section 3.3.

3.2.4.2 Visualization Tasks

Besides the higher-level analytic tasks related to sentiment analysis model, we have included a number of more concrete representation and interaction tasks that are directly supported by sentiment visualization techniques.

For example, the category ★ **Region of Interest** includes techniques which support automatic detection and highlighting of interesting or anomalous items [61, 93, 272, 491, 492], regions [345], or peaks [281, 361] (overall, 7% of the total set of techniques). Time Density Plots by Rohrdantz et al. [345] allow the users to focus on sequences of subjective text documents related to the specific features (aspects) in customer reviews and RSS news streams. The technique supports visual analysis of interesting data regions and automatic extraction of such patterns. The actual visual representation combines a series of bar glyphs with an area chart for the corresponding time region. TargetVue by Cao et al. [61] uses sentiment polarity values to analyze and visualize anomalous user behavior analysis in social media and email communications, employing glyphs to represent detected anomalous users.

A lot of techniques in our survey (62%) are related to **Clustering/Classification/Categorization**. Here, we have identified the techniques that involve additional (semi-)automatic tagging or grouping of data elements *besides* the actual sentiment classification, represented or facilitated by the visualization. The work by Oelke et al. [312] on opinion analysis for customer reviews uses several techniques for visual analysis of aspects/features, including a matrix-based visual summary report and a circular correlation map (a sort of combination of a node-link diagram and a parallel coordinates plot). The authors also introduce a novel technique for the visual analysis of opinion clusters that combines a Voronoi diagram with thumbnails containing cluster details as tables. Brew et al. [49] support the temporal sentiment polarity analysis for groups of Twitter users with their SentireCrowds system. They cluster users into groups using tweet contents and calculate the aggregated polarity values for each cluster / time step. SentireCrowds provides an overview of the overall sentiment over time with an area chart as well as multiple treemaps for individual time steps which represent the polarity and salient keywords for user clusters. Kim and Lee [221] discuss a dimensionality reduction technique called Semi-Supervised Laplacian Eigenmaps which they apply to the customer reviews data. The method involves extraction of features (terms) related to POSITIVE/NEGATIVE categories and dimensionality reduction of the feature space based on graph and matrix computations. The resulting 2D embedding of reviews, which highlights the clustering in the data, is visualized with a color-coded scatterplot. ToPIN by Sung et al. [408] identifies topics in student comments by using a clustering algorithm and then represents them as nodes. The average polarity of comments belonging to the same cluster is encoded with node brightness.

Comparison of several entities is facilitated by most visualization techniques in our survey (93%). For instance, Xu et al. [483] propose a method for opinion analysis of product reviews that directly takes the comparisons of particular features into account. The resulting visualization uses a node-link diagram to represent a probabilistic graphical model as a bipartite graph, where the polarity and direction of comparison is displayed for each feature. Kuksenok et al. [245] use a timeline visualization to represent occurrences of several affect categories in their annotated data set and to identify relationships between such categories. One of the techniques used in the SocialBrands system by Liu et al. [268] is BrandWheel, which visualizes the scores of various brand aspects estimated by the analysis of social media posts and employees' reviews. The system is capable of simultaneously displaying two BrandWheels for comparison purposes, and even visualizing their differences as a derived representation. A similar comparison mode was previously used in EmotionWatch by Kempter et al. [215] to compare emotional reactions to a pair of topics or named entities on Twitter. Lariat by Chen et al. [74] is designed for visual comparison of Twitter queries using multiple attributes, including polarity analysis results. Finally, Gao et al. [145] describe an interactive interface which positions the emotion classification results (represented with bubbles and bar charts) of Reddit comments by the supporters of two competing political candidates side by side for comparison purposes.

Another task supported by the absolute majority of techniques (96%) is **Overview**—"The big picture" achieved by displaying (1) a significant portion of the data set or (2) aggregated representations. The former (1) is the case for techniques such as City Sentiment by Wu et al. [476], which displays a grid of bubble glyphs representing Chinese cities. Each glyph encodes the number of Weibo posts from the respective city as well as the overall detected polarity. BrandSediments introduced in the SocialBrands system by Liu et al. [268] provide an overview of brand traits/aspects such as SINCERITY and EXCITEMENT by laying out clouds of bubbles (which represent individual brands) along the corresponding axes. (2) Aggregated representations are used to provide overview by techniques such as AMAZING by Miao et al. [288] or Westeros Sentinel by Scharl et al. [360], which include pie charts to provide a summary of polarity values distribution. SentiWheel by Gali et al. [139] provides daily summaries of the sentiment on several Canadian banks in social media posts with a variation of a sunburst diagram which is also capable to display relations based on shared keywords.

Few of the sentiment visualization techniques (16%) support the task of **Monitoring**, which could be described as dynamic observation and alerts on changes in the data. Some of the techniques use continuously updated timeline metaphor to represent the changes in data [228, 230, 469], others involve dedicated representations heavily relying on animation [115, 139, 150, 229] or combine multiple representations [93, 281, 360]. Calderon et al. [55] discuss how combining several visual representations can facilitate the monitoring task for streaming sentiment data.

We have used the category **Navigation/Exploration** for the 67% of techniques which support interactive exploration of the data. Most of such techniques support details on demand, typically by providing the user with a detailed (text) view for the selected timeline [53, 111] or map [255, 272] region. Another option is to display specific analysis details for a selected/hovered item [78, 164, 425].

Finally, we have added the category **Uncertainty Tackling** for sentiment visualization techniques that make use of uncertainty present in data and/or computations (currently supported by roughly 9% of techniques). In several works, UNCERTAIN [17], UNDEFINED [356], or CONTROVERSIAL [103] sentiment is used as a separate category besides NEUTRAL. Gamon et al. [141] use the certainty level of their subjectivity classifier to filter out blog posts which should not be represented in the visualization with the threshold of 50% certainty. OpinionSeer by Wu et al. [481] is a complex system for the visual analysis of customer feedback that extracts polarity values associated with certain features (aspects). The further computational steps include uncertainty modeling and aggregation of opinions based on a model called "subjective logic". The authors provide a visualization that combines multiple radial views as well as a scatterplot laid out inside of a triangle to represent temporal, geospatial, aspect, sentiment, uncertainty, and customer profile data at multiple levels of details. Rohrdantz et al. [345] estimate the uncertainty of their sentiment analysis method and encode such values directly in their Time Density Plots alongside the polarity and data entry counts. The work of Makki et al. [278] focuses on sentiment lexicon expansion/annotation. Their aspect-based sentiment analysis model predicts the polarity values of individual terms. The users can then validate and edit the predicted sentiment values by using several visual representations: a tree cloud and a scatterplot with embedded word clouds. Uncertainty in this case is related to the ambiguous polarity predictions, which are highlighted in yellow in the tree cloud representation. The VA framework for event cueing by Lu et al. [272] involves the polarity analysis of RSS news messages based on ML methods. The authors address the task of uncertainty visualization with stream graphs: stacked layers encode the volumes for several sentiment classification certainty levels over time. The previous work from the same authors [270] also uses blurred map glyphs to represent uncertainty of Twitter data polarity analysis for disaster scenarios. Blur or fuzziness is also used to represent uncertainty of emotion classification results in the recent CrystalBall system by Cho et al. [80] that focuses on event discovery and prediction in Twitter data. Finally, TExVis by Humayoun et al. [186] uses opacity of chord diagram arcs to represent confidence of polarity classification results for tweets.

3.2.5 Visualization Aspects

The final two groups of our categorization are related to specific aspects of representing sentiment data visually.

3.2.5.1 Visual Variable

One of the interesting research questions that we aimed to answer with this survey was related to how sentiment values are usually represented by various techniques. Based on the work by Bertin [35], we have introduced several categories for the visual variables used to encode sentiment.

By far the most popular choice is **Color**, used by 87% of techniques. Most techniques that focus on polarity values use green for positive and red for negative polarity [429, 470], although several techniques reverse this color map to use green/blue as a cold hue corresponding to negative and orange/red for positive [55, 75, 135, 180, 182, 476, 477, 488]. The techniques related to emotions or affect categories other than positive/neutral/negative either use an ordinal set of colors representing categories [115, 209, 492] or interpolate the colors based on the corresponding dimensional model [10, 448]. In some cases, opacity is also used to conduct the level of confidence or uncertainty related to sentiment classification [186].

Position/Orientation holds the second rank with the support of 50% techniques. The most obvious example here would be a line plot whose vertical axis corresponds to the aggregated polarity value [69, 412]. Vista by Hoeber et al. [177] uses line plots to represent counts of positive, neutral, and negative tweets over time. In other cases, horizontal position is used to encode the polarity [77, 278]. For instance, Eventscapes by Adams et al. [2] is a system for visual analysis of events, topics, and emotions (moods) in RSS news feeds. The main visual representation of Eventscapes uses a timeline metaphor with document thumbnails laid out linearly according to their timestamps and valence values (vertical and horizontal axis, respectively). ClasSense Morale Graph by Jiranantanagorn and Shen [202] uses the vertical axis of a temporal bubble chart to represent the average polarity of student comments made during an online lecture. Also, several other techniques which use a gauge metaphor rely on orientation rather than position [470, 495]. For example, the work by Kherwa et al. [220] includes gauge glyphs representing the aspect-level polarity detected in customer reviews.

Size/Area is used to represent sentiment by 44% of techniques in our survey. The basic examples include the techniques which use bar charts [11, 296], pie charts [96, 281], bubble charts [179], area/stream charts [228, 263], or filled star plots [215, 304]. More unusual cases in this category include BLEWS by Gamon et al. [141], which encodes the subjectivity level with amount of glow, and Semantize by Wecker et al. [467], which adjusts the font size of intensifying and diminutive words detected in text.

The remaining choices of visual variable are much less popular. **Shape** is used by 11% of techniques. In several cases, the corresponding techniques use a glyph [262] or glyph-like [150] representation. The work by Lee et al. [250] on analysis of financial blog posts uses simple glyphs to represent polarity values in text lists alongside line plots and bar charts. Tour-pedia by Cresci et al. [90] analyses customer reviews of touristic locations and visualizes them by embedding smiley glyphs (based on the average review polarity) in a map. Smiley glyphs associated with five levels of polarity (and double-encoded with color) are also used as one of the representations in the NewsTone tool by Harris [165] to indicate the sentiment of news articles. Kim and Lee [221] use several marker shapes as well as color coding to differentiate between labeled/unlabeled positive and negative reviews in their scatterplot representation. The basic marker shapes such as triangles and diamonds are also used in the Storyteller tool by van Meersbergen et al. [439] to represent categories in bar charts and scatterplots, including multiple categories of affect. Other techniques use the contour of a heat map or similar representation to convey emotion values [10, 254]. Kempter at al. [215] conduct an emotion analysis of social media texts with 20 emotion categories. Their visualization system EmotionWatch uses a star plot with filled area to represent the emotions detected for the selected time interval. Munezero et al. [304] focus on the temporal emotion analysis and visualization for an individual Twitter user with their tool EmoTwitter. One of the visual representation used for the resulting eight basic emotions is a star plot with filled area.

Finally, we have discovered that only five techniques in our complete data set (3%) use **Texture/Pattern** to represent sentiment. Gali et al. [139] introduce three techniques involving polarity detection in timestamped social media posts related to five Canadian banks. One of these techniques, Emotional Tapestry, generates monthly summaries as woven patterns different for various sentiments, which can be combined for several banks at a time. Zhang et al. [489] introduce a visualization technique for temporal visualization of social media sentiment that is based on the Electron Cloud Model. The authors calculate the polarity of posts over time for individual users and then visualize this data using a special layout algorithm, which results in a texture-like rendering of trajectory lines. Kuang et al. [233] describe ImgWordle, a visualization tool that is designed for social media monitoring. One of its visual representations is a choropleth map that provides an overview of frequency and sentiment polarity of posts for each region. The authors use textures to represent sentiment, since the color channel is used to convey topic data. Finally, Krcadinac et al. [229] propose an artistic visualization technique for chat conversations called Synemania. They analyze the emotions present in chat messages using the Ekman's six emotions and then use this data in an animated particle simulation. The visualization can therefore be used for monitoring of emotions during a conversation by observing the overall texture and colors.

3.2.5.2 Visual Representation

The last group in our categorization includes visual representations (or metaphors) that make use of sentiment. The statistics in this group are not so heterogeneous as for visual variable, for instance: the categories described below are supported by 20% to 42%. The majority of techniques in our survey (69%) have more than one category assigned—in many cases, novel or complex representations combine several traits, or multiple coordinated views are used to represent data that includes sentiment information.

The first category in this group is **Line Plot / River**, supported by 40% of techniques. Basic line plots/charts have been used primarily for temporal data in sentiment visualizations since earlier works such as SATISFI [412], MoodViews [291], and the work by Claster et al. [81] up to the recent techniques such as Westeros Sentinel [360] and TSViz [340]. We also include the techniques that use area charts in this group since they usually convey the same data as line plots, for example, the work by Fukuhara et al. [137] or Lingoscope by Diakopoulos et al. [104]. Then there are river-like representations [54, 167] that range from a simpler stacked area chart used, for instance, in VisTravel by Li et al. [255], to stream graphs used in VISA by Duan et al. [111], PEARL by Zhao et al. [492], and OpinionFlow by Wu et al. [480]. The flow metaphor is also used by a recent work by Wang et al. [452]. Their system called IdeaFlow combines the ideas of storyline visualizations and topic modeling to facilitate the lead-lag analysis of ideas discovered in social media; the results of opinion mining are integrated in the visual representation to convey polarity towards certain ideas. Finally, we also include several techniques in this category which do not use a standard line plot per se, but rather representations which consist of line segments or contours drawn in a certain coordinate system. The corresponding examples include a star plot used in Affect Inspector by Subasic and Huettner [407], a line plot variation used in SentiVis by Di Caro and Grella [102], and a parallel coordinates plot [173] used by Zhao et al. [494].

The second category in the visual representations group is **Pixel/Area/ Matrix** (used by 42% of techniques). Here, we have tried to collect the techniques which use space-filling approaches and other representations which rely on the size/area variable. One of the basic representations here is a pie chart: for instance, AMAZING by Miao et al. [288], TwitInfo by Marcus et al. [281], and 5W Summarization by Das et al. [96] use pie charts to provide an overall summary about the polarity distribution. Kumamoto et al. [246] describe an emotion detection tool for personal Twitter data that uses six emotion categories organized in polar pairs. The tool presents a simple visualization consisting of line charts and pie charts for the individual user's data over time. Many other techniques use various forms of bar charts, including regular histograms used in FAVe by Guzman et al. [161], VisOHC by Kwon et al. [247], and GeoSentiment by Pino et al. [324], and stacked bar charts used by Dehiya and Mueller [101] and Mohammad et al. [296]. Wanner et al. [456] introduce a visualization system for internet forum data called ForAVis. The main visual representation used by the authors comprises pixel-based stacked bars that are organized and colored according to the selected set of features, including sentiment polarity. Another example of this category is a tree map, which is used in Pulse by Gamon et al. [140] and SentireCrowds by Brew et al. [49]. Wu et al. [477] use a cartogram for their system City Flow to represent sentiment of Weibo posts originating from specific cities. The sizes of nodes represented by squares are calculated similarly to a regular tree map, but the layout takes the geographical positions of cities into account.

Various forms of **Node-Link** representations are used by sentiment visualization techniques (20%) on their own, e.g., in works of Kang and Ren [209], Small [394], or Makki et al. [278], or in combination with other representations such as a map (e.g., in Whisper by Cao et al. [59] or the work by Zhang et al. [489]) or a (stacked) bar chart (for instance, in the work by Xu et al. [483] or the works by Hoque and Carenini [180,182]). Several techniques use arc diagram variations, e.g., the work by Chen et al. [73] or News Flow by Braşoveanu et al. [45]. Arc diagrams are also used in the recent work by Fu et al. [136], whose system iForum uses multiple representations for visual analysis of MOOC forum data. Individual threads are represented by a combination of a river and an arc diagram called Thread River, and this representation can be user-configured to display sentiment analysis results.

We have used the next category for representations that use multiple visual items (such as dots, bubbles, glyphs, or words/tags) to give rise to associations with **Clouds/Galaxies** (supported by 27% of the techniques set). Some of the techniques in this category use various forms of word/tag clouds, for instance, Review Spotlight by Yatani et al. [485] and webLyzard by Scharl et al. [361]. Fisheye Word Cloud by Wang et al. [449] is a technique for temporal sentiment visualization that detects the polarity of individual key terms in a set of Twitter posts. The visual representation is based on a word cloud whose layout takes the temporal order into account. The technique makes heavy use of *focus+context* for interactive exploration. Other techniques rely on clouds of dots or similar markers, e.g., RadViz-based Attribute Astrolabe used in SentiView by Wang et al. [447] and the work by Kim and Lee [221]. Opinion Zoom by Marrese-Taylor et al. [283] provides a lightweight visualization of customer reviews using an aspect-based sentiment analysis model. The system uses basic visual representations such as bar charts and bubble charts. Lu et al. [271] describe a VA framework for classification and prediction that uses sentiment analysis to predict movie gross based on social media texts and movie reviews. Polarity values are used in several visualizations used in the framework, namely, word clouds and temporal bubble charts. Hohman et al. [179] also use a form of a temporal bubble chart to represent counts of affect-bearing words in the subtitles for the "Game of Thrones" TV series. Finally, the galaxy metaphor is also used in a more literate sense in CosMovis by Ha et al. [162], a system for emotion analysis in movie reviews which displays a constellation map of movies positioned according to the specific affect-bearing words detected in reviews. The visualization also contains the centroids of corresponding word clusters and artistic representations of emotions as constellations.

The category **Maps** includes techniques (23%) which use either (1) an actual geographical map or (2) an abstract map which somehow allows the user to identify interesting regions or peaks in the overall landscape. In the former case (1), the maps are often augmented with markers (e.g., in TwitInfo by Marcus et al. [281]) or overlays (e.g., in the work by Zhang et al. [489]). The visualization of public opinions on educational institutions in eduMRS–II by Qiu et al. [332] provides a map with overlaid circles/bubbles which represent the aggregated positive or negative polarity. A similar approach is used by Dai and Prout [95] to represent the aggregated positive sentiment on the Super Bowl teams extracted from Twitter. Tweetviz by Sijtsma et al. [378] uses a map view to represent business locations reviewed in tweets and encodes the review polarity with the map marker color. Some techniques use the choropleth approach, for instance, MoodLens by Zhao et al. [493] and ImgWordle by Kuang et al. [233]. The work by Yu et al. [488] uses a choropleth map among other representations to display sentiment analysis results for Weibo microblogs in China. Zhao et al. [495] introduce Social Sentiment Sensor, a visualization system for monitoring topics and emotions in microblog streams. The resulting values are aggregated for a specific topic or geographical region and visualized with line plots, pie charts, bar charts, choropleth maps, and gauge-based glyphs. (2) Some of the abstract map examples in this category include heatmap-like representations. Sentimap by Hennig et al. [175] creates a heatmap for temporal sentiment data based on specific Twitter search terms (hashtags). A later technique, Cluster Heat Map [174], uses an algorithm based on dimensionality reduction to change the vertical layout of the heatmap. This facilitates the discovery of patterns and clusters in the data. Finally, some techniques use a combination of ideas, for instance, Hao et al. [164] as well as Steed et al. [401] create a pixel-based heatmap on the top of a geographical map. In contrast, GeoSentiment by Pino et al. [324] uses a regular heatmap in the overlay, which is calculated for aspect-based sentiment analysis results for social media and news data.

While most of the techniques in our survey visualize sentiment extracted from text data, some of them (28%) use **Text** as one of visual representations. Techniques such as Ink Blots by Abbasi and Chen [1], Jigsaw by Görg et al. [156], and Semantize by Wecker et al. [467] visualize the content of a text document with words, phrases, or regions highlighted based on the sentiment analysis results. In addition to such an approach, LDA-based sentiment visualization by Chen et al. [77] presents an overview of sentiment-bearing word pairs for a selected topic in hotel reviews by positioning them in a grid layout. The horizontal position encodes the polarity in this case. Nokia Internet Pulse by Kaye et al. [210] organizes key terms from tweets into vertical stacks / columns following the temporal order horizontally and applies color coding to convey the average polarity associated with such terms. Typographic Set Graph by Brath and Banissi [47] demonstrates how multiple font attributes such as font family, style, and weight can be used simultaneously with color coding to represent membership of words in 10 sets corresponding to emotion and polarity categories.

The final representation in this group is **Glyph/Icon**, which is supported by 35% of techniques. Multiple techniques in our survey use glyphs in combination or as part of other visual representations. For instance, Wanner et al. [457] propose a visualization of RSS feeds comprising series of curved bar glyphs that encode temporal, topical (whether the RSS news is related to the Democratic or Republican party), and polarity information. A later work by Wanner et al. [459], Topic Tracker, is a system for temporal visual analysis of Twitter streams that combines topic monitoring and gradual sentiment polarity detection. The authors use basic color-coded triangle glyphs representing the timestamp and polarity of individual tweets. The glyphs are positioned in a dense fashion, and the final result resembles pixel-based representations. FluxFlow by Zhao et al. [491] is a VA system for investigating anomalous patterns of information spreading on social media, namely, retweeting threads on Twitter. The system uses lexical matching of emotion-bearing words as one of the features for estimating the anomaly score and visualizes these values as part of *thread glyphs*. The recent VisForum system by Fu et al. [135] uses glyphs to represent the activity of individual users and groups in online forums, and it encodes the average polarity of group discussions with colors using the purple hue for negative and brown for positive. Several other techniques use gauge-like glyphs, for instance, VIBES by Wensel et al. [470], the work by Kherwa et al. [220], or Social Sentiment Sensor by Zhao et al. [495]. Even smiley faces are used in visualizations as glyphs (e.g., in Affect Color Bar by Liu et al. [262], Tour-pedia by Cresci et al. [90], and NewsTone by Harris [165]), which in a way is taken to the next level by the actual faces of virtual avatars by Gobron et al. [150].

3.2.6 Discussion and Analysis

After several iterations of adding new techniques and refining the categorization, we have been able to summarize the state of the art in sentiment visualization based on the statistics for our data. In addition, we have investigated the relations between the categories in general.

Correlation between Categories We have conducted a correlation analysis for categories assigned to sentiment visualization techniques. Technique entries were treated as observations, and categories were treated as dimensions/variables (see the online browser discussed below for the complete categorization results). Linear correlation analysis was then used to measure the association between pairs of categories. The resulting matrix in Figure 3.8 contains Pearson's r coefficient values which reveal certain patterns and interesting cases of positive (green) and negative (red) correlation between categories. The interpretation of the coefficient values seems to differ in the literature: Cohen [82] defines the range 0.30–0.50 (absolute values) as moderate correlation and 0.51–1.00 as strong correlation; Taylor [415] mentions the corresponding ranges 0.36–0.67 and 0.68–1.00 used in earlier works; and Evans [125] defines the ranges 0.40–0.59 for moderate correlation, 0.60–0.79 for strong correlation, and 0.80–1.00 for very strong correlation. Based on this, we have focused on cases with an absolute value of 0.40 or higher.

The interesting cases with *negative correlation* mostly include categories from the same groups, implying "competition" or a kind of paradigmatic relation between them. For example, the correlation of −0.47 between reviews and social media could be explained by a general shift from data sets of well-defined product reviews to the data extracted from social media (including posts associated with some brands)—see the discussion of such temporal trends below.

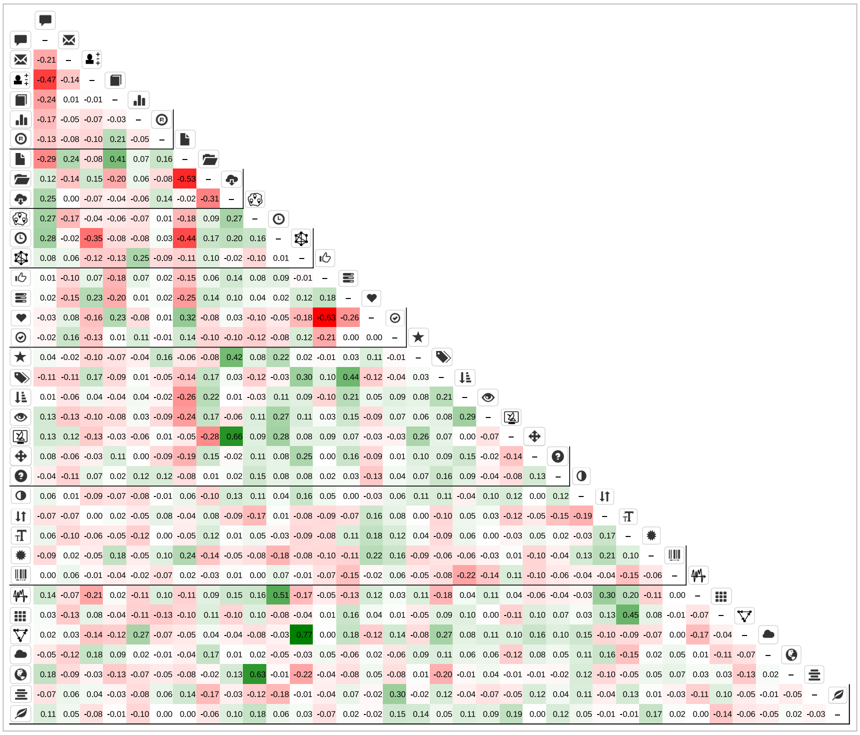


Figure 3.8: The matrix of correlation values for categories (see Table 3.5) calculated for our techniques data set.

The correlation of −0.53 between the document and corpora categories is interesting, given the discussion in Section 3.2.3. While there are many techniques using both data source types, this value could be explained by techniques focusing exclusively on individual documents or corpora. The document category is also negatively correlated to the temporal data property with a value of −0.44: while there are some techniques which treat position in text as the temporal dimension, the majority of the techniques which support time series apparently use other data sources. Finally, the strongest negative correlation in our survey is −0.63 between the analytic tasks of polarity and emotion/affect analysis, which is explained by a more specialized nature of the latter task.

The cases of *positive correlation* tend to include categories from different groups in a kind of syntagmatic relation. For instance, the correlation of 0.41 between the literature domain and the document data source type can be explained easily: the corresponding techniques focus on individual novels/poems. The positive correlation of the streams data source with such visualization tasks as region of interest and monitoring with the respective values of 0.42 and 0.66 was also expected. The analytic task of opinion mining / aspect-based sentiment analysis involves the support for various kinds of classification/clustering/categorization, hence the positive correlation of 0.44. Temporal data is often visualized using representations such as line plots or rivers, which explains the correlation of 0.51. The strongest positive correlation of 0.77 in our survey exists between network data and node-link representations, which is not surprising at all. We could also expect the correlation between geospatial data and map representations to be higher than 0.63—but we have noted in Section 3.2.5 that the latter category also includes abstract maps. Finally, the correlation of 0.45 between pixel/area/matrix representations and the visual variable of size is explained by such representations as bar charts and pie/donut charts.

Popular Approaches Table 3.5 presents the statistics for the collected data based on the final categorization. It supports our expectations of the most common aspects of existing sentiment visualization techniques. An average technique is used for visualization of temporal data (stored as corpora) from social media based on polarity analysis / subjectivity detection. The popularity of the more specific aspect-based sentiment analysis / opinion mining is explained by the existing interest for topic analysis and visualization. Based on the statistics given in Table 3.5, we can also identify a standard set of four visualization tasks relevant to sentiment, which reflect the visual information seeking mantra [374] and visual analytics mantra [212] to some extent: if possible, classify/cluster the data into groups first, then provide an overview of these results, compare interesting items, and explore them in detail. More specific visualization aspects related to variable and representation are discussed below. Also, we should note that the absolute majority of sentiment visualization techniques rely only on 2D representations (even though we have not discussed it explicitly).

Temporal Trends In addition to the overall statistics, it is also interesting to analyze temporal trends with regard to the occurrence of individual categories in our collected data set. Figure 3.9 comprises sparkline-style plots based on the category counts normalized by the total technique counts for each year (for example, 6 out of 29 techniques from 2016 support streaming data). It allows us to detect global trends and compare trends within each group of categories. For instance, it confirms our previous statements about popularity of social media and decreasing role of reviews as data sources. The popular approaches discussed above demonstrate stable support throughout the years, but it is also interesting to trace the usage of underrepresented categories over time, as discussed below.

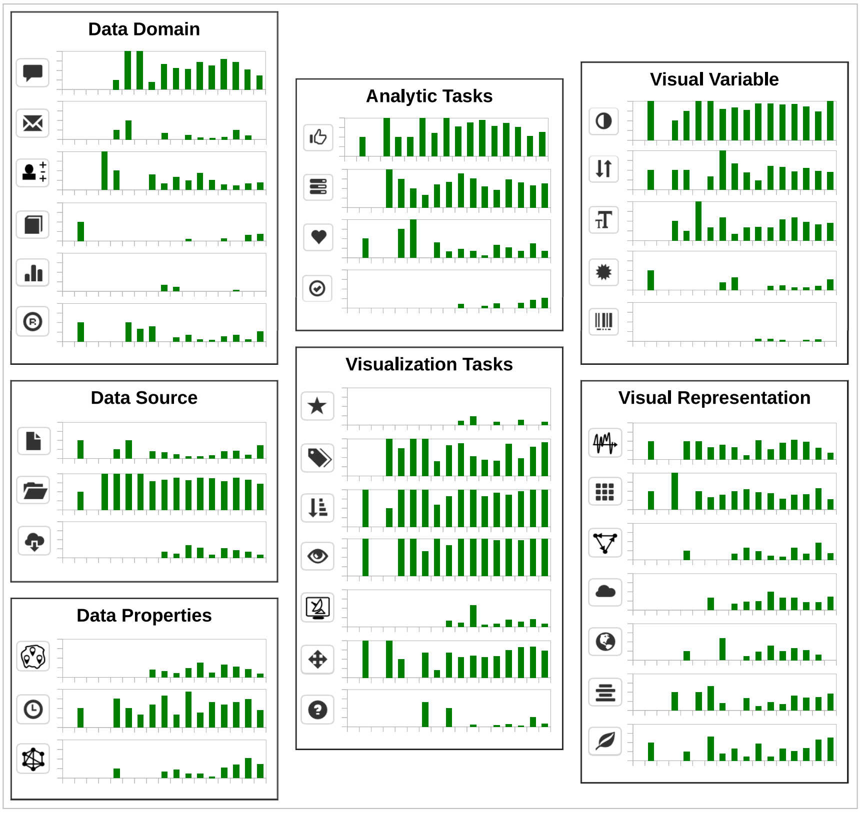


Figure 3.9: Temporal sparkline-style plots representing relative popularity of categories in our data over time (see Table 3.5 for the legend). The values are relative to the total technique counts for the corresponding years (cf. Figure 3.6).

Underrepresented Categories Emotion/affect analysis is supported by a relative minority of techniques, and only a few techniques address the issues of stance analysis, as discussed in Section 3.3. According to the temporal trends (see Figure 3.9), the interest has fluctuated over the years for the former task and started to emerge only recently for the latter. These analytical tasks present multiple opportunities for future research with applications in several domains (social media, literature, etc.). As for visualization tasks, uncertainty tackling is currently underrepresented. This also presents future research opportunities, since uncertainty is an inherent aspect of many popular ML models (e.g., SVM or CRF) as well as data sets characterized as "Big Data". The increasing interest for such data will for sure also affect the number of techniques supporting streaming data sources as well as the related visualization task of monitoring. While such data sets are mostly associated with online sources, we have been surprised by the statistics for several other domains. For instance, very few techniques (mostly from the past) address the literature domain as opposed to the overall growth of interest for digital humanities; also note the low number of techniques that focus on a single document. We also had to exclude the domain of patents from our categorization compared to the more general design space of text categorization (cf. Section 3.1.1) since no detected techniques support it. This can be explained by the generally formal and objective style of language used in such texts—on the other hand, several techniques already support the related domain of scientific articles.

Visual Representations The statistics on the visual variable state that color is the most common visual channel to convey sentiment/emotion, which was expected by us even before collecting the data. The rather large number of techniques using position/orientation and size/area to encode sentiment can be explained by the usage of line plots and stream graphs for temporal sentiment data, as well as pie charts and bar charts for simple visual summaries which are often used by techniques originating in non-visualization disciplines. This leads us to the issue of categorizing the visual representations used for sentiment into simple and complex, which was initially one of our intentions. The existing work investigating complex representations [427, 455, 487] refrains from providing exact definitions of such, though. Shamim et al. [371] have already raised the issue of evaluating sentiment visualization techniques including the perception aspects—this problem presents interesting opportunities for future research.

Interactive Exploration with a Survey Browser We have developed an interactive survey browser similar to our TextVis Browser (cf. Section 3.1.2) to accompany the work discussed in this section and used it extensively ourselves while working on the survey. Figure 3.10 demonstrates its user interface that is focused on individual technique thumbnails and the interaction panel comprising category filters and a search field. In general, it follows the design decisions used by several existing browsers [29, 235, 236, 364, 421]. SentimentVis Browser is implemented as a client-side web application using HTML, JavaScript, and D3 [94]. After loading the page, the user is presented with a list of visualization technique thumbnails organized in a grid. The entries are ordered by publication year first and then by the prime author's surname. Clicking on a thumbnail opens a dialog box with details such as a complete bibliographical reference, a URL link to the source publication webpage (if available), a BiBTeX file link, and a complete list of categories assigned to the corresponding technique. The categories are also presented in form of filters in the main interaction panel on the left. Additionally, the panel includes a text search field, a time range slider, and a histogram showing the temporal distribution of techniques before and after filtering. The users can also access category statistics via the "About" dialog (cf.

Figure 3.10: The web-based user interface of our visual survey called *SentimentVis Browser* ("Sentiment Visualization Browser") that consists of the technique thumbnails grid and the interaction panel. Details for a selected entry are shown by clicking on a thumbnail image in the main view. The survey contains **167** categorized visualization techniques as of February 6, 2019.

Table 3.6: Authorship count distribution for sentiment visualization techniques

#techniques	1	2	3	4	5	6	7	9	13
#authors	439	48	24	6	8	3	1	1	1

Note: the current data set includes **167** techniques and **531** authors in total.

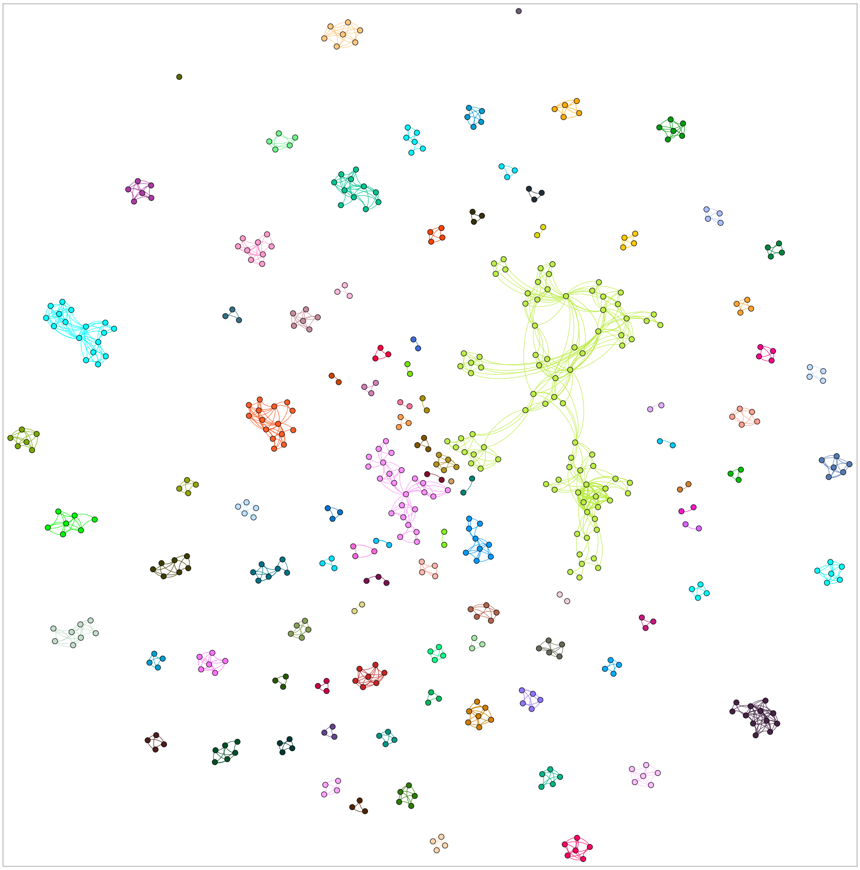


Figure 3.11: Co-authorship network for the sentiment visualization survey entries (as of February 6, 2019) visualized in Gephi with force-directed layout algorithms. Note the large connected component in the center (in green) containing **86** author nodes.

Table 3.5) and a summary table with an overview of the complete categorization. We encourage the users to submit additional entries to SentimentVis Browser via a form available from the top panel—the process is not entirely automated, though, since we intend to continue careful curation of the survey.

Authorship and Co-Authorship Statistics Similar to the analyses for the text visualization techniques and publications data set discussed in Section 3.1.3, we have extracted authorship and co-authorship metadata for the collected sentiment visualization techniques. The top three authors with regard to number of techniques are Daniel A. Keim (13 entries), Christian Rohrdantz (9 entries), and Huamin Qu (7 entries), similar to the statistics for more general text visualization field; they are followed by several authors with 6 and 5 techniques. The comparison of Table 3.6 with Table 3.3 demonstrates rather similar tendencies with regard to the core group of contributors to the field. Switching from authorship to co-authorship metadata analyses, the co-authorship network for sentiment visualization techniques[6] includes 531 nodes and 1,189 edges, and the largest betweenness centrality values in this network are associated with Christopher Collins, Daniel A. Keim, Jian Zhao, Huamin Qu, and Nan Cao. Similar to the network for text visualization in Figure 3.4, the authors who contributed to the sentiment visualization field are mainly grouped in small connected components with a notable exception of a large component including 86 collaborators, as seen in Figure 3.11.

The insights about the current structure of the research community working on sentiment visualization conclude the discussion of this problem, and we can now focus on an even more specific problem of stance visualization.

3.3 Stance Visualization

As part of our survey of the sentiment visualization field discussed in the previous section, we can establish the task of **Stance Analysis** to represent analyses that encompass not only sentiment/affect, but also other categories of subjectivity/evaluation expressed in text, for instance, AGREEMENT, CONTRAST, CERTAINTY, or JUDGMENT. Currently, very few visualization techniques support this task (10% of the complete sentiment visualization techniques set, or even 7% if excluding our own contributions), and we discuss them in detail in the rest of this section.

Small [394] discusses sentiment analysis and visualization of citations in scientific literature with Maps of Science. His approach results in classification of multiple categories beyond the standard polarity-related ones. For instance, UNCERTAINTY and DIFFERENTIATION/CONTRAST can be considered as categories of stance. The results of analysis are used to calculate the layout of a node-link diagram which resembles a map.

Attitude Radial Plot by Almutairi [10] visualizes the contents of a text document as a 2D/3D map using the dimensions of AFFECT, JUDGMENT, and APPRECIATION as well as the position of text fragments in the document.

[6] `http://gmap.cs.arizona.edu/map/9144/` (last accessed in February 2019)

Table 3.7: The existing stance visualization approaches in the design space of sentiment visualization techniques

Technique	Online Social Media	Communication	Reviews / (Medical) Reports	Literature/Poems	Scientific Articles / Papers	Editorial Media	Document	Corpora	Streams	Geospatial	Time Series	Networks	Polarity Analysis / Subjectivity Detection	Opinion Mining / Aspect-based Sentiment Analysis	Emotion/Affect Analysis	Stance Analysis	Region of Interest	Clustering/Classification/ Categorization	Comparison	Overview	Monitoring	Navigation/Exploration	Uncertainty Tackling	Color	Position/Orientation	Size	Shape	Texture/Pattern	Line Plot / River	Pixel/Area/Matrix	Node-Link	Clouds/Galaxies	Maps	Text	Glyph/Icon
Maps of Science (2011) [394]					●			●				●	●	●		●		●	●	●					●						●				
Attitude Radial Plot (2013) [10]						●	●						●			●			●	●				●	●	●	●			●			●		
Lingoscope (2014) [104]	●							●			●		●	●	●	●			●	●		●		●	●	●			●	●				●	
@AM (2014) [307]							●						●		●	●			●	●				●	●	●			●	●				●	
Oil Spill Vis (2014) [425]	●							●			●				●	●			●	●		●		●		●			●						
Proposals and Arguments Vis (2016) [31]	●	●					●					●		●		●			●	●				●							●			●	
SemEval Dashboard (2016) [296]	●							●					●	●		●			●	●		●		●		●				●				●	
ConToVi (2016) [115]		●						●			●			●		●		●	●	●	●	●		●	●	●			●						●
NEREx (2017) [116]		●						●			●		●	●		●		●	●	●	●	●		●			●				●			●	●
Feature Alignment Vis (2017) [200]		●						●			●	●				●		●	●	●		●		●	●						●			●	●
Scalable Stance Vis (2018) [70]	●							●					●	●		●		●	●	●				●	●	●	●			●					●

Note: Supported categories are marked by ●.

Lingoscope by Diakopoulos et al. [104] focuses on an interesting problem of analyzing how opinions and arguments are framed in online blog discourse on the topic of climate change. The tool makes use of key terms of VIRTUE and VICE from GeneralInquirer [146] in order to assess moral APPROVAL and DISAPPROVAL expressed in the texts. The results associated with specific topic terms and search queries are presented using bar charts and area charts, and a text view with highlighted terms is available for close reading.

Neviarouskaya et al. [307] visualize the results of document analysis with their @AM model, including the JUDGMENT and APPRECIATION aspects which go beyond typical polarity and emotion categories.

Torkildson et al. [425] complement Ekman's six emotions with two additional categories representing SUPPORT and ACCUSATION for their visualization of social media posts on the Gulf Oil Spill crisis.

The technique by Bembenik and Andruszkiewicz [31] takes an unstructured text document as input, extracts proposals and arguments alongside their polarity, and visualizes this data using a node-link diagram.

Mohammad et al. [296] introduce a dashboard visualization of the SemEval-2016 stance-annotated data set which provides an overview of the data with regard to class distributions by using several bar charts, a tree map, and confusion matrices. In their work, stance is treated as AGREEMENT/DISAGREEMENT with a certain topic besides the expressed sentiment.

El-Assady et al. [115] provide an animation-based visualization of conversation transcripts, e.g., political debates transcripts, with their system ConToVi. The system supports several stance categories related to argumentation as well as sentiment, CERTAINTY, and POLITENESS. It allows the users to monitor the stance of individual speakers with regard to specific topics. ConToVi is followed up by NEREx [116], an approach which focuses mostly on the networks of extracted named entities and represents them with node-link diagrams. NEREx supports polarity categories as well as POLITENESS, which are represented with node glyphs combining icons and color-coded backgrounds. This work is complemented with an approach by Jentner et al. [200] for interactive alignment of text features for text transcripts. Here, politeness as well as several other features/categories related to stance such as AGREEMENT, BARGAINING, and CONSENSUS are represented with color-coded nodes.

Finally, Chamberlain et al. [70] propose a scalable, lightweight visualization of sentiment and stance that can be embedded into web pages alongside tables or lists with detailed textual descriptions. Their approach uses the stance categories of SUPPORT, NEUTRAL, and OBJECT alongside three sentiment polarity categories and represents the corresponding numerical values with bar charts aligned in a specific order. The use case scenario for this technique is summarization of opinions provided in the public comments for e-government data.

We have summarized the existing stance visualization approaches in Table 3.7 with regard to the categorization introduced in the previous section. These approaches (including the very recent ones) address multiple areas of our design space, for instance, with regard to using multiple types of visual channels & representations or supporting the standard visualization tasks of overview and comparison. At the same time, there clearly are gaps in the existing research on stance visualization with regard to supporting various data domains, data types & properties, and visualization tasks. Furthermore, the challenges related to representing multiple stance categories at the same time (for instance, the output of a multi-label classifier) are not adequately addressed by most of these approaches, which provides opportunities for further contributions in stance visualization.

3.4 Summary

In this chapter, we have discussed the current state of the art in the field of text visualization and then proceeded by narrowing the scope of our inquiry to the more specialized problems of sentiment and stance visualization. We have introduced a categorization of text visualization techniques motivated by earlier surveys, implemented an online survey browser, and collected & categorized more than 400 techniques over the years to achieve the overview of this field. Since our original publication on this topic [236], our categorization and the corresponding data were mentioned by several survey and meta-analysis papers related to text visualization, including the book by Cao and Cui [58], the survey of surveys (SoS) by Alharbi and Laramee [9], and the recent meta-analysis by Liu et al. [264]. Several other surveys and meta-analyses related to text visualization have also emerged, including the surveys on text visualization and visual text analysis for digital humanities by Jänicke et al. [197,198], the taxonomy of target-problem space for visual text analytics by Tofiloski et al. [419], the survey of topic- and time-oriented visual text analytics techniques by Dou and Liu [105], the survey of visual approaches for analyzing scientific literature and patents by Federico et al. [127], and the surveys of social media visual analytics approaches by Wu et al. [478] and Chen et al. [76]. The categorizations used by these authors, including the descriptions of specific domain-related tasks and methods, might be used to refine our categorization as part of the future work. The great interest for integration with ML and AI methods, which emerged in the visualization community in the past couple of years, could also inspire extensions for our categorization; for this, in addition to the aforementioned surveys by Chen et al. [76] and Liu et al. [264], several recent publications could be beneficial, including the works on predictive visual analytics by Lu et al. [269], integration of ML and VA by Endert et al. [120], visual interaction with DR by Sacha et al. [351], interactive analysis of word embeddings by Heimerl and Gleicher [171], and UI design for interactive ML by Dudley and Kristensson [112].

Our categorization of text visualization techniques served as a foundation for the design space of sentiment and stance visualization techniques, which was discussed in detail in this chapter. While the overall number of sentiment visualization techniques found by us in peer-reviewed literature is much smaller than the respective number for text visualization in general, the statistics about the publication dates, outlets, and authors provide evidence for the interest for this problem in the visualization community as well as in other disciplines. The analysis of the existing techniques with regard to our categorization has provided us with insights about the state of the art in this field, including the profile of a typical technique and the gaps in support for various categories. This contribution opens up opportunities for future work involving novel sentiment visualization techniques.

Finally, we have identified the existing approaches supporting stance visualization and positioned them in our design space in this chapter. In comparison to 430 and 167 techniques included in more general text and sentiment visualization data sets, respectively, the current list of techniques supporting stance visualization in our survey includes 16 entries (excluding our own work, only 11). While some of these techniques follow the approach for stance analysis and classification dominant in CL/NLP (that is, focusing on a small set of mutually exclusive categories such as AGREEMENT/DISAGREEMENT or PRO/CONTRA with regard to a specific topic of interest), there is definitely room for further research on stance visualization with different (and more extensive) sets of categories, support for further data types, visualization tasks, and representations. Our own work on stance visualization addressing some of these challenges will be introduced in the remainder of this dissertation, and the contributions will be positioned in the context of the categorization similar to Table 3.7 in the final chapter.

Chapter 4

Visualization of Sentiment and Stance Markers with uVSAT

Contents

4.1 Background 78
4.2 Related Work 78
4.2.1 Time-Dependent Text Visualization 79
4.2.2 Sentiment Visualization 79
4.2.3 Visualization for Linguistic Research 80
4.2.4 Visual Analytics for Sentiment Research 80
4.3 Architecture 81
4.3.1 System Architecture and Workflow 81
4.3.2 Data Model 81
4.4 Requirements Analysis 83
4.4.1 Analysis Challenges 83
4.4.2 Analytical Tasks 84
4.5 Visualization Approach 85
4.5.1 Timeline View 85
4.5.2 Document Views 91
4.5.3 Aggregation Charts 93
4.5.4 Marker and Document Export 94
4.6 Case Study: Linguistics Research 96
4.6.1 Timeline Data Analysis 96
4.6.2 Identifying the Document of Interest 98
4.6.3 Identifying the Markers of ANGER 99
4.6.4 Final Document Analysis 99
4.6.5 Case Study Summary 101
4.7 Expert Reviews and Discussion 102
4.7.1 Domain Expert Reviews 102
4.7.2 Performance and Scalability 104
4.7.3 Lessons Learned 104
4.8 Summary 105

Examples provided in the previous chapters demonstrate how text data available in digital form can be used for computational and visual analyses, in particular, various forms of sentiment analysis and visualization. This also presents an opportunity for researchers that are interested in stance analysis.

To reiterate, *stance* is a relatively broad concept in linguistics [121] related to (inter-)subjectivity expressed in text or human conversation, e.g., attitudes, feelings, perspectives, or judgments. Research on stance includes both theoretical efforts (related to the definition and the knowledge about the nature of this phenomenon) and practical efforts (related to collecting evidence and explaining the means of taking stance), and it can lead to various text analytics applications. The practical tasks require processing large quantities of textual data that is infeasible for manual investigation, e.g., providing a temporal overview of stance usage in social media, retrieving the corresponding text data relevant to stance phenomena, or analyzing the occurrences of stance expressions. Therefore, stance researchers are interested in automated ways of text processing that can be offered by researchers from the field of computational linguistics (CL) or natural language processing (NLP).

Our research project StaViCTA addressed this challenge and aimed to produce a refined theory of stance, efficient interactive visualization and computational techniques for its analysis, and solutions for specific applications. Due to the early stage of research in stance analysis, the project itself followed an iterative progress plan. Therefore, we initially focused on the categories usually detected by the means of sentiment analysis. We considered categories of emotion as underlying aspects of linguistic stance in order to support the construction of the stance model in general; besides emotions, our approach supported stance categories CERTAINTY and UNCERTAINTY with the same computational method.

In this chapter [243][1], we focus on the exploration of social media documents (in English) and the collection of a training data set which was used later in the project to develop appropriate machine learning (ML) approaches. The composed training data consists of text chunks, called *utterances*, that are associated with specific expressions of stance (see Figure 4.1). These utterances can be used for both NLP purposes and manual linguistic investigation; we denote them by *stance markers*. This collection of relevant stance markers is the basis for a refined theory and sophisticated NLP models for stance analysis in general.

Here, we present our tool called uVSAT that can help stance researchers to identify candidate documents that may contain stance expressions, analyze the document texts, and export the new stance markers (as introduced in our previous poster abstract [237]). uVSAT supports the research task of how we

[1]This chapter is based on the following publication: Kostiantyn Kucher, Teri Schamp-Bjerede, Andreas Kerren, Carita Paradis, and Magnus Sahlgren. Visual analysis of online social media to open up the investigation of stance phenomena. Information Visualization, 15(2):93–116, April 2016. SAGE Publications. doi:10.1177/1473871615575079 © 2015 The Authors.

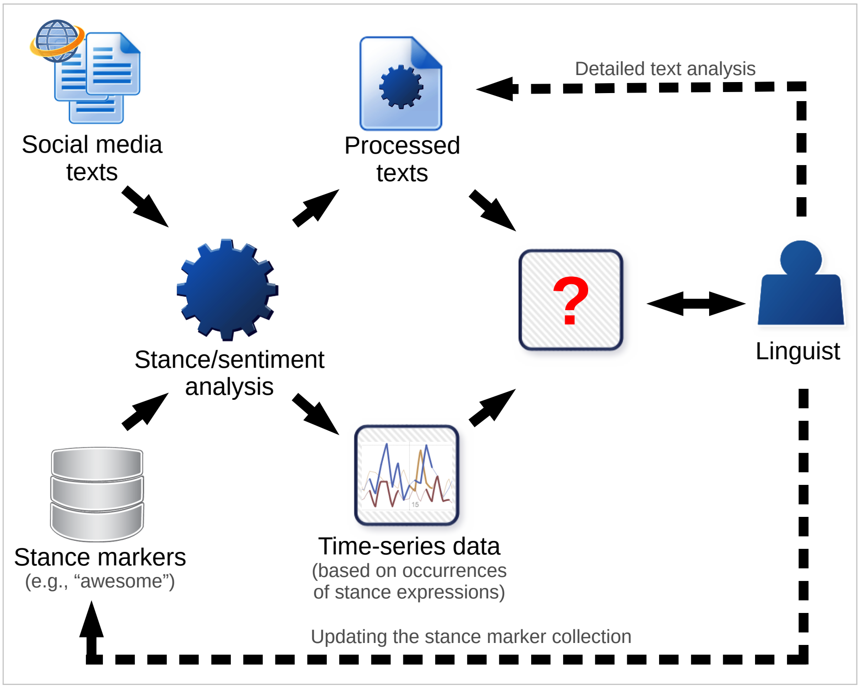


Figure 4.1: The diagram gives an overview of the underlying research problems from the user perspective. To succeed with the analysis of stance, linguists require means to analyze and interact with the output of NLP algorithms as well as means of further manual investigation. These means are still missing in the analysis loop and are indicated by the red question mark. The dashed edges denote the user operations that depend on the results of interactive visual analysis. *Reprinted from [243] © 2015 The Authors.*

can study the use and patterns of stance meanings and stance expressions in human communication over time in order to investigate what stance markers and stance markings are used when, why, how, where, and in what type of dialogic sequences related to the contexts where they occur. Our effort described in this chapter is meant to complement the existing techniques for stance analysis based on manual close reading and traditional linguistic tools by introducing a VA approach to this problem, while not providing a completely automatic stance analysis yet.

The remainder of this chapter is organized as follows: the next section provides a short background discussion of sentiment and stance analysis from the perspective of linguistics and NLP. Section 4.2 covers the related work in text visualization, including work dedicated to sentiment analysis visualization. After that, we explain the system architecture and data model in Section 4.3 as well as user tasks supported by uVSAT in Section 4.4. Then, we describe in detail our visualization and interaction approaches for this tool in Section 4.5. The subsequent Section 4.6 discusses a case study from the linguistics domain based on exploration of data with regard to ANGER sentiment. Section 4.7 provides the results of a domain expert review and our reflections about the tool. Finally, we summarize the contributions and future work in the last section.

4.1 Background

While theoretical background and computational methods for both sentiment and stance analyses were discussed in detail in Chapter 2, we need to make several clarifications about the approach taken for uVSAT. As opposed to some of more complex approaches based on ML, we opt for a simplistic approach to sentiment classification for the purposes of the visualization tool in order to preserve transparency and simplicity. As previously noted, we have chosen to address stance through subcategories. More specifically in uVSAT, these are based on Ekman's "Big Six" emotions [114], employing the NLP solution of simple lexical matching over lists of attitude terms (which we call stance markers, as already mentioned in the previous section). In addition to the categories of emotion, our approach also supports stance categories CERTAINTY and UNCERTAINTY [357]. The main goal at this stage of the project was to facilitate experiments to further improve our understanding of stance in general and our analysis techniques in particular. While our method of sentiment analysis is used for uVSAT simple, such a lexical-based approach is still widely used by visualization and visual analytics solutions [458, 492], especially the ones aiming for high performance when processing large amounts of input data [213]. There are also several examples of combining both lexical-based and machine learning-based approaches for sentiment analysis that report similar [309] or even surprisingly good [156] results when using the lexical approach.

4.2 Related Work

Our tool uVSAT was designed to visualize and interact with large text data sources as well as the results of automatic text processing which include time series. There have recently been multiple works dedicated to text visualization and analytics of social media. While a detailed discussion of existing work in text visualization and, more specifically, sentiment visualization is provided in Chapter 3, we will discuss several groups of works most relevant to uVSAT from various aspects in this section.

4.2.1 Time-Dependent Text Visualization

A good number of such works address temporal aspects to visualize events, topic competition/evolution, or other time-dependent data. While some of them introduce novel metaphors for visual encoding, multiple techniques combine well-known representations such as line plots, river metaphors, or animated force-directed graphs. Havre et al. [167] introduce ThemeRiver, the original technique for temporal data visualization based on a river metaphor that is designed to depict topic evolution in document collections. Dou et al. [107] combine trees, text tags, and rivers in their HierarchicalTopics system to visualize the temporal evolution of topics in corpora. Others, such as Xu et al. [484], combine line plots, stacked charts, and word clouds to depict topic competition in social media document collections. To support real-time monitoring of streaming Twitter data relying on automatic text classification, Bosch et al. [41] use a timeline, word clouds, glyphs, and maps in the ScatterBlogs2 system. For the work described in this chapter, we decided to choose simple visual representations (line plots, text tags, and bubble charts) for the data available to us.

4.2.2 Sentiment Visualization

While specific problems (and the corresponding analysis techniques) such as topic modeling and event detection have been very popular in text visualization, the interest for sentiment analysis and visualization is also arising in the VA community. Liu et al. [261] as well as Oelke et al. [312] describe visualizations for opinion mining of reviews. Wanner et al. [457], Cui et al. [93], and Rohrdantz et al. [345] present approaches for visual sentiment analysis that support temporal data. Görg et al. [156] describe fluid integration of sentiment analysis as well as other computational text analyses with interactive visualizations in their system Jigsaw. Online social media data is used for visual sentiment analysis by Wanner et al. [459], Zhang et al. [489], and Hao et al. [164]. SentiView, introduced by Wang et al. [447], not only facilitates temporal sentiment analysis, but also augments it with relation analysis based on graph representation—this is relevant to our long-term research goals involving intersubjectivity and stance analysis. The work of Zhao et al. [492] describes PEARL, a visual analytics system for multidimensional personal emotion/sentiment visualization of Twitter posts over time and uses an approach similar to ours (based on lexical matching of *emotional words* pertaining to eight emotion categories and three additional emotion dimensions)—however, our work focuses on the analysis and visualization of data related to multiple posters/authors and sources, and we are interested in categories beyond emotions/sentiment. In general, most of the discussed works involve sentiment analysis as a *means* rather than the *object* of research. Our approach, in contrast to theirs, focuses on the analysis of sentiment to bootstrap the research on visual stance analysis. This leads us to the involvement of experts in linguistics as users and the discussion of existing visualization approaches related to the domain of linguistics.

4.2.3 Visualization for Linguistic Research

InfoVis and VA techniques have been used to facilitate tasks such as the analysis of corpora (e.g. Compus by Fekete and Dufournaud [128], CorpusSeparator by Correll et al. [88], Text Variation Explorer by Siirtola et al. [377] and those techniques proposed by Regan and Becker [336]), the analysis of relations/re-use (e.g., ShakerVis by Geng et al. [147] and techniques proposed by Jänicke et al. [199]), and lexical analysis (for instance, the study described by Rohrdantz et al. [346]). An additional category of tasks that is worthy of mention is related to semantics: while numerous text visualization techniques use topic modeling, experts in computational linguistics use visualization to facilitate their research on this subject. For instance, Kabán and Girolami [207] visualize their own model of dynamically evolving text collections. Another task related to stance analysis is discourse analysis. Existing work on visualization of discourse includes the graph-based approach by Brandes and Corman [44], Conceptual Recurrence Plots by Angus et al. [16], as well as several works that focus on discourse in online social media: Lingoscope by Diakopoulos et al. [104] or ConVis by Hoque and Carenini [180].

4.2.4 Visual Analytics for Sentiment Research

Finally, the work that is most relevant to our approach described in this chapter is dedicated to sentiment visualization which facilitates the research on sentiment for linguists. Gregory et al. [157] conduct visual sentiment analysis of document collection with regard to *affect bearing* words. Their approach involves eight affect categories (POSITIVE, NEGATIVE, VIRTUE, VICE, PLEASURE, PAIN, POWER COOPERATIVE, POWER CONFLICT) and uses IN-SPIRE for visualization purposes. The work of Makki et al. [278] focuses on sentiment lexicon refinement from reviews data set which involves user input via interactive visualization. Their sentiment analysis is based on a standard POSITIVE-NEGATIVE dichotomy. The two major differences between these works and our proposed approach in uVSAT are the involvement of online social media text data (which is dynamic with regard to analysis sessions and available for temporal analysis) and the choice of sentiment categories (which is a base for the further analysis of stance).

To the best of our knowledge, the problem of stance analysis and visualization had not been explicitly addressed by work in visual analytics or information visualization before the contribution discussed here. Therefore, we wanted to raise the awareness of the InfoVis and VA communities with the work presented in the current chapter by building on the discussed work in text visualization for sentiment analysis as well as existing work on visual text analytics for linguists.

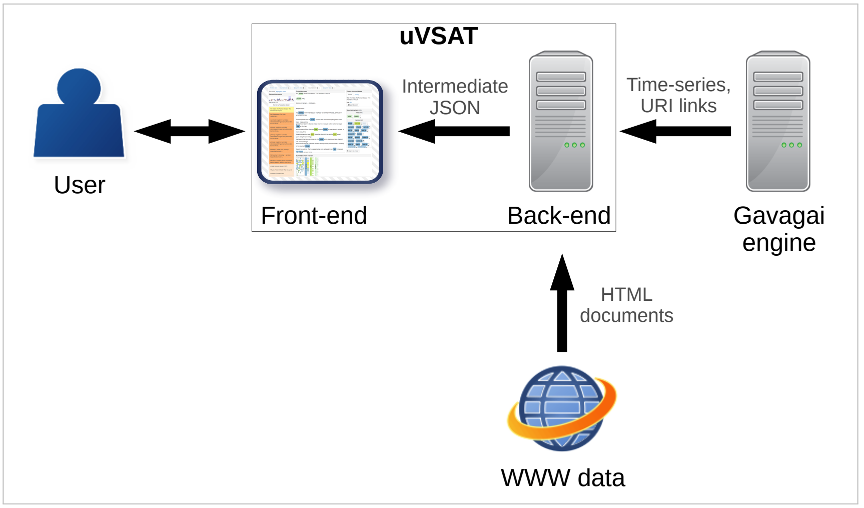


Figure 4.2: The architecture of uVSAT comprises front-end and back-end tiers that communicate with external servers. *Reprinted from [243] © 2015 The Authors.*

4.3 Architecture

In this section, we provide an overview of our software implementation architecture and explain how the sentiment and stance analysis methods are applied to the data available from our partners and remote sources.

4.3.1 System Architecture and Workflow

Figure 4.2 displays the overall architecture of uVSAT that is implemented as a web application. The back end consists of a (visualization) server application implemented in Java with Spark Web Framework [398] that communicates with the Gavagai computing server, fetches the HTML content from URI links, processes the text data and communicates the results in JSON [206] format to the client(s). The front end is implemented in HTML/JavaScript with Bootstrap [38], jQuery [205], D3 [94], and Rickshaw [338] libraries, and it only requires a modern web browser to be used. While the major and cost-intensive computational analyses are processed by the Gavagai and visualization servers, several minor analyses (which do not require intense computations for large amounts of data) are implemented on the client side.

4.3.2 Data Model

uVSAT has been designed to use time series data from external providers through a RESTful API [131], as well as to fetch and process corresponding HTML data from respective web servers. Currently, we use time series data only from our collaboration partners at Gavagai (though we plan to support other data sources in the future). Gavagai analyzes text data from multiple sources, but for the purposes of the system presented in this chapter, they use the data fetched from various blogs and forums.

As mentioned in the background section, we focus on the simplest possible type of stance analysis, i.e., counting the occurrences of sentiment and stance terms/keywords in documents that mention specific target terms. This simple approach allows our partners to support the analysis of large amounts of text data, up to 15 million documents per day. Here, a *target* can be anything of interest: a person, a brand, a company, a location, an event, or even something abstract like a concept or an idea—as long as it can be defined by a set of keywords (also denoted by *target terms* in the context of our tool). Our present set of targets T includes the following[2]:

$$T = \{\mathsf{diet}, \mathsf{weapons}, \mathsf{Hobbit}, \mathsf{Coca\text{-}Cola}, \mathsf{Pepsi}\}$$

To detect documents associated with stance, we consider specific markers relevant to sentiment and (un)certainty from several available sources (WordNet-Affect [404], GeneralInquirer [146], and CompassDeRose [86]), while refining those marker lists is one of the purposes of uVSAT (since the sources above do not differentiate stance from sentiment, etc.). Our choice of analyzed stance categories (also denoted by *observers* in the context of our tool[3]) includes the Big Six emotions (see Section 2.1) as well as two other categories:

$$O = \{\textsc{anger}, \textsc{joy/happiness}, \textsc{fear}, \textsc{sadness}, \textsc{disgust}, \textsc{surprise}, \\ \textsc{certainty}, \textsc{uncertainty}\}$$

As an example, weapons is a monitored target which is defined by a list of 3,771 keywords, harvested from the Wikipedia lists of weapons[4]. Whenever one of these keywords is mentioned in open online media, the entire utterance containing the keyword is analyzed for occurrences of stance markers. Here, *utterance* is simply defined as a sequence of text defined by delimiter symbols, for instance, the text fragment

> *"I am so sick of people who sell such rifles and so sick of people who buy this distasteful weapon."*

[2]Here and below in this chapter, sans serif font is used to indicate targets of interest.

[3]The term "observer" originates from the Gavagai API, and its meaning is related to the corresponding software design pattern with regard to "observing" a signal in the processed data stream. The term was introduced in uVSAT early during the development process and remained in usage to avoid additional confusion from both developers and users.

[4]`http://en.wikipedia.org/wiki/Lists_of_weapons` (last accessed in February 2019)

contains 2 occurrences of the stance marker "sick of" and 1 occurrence of "distasteful", generating a *polarization value* of 3 for the target weapons for observer DISGUST.

To summarize the description of n targets, m observers, and their possible combinations, we can describe the hierarchical structure of the data as $\{(T_i, \{O_{i1}, \ldots, O_{ij}\}) \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ for targets $T_i \in T$ and the corresponding observers $O_{ik} \in O$, for instance, (Hobbit, {DISGUST, ANGER, ...}).

The occurrence counts are aggregated for each target-observer combination (T_i, O_{ik})—e.g. Hobbit/DISGUST or Hobbit/ANGER[5]—over a specific timeframe which is presently set to one hour. Thus, all occurrence counts for a specific stance category within this timeframe $[t_1; t_2]$ are summed, resulting in an hourly value v for each combination. These values are then retrieved and visualized by uVSAT as time series. Because of this aggregation step (which is necessary to reduce the complexity and computational demands), the time series data describes the general tendencies with regard to stance, but it does not directly provide any details about the distribution of specific markers. Therefore, further exploration of the original text documents is required from the users.

The Gavagai API also provides URIs to the documents used to calculate the polarization values (taking (T_i, O_{ik}, t_1, t_2) as arguments and returning sets of URIs), although the corresponding HTML content has to be downloaded and processed on our side. Unfortunately, the total amount of available data makes it infeasible for the VA tool to prefetch everything.
Therefore, we limit ourselves to queries for specified sets of target-observer combinations across interactively selected time intervals (although we plan to support streaming data in the future).

4.4 Requirements Analysis

After the introduction of the fundamentals and research gaps of visual stance analytics including a short discussion of the origin and structure of available data sets, we are able to take a closer look at the actual analysis challenges and most important tasks that uVSAT should address. They are based on extensive discussions with our collaboration partners in linguistics and CL.

4.4.1 Analysis Challenges

We have designed uVSAT to facilitate users with answering the following questions:

Q1 How do the calculated values for targets/observers change over time? What are the overall temporal trends?

[5]Note that we equivalently use the notations (T_i, O_{ik}) and T_i/O_{ik}.

Q2 How to identify "interesting regions" in multiple time series which span over long intervals of time? How to reduce the visual complexity with regard to noisy data?

Q3 What are the original documents associated with the values for targets/observers? How to identify the most interesting documents with regard to stance analysis?

Q4 How are markers distributed in a particular document?

Q5 How are specific markers distributed in the retrieved sets of documents? How to identify the documents with a large number of markers or the documents which contain a lot of unique marker types?

Q6 How to handle a long analysis session involving multiple time intervals and document sets? How to recover a previously discarded document set? How to navigate quickly to a previously analyzed document set?

Q7 Are there any relationships between analyzed document sets?

Q8 How to use particular marker, document, or document set analysis results for further investigation?

4.4.2 Analytical Tasks

These questions and problems can be mapped to the following categories of high-level (analytical) tasks:

T1 *Time series analysis:* compare the values for various targets and observers (Q1, Q2), explore trends (Q1, Q2), identify interesting regions for further investigation (Q2).

T2 *Document sets navigation:* query for the documents associated with selected observers / time intervals (Q3), keep track of related queries (Q7), and navigate the queries history (Q6).

T3 *Document sets analysis:* explore the retrieved document sets (Q3) and reveal the general trends by using data aggregation (Q5).

T4 *Document navigation:* query for specific documents either explicitly (Q6) or while navigating enclosing document sets (Q3) and aggregated data (Q5).

T5 *Document analysis:* explore the text content and stance markers distribution in a selected document (Q4), export the static content for manual investigation (Q8).

T6 *Stance marker collection:* export the selected utterances (or parts of them) as new markers (Q8).

In the following section, we discuss our visualization approach in detail, justify the design decisions, and refer back to the above listed research questions and tasks.

4.5 Visualization Approach

The graphical user interface (GUI) of our tool[6] offers a tab-oriented design with two types of tabs (see Figure 4.3 and Figure 4.4): a single timeline view tab that is used to work with an arbitrary number of timeline plots, and multiple document view tabs that are opened by the user when fetching the document URIs for selected time intervals. As the timeline view is the entry point of all visual analyses supported by our approach, we start our discussion with this view.

4.5.1 Timeline View

The timeline view tab (see Figure 4.3) provides the users with the interfaces for exploring time series data for selected targets/observers and specified time intervals. Note that fetching the input data to be analyzed—i.e., the initial selection of specific targets, observers, and time ranges—from the Gavagai server is done via a simple dialog box as explained in our case study (see Section 4.6). In this section, we concentrate on overall design aspects including visual representation and interaction possibilities.

4.5.1.1 Color Coding Considerations

Before we address the particular representations, we have to explain the color coding scheme used for the timeline view as well as document views. As mentioned in Section 4.3.2, the analyses supported by our tool involve combinations of targets T_i and specific observers O_{ik}. So, the resulting hierarchical data structure for one specific target might be $(\text{diet}, \{\text{ANGER}, \text{JOY}, \dots\})$, for instance. The time series data fetched from our partners is organized this way with the focus on target-observer combinations, and our initial choice of the color coding was based on the decision to provide a unique color for each combination. However, this approach had two issues: first, the sheer number of combinations (45 entries in our present set of target/observer combinations) made it difficult to use a color scheme that would facilitate the users' perception of the data. And second, that color scheme was not related to the scheme for document views (described below), so the users could easily lose the mental map when switching between the view tabs.

[6] A demo video for uVSAT is available at `https://vimeo.com/128357373` (last accessed in February 2019).

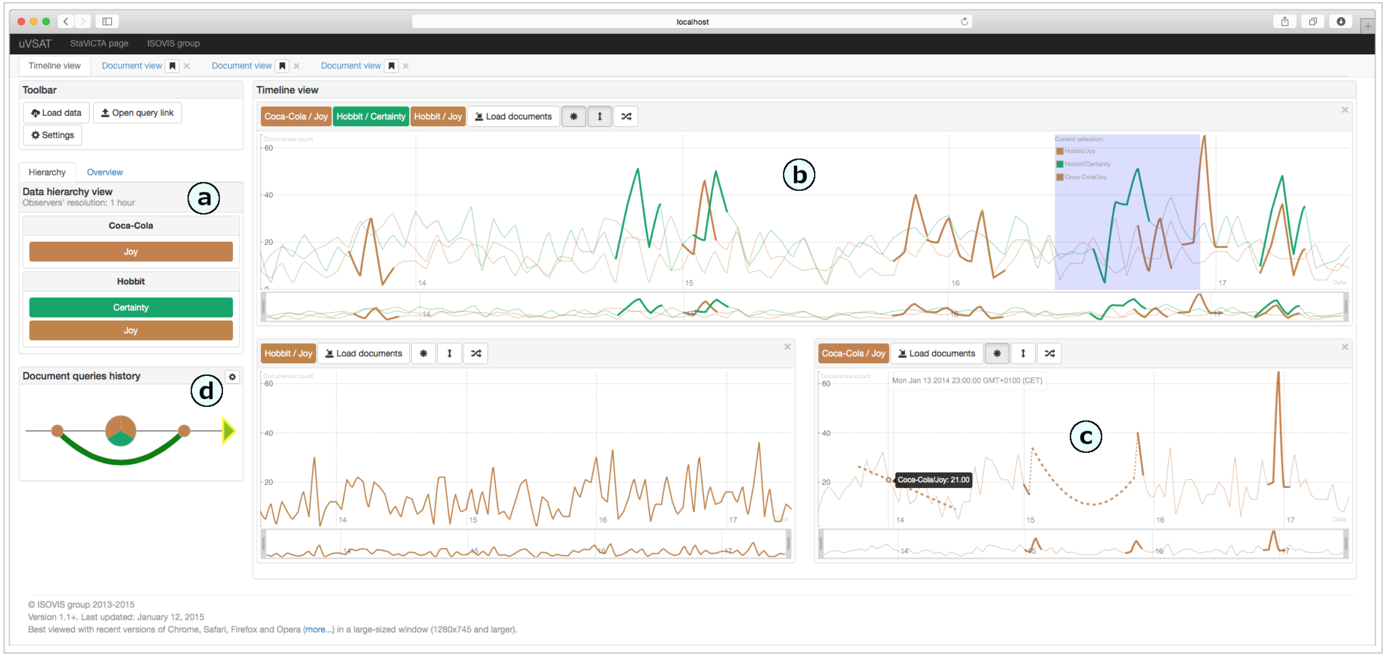


Figure 4.3: The screenshot of our tool shows the *timeline view*. Users start by loading time series data associated with the two targets Hobbit and Coca-Cola and two observers CERTAINTY and JOY. Then they can explore the data and select time intervals for specific target-observer combinations (see the blue-shaded area in (b)) in order to further analyze the corresponding documents. The screenshot demonstrates (a) the data hierarchy view, (b) timeline plots, (c) trend lines, and (d) the history diagram. *Reprinted from [243] © 2015 The Authors.*

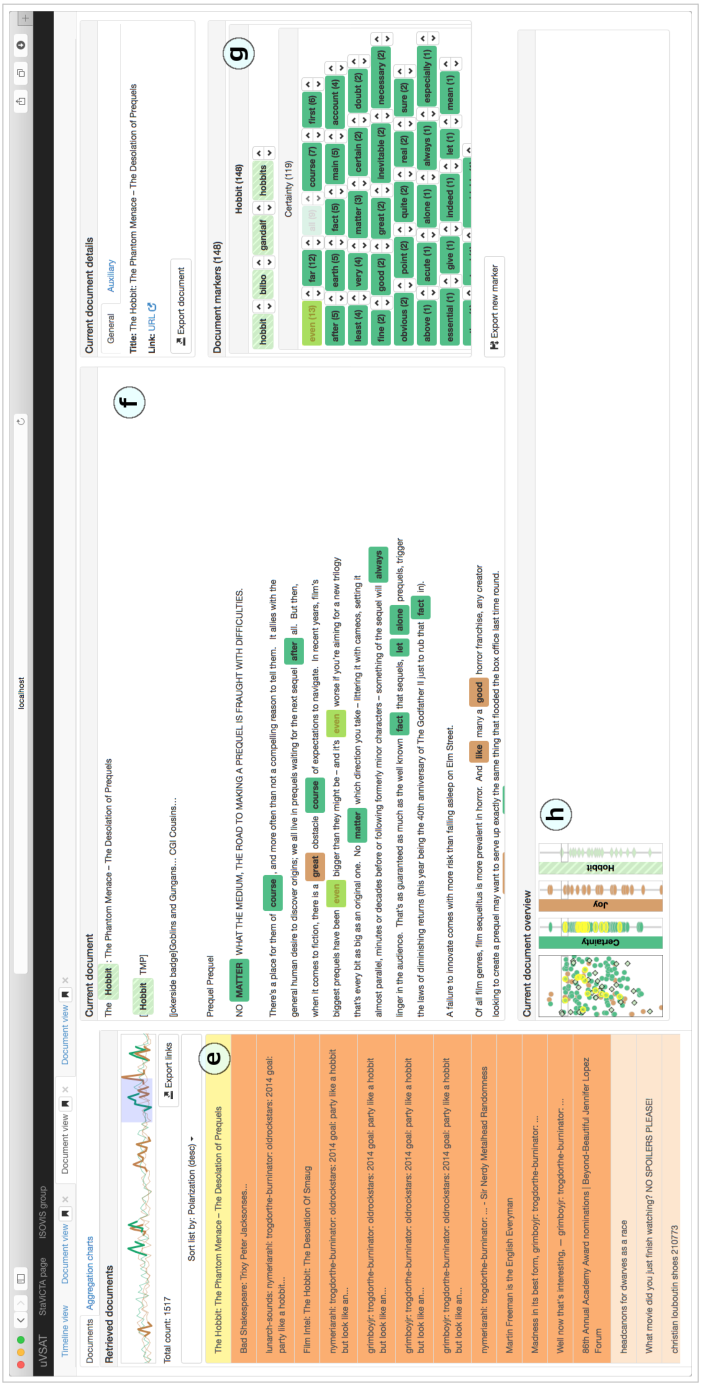


Figure 4.4: The screenshot displays our *document view* for a selected time interval (cf. Figure 4.3) and the target-observer combinations Coca-Cola/JOY, Hobbit/JOY and Hobbit/CERTAINTY. Here, users can explore the resulting set of documents (on the left) and are able to analyze the contents of specific documents (middle) with regard to the occurrence of stance markers summarized on the right hand side. The screenshot demonstrates (e) the document links list, (f) the current document view, (g) the document marker view, and (h) the current document overview. *Reprinted from [243] © 2015 The Authors.*

The analyses employed by document views (see Section 4.5.2) concentrate on the observers, i.e. stance categories, and do not differentiate between observers related to various targets. This had an implication that the color coding for document views was initially based on ColorBrewer [85], and it contained separate colors for observers and targets.

Afterwards, we have changed the color coding used for the timeline view in accordance to the TreeColors approach [416]. To generate the colors, we have *inverted* our hierarchy to the form $\{(O_j, \{T_{j1}, \ldots, T_{ji}\}) \mid 1 \leq j \leq m, 1 \leq i \leq n\}$, for instance, $(\text{JOY}, \{\text{diet}, \text{Hobbit}, \ldots\})$, and then used the TreeColors package. The resulting color coding aims to assign different observers distinct color hues; however, it is not perfect, since there are still too many of those. The colors assigned to target-observer combinations pertaining to the same observer have rather similar hues. This, on the one hand, makes it simple to spot such similar combinations. On the other hand, though, it makes it difficult to discern such plots—this is partially alleviated by interaction techniques such as details on hover and filtering. Overall, the main benefit of this approach is that it allows of using the same color hues for observers across the timeline and document views, which helps to preserve the users' mental map.

4.5.1.2 Data Hierarchy View

After the input data has been loaded, the users are provided with the data hierarchy view displayed in Figure 4.3(a) that shows the hierarchical structure of the available target-observer combinations. Users can also open a tab with sparkline-like [432] "overview plots" (see Figure 4.10) for *all* fetched time series which are similar to regular timeline plots with highlighted *regions of interest (ROI)* (see below). These overview plots support a simple way to compare the time series and to find more general patterns in the data (research questions Q1 and Q2). As soon as interesting target-observer combinations are found, the user may want to investigate this data in detail and drag-and-drop the entries from the data hierarchy view onto the main part of the tab. Then, uVSAT displays the timeline plots for the chosen combinations. For instance in Figure 4.3, a user has selected three views where several target-observer combinations are visualized.

4.5.1.3 Timeline Plots

uVSAT uses a standard line plot representation for time series data (see Figure 4.3(b)) and supports usual interaction techniques for such plots (research question Q1). We have chosen this visual representation as our domain experts are already familiar with it. In addition, line plots can be easily extended with additional graphical features. Details on hover, plot overview, and scroll & zoom are provided by default by the Rickshaw component. Users are also able to filter the plots with regard to visible target-observer combinations by switching on and off the corresponding labels. Our tool supports multiple plots displayed on the same canvas (users can drag-and-drop additional items from the data hierarchy view) or separately (users can drag the plot containers to change the timeline view layout). For the comparison of several plots displayed side by side, users can control the automatic vertical scaling—by default, plots are scaled to fit the containers. This functionality was explicitly wished by our domain experts.

4.5.1.4 ROI Highlighting

To facilitate the search for regions of interest, our tool also supports automatic ROI highlighting (research question Q2). Currently, we use a basic ad-hoc algorithm for marking the regions of interest based on outlier/differential analysis. As a first step of the algorithm, time series points x_i are marked which differ substantially (with regard to threshold parameters θ_1 and θ_2) either from the mean value μ_x (standard deviation σ_x is used for comparison), or from the preceding point (judging by the first derivative x_i'):

$$A = \{x_i : |x_i - \mu_x| > \theta_1 \sigma_x \vee |x_i'| > \theta_2 \max_j(|x_j'|)\}$$

Since the source time series data is in general noisy, A will result in multiple regions of small size (comprising only one or several points). Therefore, in the second step we smooth the results by marking neighboring points as parts of ROI, which will result in contiguous regions:

$$ROI = A \cup \{x_i : (x_{i-1} \in A) \vee (x_{i+1} \in A)\}$$

Regions of interest are highlighted by thick line segments (cf. Figure 4.3(b)). The algorithm parameters θ_1 and θ_2 can be adjusted by the user, which can be used to partially alleviate the problem of noisy data or to increase/reduce the number of highlighted regions to focus on.

4.5.1.5 Trend Analysis

Users have several options of conducting trend analyses over selected time intervals for specified observers (see Figure 4.3(c)). uVSAT supports linear and quadratic time series trend analysis based on polynomial regression (calculated with the Ordinary Least Squares (OLS) method [365]). We implemented two variations: one can choose to either render trends as overlay plots (cf. Figure 4.5(a)) or to substitute selected timeline plot segments with trend lines (cf. Figure 4.5(b)) to reduce the visual complexity of the displayed data (research questions Q1 and Q2). Trend lines are easily distinguishable by the use of dashed line style. Even information about the predicted value change at the current trend rate and a button for removing trend lines are available on hover.

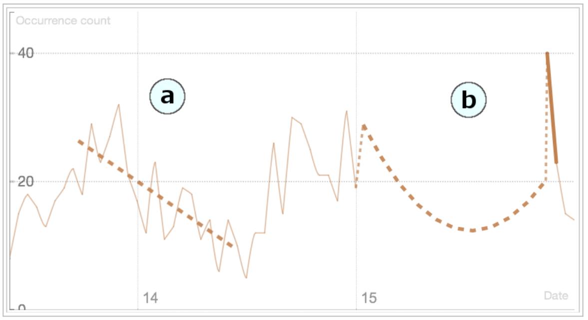


Figure 4.5: Trends in uVSAT can either be displayed either (a) as overlay plots or (b) instead of original plot segments. *Reprinted from [243] © 2015 The Authors.*

4.5.1.6 Document URI Links Queries

As soon as the user is more interested in concrete documents whose frequencies are represented by the different time-plots, he/she can select time intervals for specific sets of observers and load the corresponding URI links to the documents (research question Q3). In this case, a new document view tab is created and a thumbnail of the line plot used for the query is displayed in this new view in order to preserve the mental map. An example of this thumbnail can be seen in Figure 4.4 in the left upper corner.

4.5.1.7 History Diagram

Since the workflow of uVSAT involves multiple document view tabs that also may be closed by a user during the analysis process, the need for overview and control of such user actions arises. Our interactive history diagram (see Figure 4.3(d) and Figure 4.6) provides an overview of the document URI queries sequence, their results, and relations to each other (research questions Q6 and Q7).

In this diagram that supports the so-called analysis provenance [219], nodes represent URI queries and edges represent the detected relations between corresponding query results (this partially resembles the visualization approach described by Cernea et al. [68]). The size of every node is proportional to the number of URI links retrieved for the corresponding query. Nodes are represented by glyphs similar to pie charts (though only qualitative information about relevant observers is used), following the same color coding of observers as the timeline plots. The currently selected node is highlighted in yellow. Since the diagram is used for history navigation, it also contains a dedicated node (depicted by a triangle) that represents the up-to-date interface state. Edges connect only nodes whose query results contain common subsets of URI links. The size of common subsets (i.e., Jaccard similarity of link sets [163]) is mapped to edge opacity, thickness, or both of these attributes (selected as a user setting). The

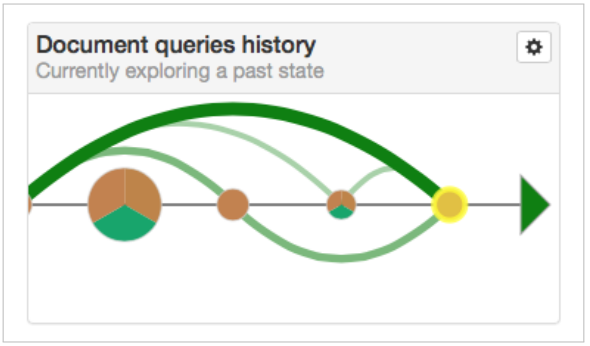


Figure 4.6: The history diagram allows users to keep track of document queries and navigate between interface states. *Reprinted from [243] © 2015 The Authors.*

layout of the history diagram is based on arc diagrams by Wattenberg [464]: nodes are simply aligned along a horizontal axis in the order of corresponding queries, and edges are rendered as curved arcs. We apply a random-order greedy heuristic described by He et al. [168] to decrease the number of edge crossings when allocating edges to the upper/lower part of the drawing.

The interactive history diagram covers the following functionalities: every time a user issues a URI links query that leads to the creation of a new document view tab, the state of this new tab and the timeline view tab are saved, and a corresponding node is added to the history diagram. When the user clicks on a history node, the timeline view tab state is restored, a document view tab with corresponding state is either created or brought into focus (if currently present), and the user actions temporarily stop affecting the history state (e.g. issuing a new query will not add the resulting state to history)—we have chosen such behavior to keep the history sequential. When the user clicks on the triangle, the previously saved up-to-date state is restored. Under circumstances, this can lead to some document view tabs getting closed.

4.5.2 Document Views

A document view tab (see Figure 4.4) basically consists of two areas. The left (smaller) area provides information about all documents fetched based on the selection made with a timeline plot. Thus, it shows the aforementioned line plot thumbnail used for the query as well as a list of links (see Figure 4.4(e)) to HTML documents (blog posts, forum messages, etc.) that were marked as associated with a specific target-observer combination. Users can filter the list by URI domain and sort it by the timestamp value or by polarization value (as reported by the Gavagai server). Polarization values are also used for the color coding of list entries (research question Q3).

By selecting a link from the list, the corresponding document content is fetched, processed at the (visualization) server side, and rendered at the client side. If the content is not available at this time, the corresponding list entry is marked. The document data at this stage is raw HTML which affects the analysis. This is because the source code comments and metadata (such as keywords) often contain text irrelevant to the document content. To direct the user's focus on textual document data, uVSAT renders the HTML content as plain text by using the Jericho library [201]. All data and analysis results related to the single focus document are shown in the second area on the right hand side of the document list. This area integrates four subviews: the current document view, the current document details view (not further discussed here), the document marker view, and the current document overview.

It should be noted that uVSAT also provides an opportunity to copy the query link for a given document view tab and to use it in later analysis sessions by opening a tab with identical contents (research question Q6).

4.5.2.1 Current Document View

Figure 4.4(f) displays the text representation of a document. The stance markers and target terms are highlighted and support brushing in coordination with the other views (research question Q4). The motivation for the color coding for document view tabs was described above: it uses a scheme with 8 colors for stance markers and a separate scheme with 5 colors based on ColorBrewer for target terms since targets share stance markers associated with observers (categories of stance), e.g., the word "commendable" is a marker of JOY for both Hobbit and Coca-Cola. To distinguish target terms from stance markers, the former are marked by a striped background pattern.

4.5.2.2 Document Marker View

Information about stance markers (and their occurrence counts) as well as target terms detected in the current document is summarized in the document marker view (see Figure 4.4(g)). The stance markers for each observer are sorted by their counts to facilitate user investigations (note that target term occurrences do not affect the statistics, since such terms are not directly related to expressions of stance). The users can navigate the document with regard to marker/term occurrences and to filter them (research question Q4).

4.5.2.3 Current Document Overview

To give users an overview of markers/terms distributions in the current document (and an additional means of navigation), uVSAT provides several visual representations displayed in Figure 4.4(h). First of all, a 2D overview is visualized by mapping the current positions of all markers/terms onto a canvas (they are represented by circles and diamonds, respectively). The current viewport is displayed as a rectangle. This overview supports navigation by clicking on a plot item or the canvas. Additionally, a separate 1D overview for each observer and target is visualized by projecting the positions of corresponding markers/terms onto a vertical axis. Such overviews help the users to immediately perceive the distributions over the document length since the 2D overview can become cluttered in case of numerous markers/terms. 1D overviews support document navigation by clicking on plot items. Seeing such distributions is especially interesting for our domain experts, because it is important for a better understanding of stance in discourse (research question Q4), for instance, if a marker for a specific stance category mostly occurs in the context of another marker.

4.5.3 Aggregation Charts

While the techniques discussed above allow the users to analyze a selected document in detail and provide an indication of interesting documents (by polarization values), the document sets retrieved for certain queries may contain thousands of documents, and the users will benefit from a method that helps them to select documents that are interesting for further stance marker investigation (research question Q5). uVSAT addresses this problem with a technique that we call *aggregation charts*: it provides an informative overview and means of navigation for the current document set with regard to detected markers and observers (cf. Figure 4.7 and Figure 4.8).

The visual representation is based on basic bubble charts described by Viégas et al. [444]. Every item in the chart represents a single document which corresponds to the target; the color coding is based on the nominal target values. A single item is visually represented by a glyph consisting of two nested circles. The size of the outer circle is proportional to the total number of corresponding stance markers detected in the document, and the size of the inner circle (filled with a more saturated color) is proportional to the number of unique marker types detected in the document. For instance, a document with 100 occurrences of a marker “good” and 100 occurrences of a marker “bad” has only two unique marker types: “good” and “bad”.

The aggregated data used for these charts can be organized in two ways: by observer and by stance marker. In the former case, a separate chart is visualized for each observer associated with the document set. In the latter case, one individual chart is visualized for each unique marker type (belonging to present observers) that has been detected in at least one document.

Figures 4.7 and 4.8 display examples of aggregation charts visualized for a document set based on 1,517 URIs retrieved for the target-observer combinations Coca-Cola/JOY, Hobbit/JOY, and Hobbit/CERTAINTY. In Figure 4.7, the charts are organized by observer: the left chart contains items pertaining to both Coca-Cola and Hobbit, however, the right one does not contain items for Coca-Cola, since

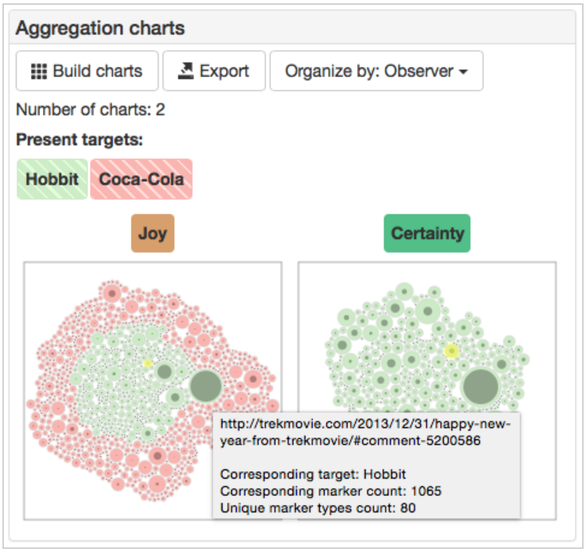


Figure 4.7: Aggregation charts *organized by observer* allow users to explore the distribution of documents with respect to the corresponding observer. *Reprinted from [243] © 2015 The Authors.*

no corresponding target-observer combination was available. The figure also shows the details for a chart item displayed on hover. An example of aggregation charts organized by stance markers is displayed in Figure 4.8. There are multiple charts sorted by the corresponding document numbers in decreasing order, and the user can browse these charts with a specific marker in mind. Details for the first chart (marker: "good") are provided in a tooltip. Here, the currently selected document is highlighted (yellow) in all charts.

Aggregation charts facilitate quick perception of the distribution of observers / stance markers in all documents, identification of documents with a large number of stance markers or unique marker types, navigation to such documents, and analysis of document properties concerning other observers / stance markers (by brushing the corresponding chart item).

4.5.4 Marker and Document Export

One aim of our visualization tool is to identify and collect relevant stance markers from a larger number of analyzed documents (research question Q8). uVSAT supports export of new stance markers from document view tabs by selecting a portion of text in the current document view (depicted in Figure 4.4(g)), assigning it with arbitrary tags, and exporting it to a JSON file. This approach allows us to collect a data set of stance markers not restricted to the categories currently used

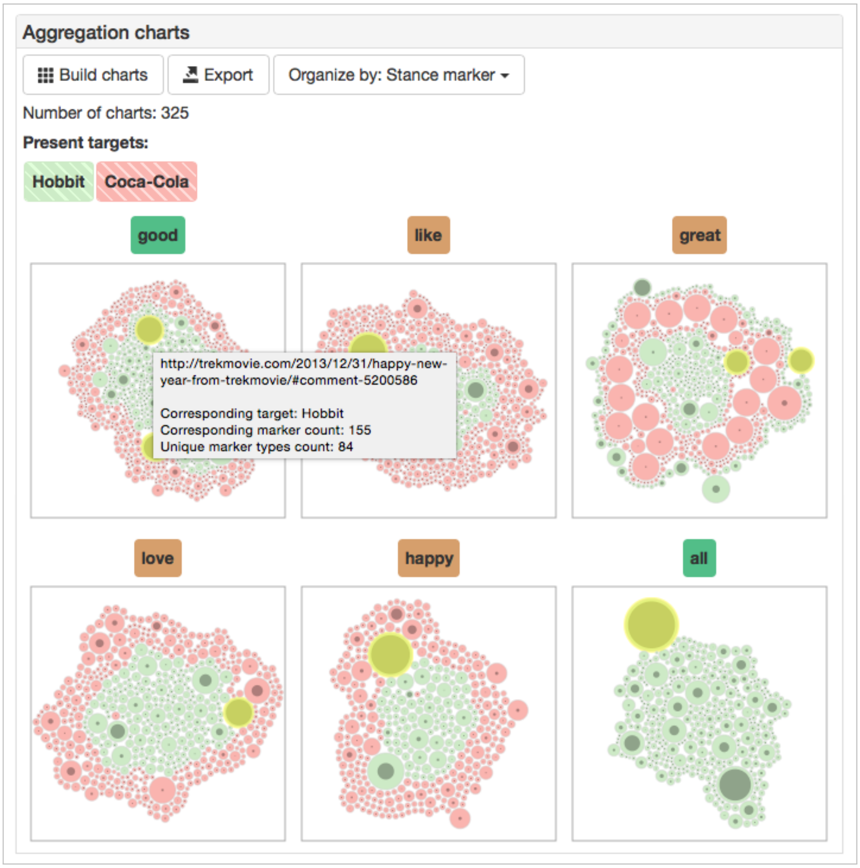


Figure 4.8: Aggregation charts *organized by marker* allow users to reverse the flow of analysis: they can concentrate on document distributions with regard to a specific interesting stance marker. *Reprinted from [243] © 2015 The Authors.*

for observers. Moreover, we are able not only to collect stance markers as short phrases (1-grams [279], 2-grams, or similar), but also to collect larger utterances which provide context for stance analysis.

Our tool also supports export of currently viewed documents and aggregation charts as static HTML pages. In the former case, the document view with highlighted stance markers and target terms, document details, hierarchical markers view, and document overview (essentially, all the data pertaining to the current document in a document view tab) are exported. In the latter case, all aggregation charts that are currently available are exported together with the corresponding document set query (used observers, selected time interval, etc.).

This feature allows users to store static data for further manual investigation or referencing, which can be especially helpful for researchers in linguistics.

4.6 Case Study: Linguistics Research

The case study described here is one in which a linguist has chosen to analyze negative sentiments of stance (focusing on ANGER) in blogs, within a limited one-week timeframe. This example illustrates how researchers in linguistics benefit from our tool when conducting stance analysis. The event chosen was the highly controversial Coca-Cola commercial presented during Super Bowl XLVIII[7] in February 2014 (03 Feb 2014 CET). The aims of the analysis are the following:

A1 analyze the overall usage of stance-related sentiments for the timespan of the scandal,

A2 identify the document with the largest number of markers of ANGER,

A3 identify the most frequently used ANGER markers,

A4 analyze how such markers are used in the previously identified document, and

A5 finalize the choice of the detected document for further linguistic research.

For performing an accurate analysis, data revealing information about the communicative forces, the attitudes to the ideas discussed at different points in time, and possible relationships between those attitudes must be made available to the researcher. By using uVSAT, the linguist is able to analyze these aspects of the social media data which would be impossible for manual stance analysis.

4.6.1 Timeline Data Analysis

First, the researcher uses the *Load data* dialog box and selects all Coca-Cola observers for the time interval 30 Jan 2014 12:00 – 06 Feb 2014 12:00 CET in order to obtain a very broad return of data (see Figure 4.9). The time series calculated for corresponding observers are loaded from Gavagai API.

By viewing the *hierarchy* and *overview tabs* (cf. Figure 4.10), the researcher verifies that all of the chosen observers have been loaded and confirms that there is sufficient data to be analyzed.

The researcher immediately notices the spike of activity on multiple plots around early hours of February 3 CET, which corresponds to the late evening of February 2 EST—the time when the advertisement was aired in USA (aim A1).

[7] `http://buzzfeed.com/ryanhatesthis/coca-colas-multi-lingual-super-bowl-ad-inspired-a-racist-mel#v3khrr` (last accessed in February 2019)

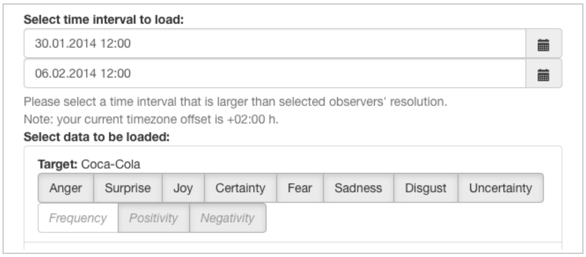


Figure 4.9: The dialog box used to select the time intervals and target-observer combinations to load time series data for. Note that there are additional observer types (FREQUENCY, POSITIVITY, NEGATIVITY) provided by Gavagai by default that are not associated with concrete stance markers (therefore, they are beyond the focus of our research). *Reprinted from [243] © 2015 The Authors.*

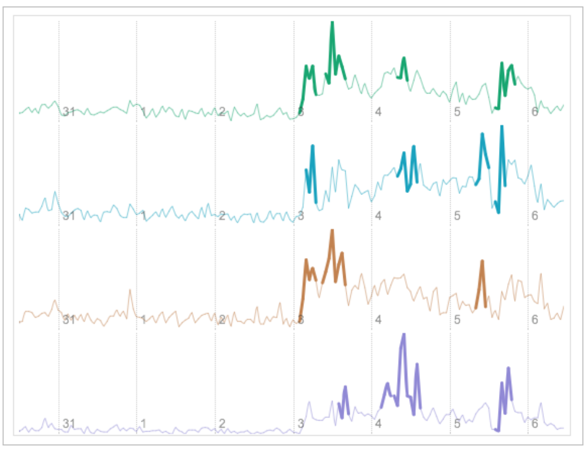


Figure 4.10: Part of the timeline overview: the plots for observers are ordered by mean value in descending order, CERTAINTY being the first. Note the spike around February 3, when the scandal occurred. *Reprinted from [243] © 2015 The Authors.*

Then, the researcher creates timeline plots by dragging-and-dropping the observer items onto the *timeline view*. By using the slider control, the researcher concentrates on the timespan 03 Feb 2014 01:00 – 03 Feb 2014 19:00 CET. To confirm a conjecture that some of the observers have extremely low counts in the current timespan (aim A1), the researcher filters them out. The remaining observers are CERTAINTY, JOY, UNCERTAINTY, and ANGER (see Figure 4.11). To start analyzing the text data, the researcher issues a request for corresponding URIs.

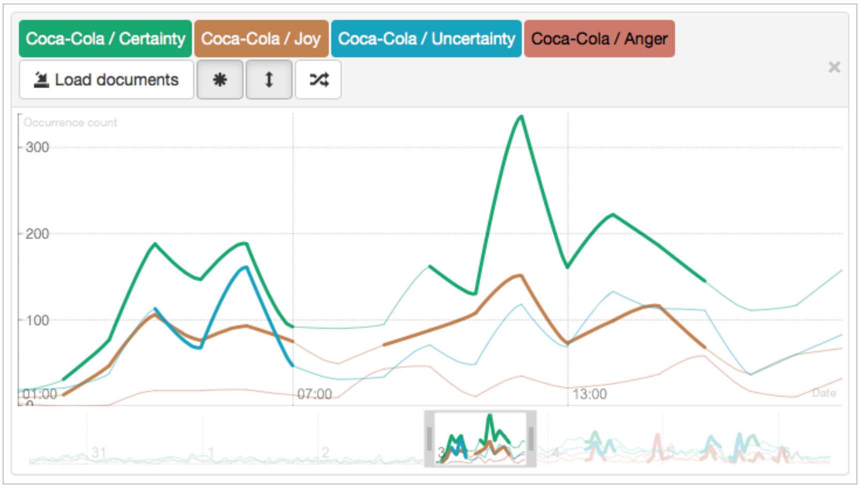


Figure 4.11: Timeline view: four observers for target Coca-Cola that are used for detailed analysis are CERTAINTY, JOY, UNCERTAINTY, and ANGER. *Reprinted from [243] © 2015 The Authors.*

4.6.2 Identifying the Document of Interest

The resulting URI set comprises 3,424 document links. While the researcher could explore this data set manually, it would take a significant amount of time to achieve aim A2. At this point, the researcher decides to build the aggregation charts for the current document set and to investigate the charts organized by observer. For this, the text document data is fetched from the respective web servers and processed by uVSAT.

The aggregation chart for ANGER (see Figure 4.12) comprises 1,948 documents which in total contain 154 unique markers of ANGER. The researcher immediately identifies two candidate documents with the largest number of corresponding markers which are represented by glyphs with the largest diameters (also, with large shaded areas which means large number of unique marker types). By hovering on these glyphs, the researcher finds out that one of them contains 142 occurrences of ANGER markers (39 unique types) and another one contains 193 occurrences (41 unique types). The researcher selects the latter glyph by clicking and loads the corresponding document.

The loaded document of interest (depicted in Figure 4.13) is a blog post[8] with a heated discussion in commentaries. To concentrate on the analysis of ANGER markers, the researcher filters out all markers of other observers. The current document overview plots at the bottom of the screenshot clearly show

[8]`http://americablog.com/2014/02/bigots-pod-coke-super-bowl-ad-singing-national-anthem.html` (last accessed in February 2019)

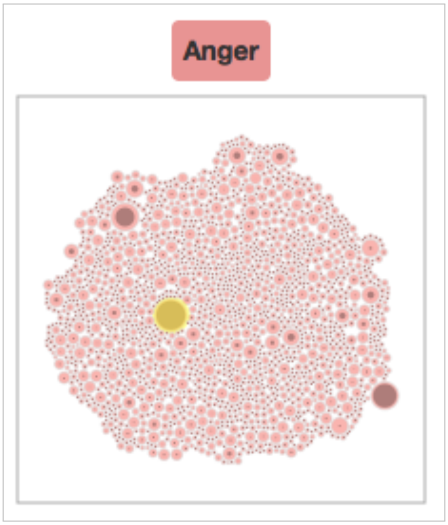


Figure 4.12: The aggregation chart for ANGER provides an opportunity to identify the document with the largest number of corresponding stance marker occurrences. There seem to be two candidate documents which are represented by large glyphs (also with large shaded area). By hovering on these glyphs, the one with larger count of markers (in this case, 193 occurrences) is identified and later used for detailed analysis. *Reprinted from [243] © 2015 The Authors.*

that the markers of ANGER, as well as the target terms of COCA-COLA, are evenly distributed throughout the entire document. To refine the analysis, the researcher needs to concentrate on specific markers.

4.6.3 Identifying the Markers of ANGER

The aggregation charts for the current document set can be organized by stance marker instead of observer. The researcher selects this option and explores the resulting set of 605 aggregation charts (one per each unique stance marker type). Since the charts are ordered by marker occurrences number in descending order, the researcher quickly identifies several most frequent markers of ANGER, thus achieving aim A3 (see Table 4.1).

4.6.4 Final Document Analysis

After identifying the most frequent markers of ANGER using the aggregation charts (here: "hate", "angry", "offended", etc.), the researcher concentrates on the previously selected document and filters out all the other markers. It turns out that some of the identified markers are also among the most frequent markers of ANGER in the document as well (cf. Table 4.2).

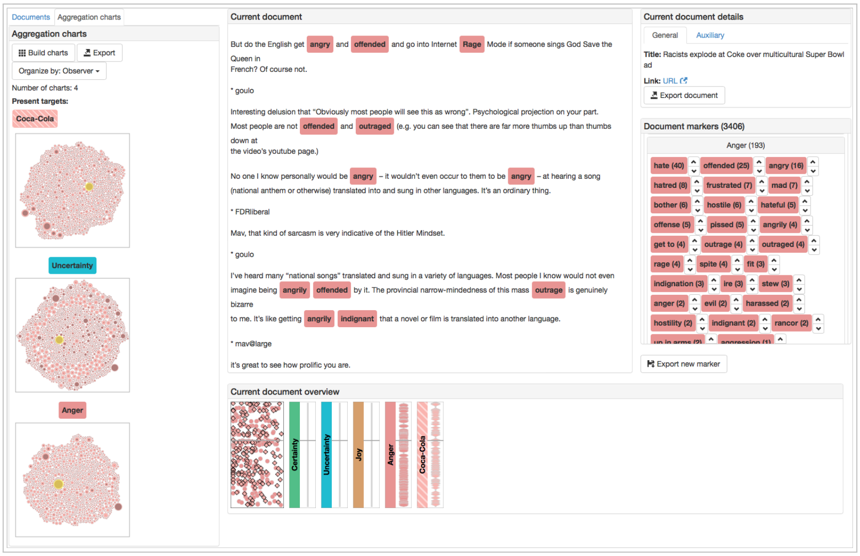


Figure 4.13: Document view for a selected document with majority of stance markers filtered out. Besides the Coca-Cola target terms, only the instances of all markers of ANGER are displayed. *Reprinted from [243] © 2015 The Authors.*

Table 4.1: Stance markers of ANGER in the documents

Marker	Corresponding documents	Unique markers in documents
hate	579	123
angry	347	113
offended	265	92
outrage	232	91
fit	206	107

Note: The table includes the most frequently used stance markers of ANGER in the document set related to the case study. This data has been discovered by investigating the details when hovering over aggregation charts' labels. *Reprinted from [243] © 2015 The Authors.*

The researcher reviews the current document overview once more (see Figure 4.14) and concludes that the identified markers are also distributed throughout this document. As the observer ANGER has the marker "hate" prolifically used, the analyst investigates further, addressing the linguistic characteristics that are employed by users that have posted these. The linguist now proceeds with a close analysis of the document giving critical attention to the markers "hate",

Table 4.2: Stance markers of ANGER in the selected document

Marker	Occurrences in document	Rank in document
hate	40	1
offended	25	2
angry	16	3
outrage	4	8
fit	3	9

Note: The table includes the number of occurrences and ranks of the previously identified stance markers of ANGER in the current document. *Reprinted from [243] © 2015 The Authors.*

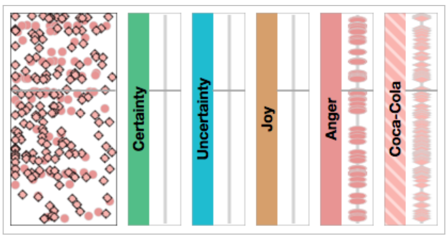


Figure 4.14: The overview for the previously selected document with only five marker types of ANGER displayed. Note that even after filtering the other ANGER markers (cf. Figure 4.13), numerous instances of these five marker types remain and they seem to be distributed throughout the whole document. *Reprinted from [243] © 2015 The Authors.*

"offended", and "angry", thus achieving aim A4. The researcher's conclusion is that the identified document is interesting for further manual linguistic analysis (e.g., with regard to the flow of the conversation) as well as for preparation of an ML training data set. By exporting the document from uVSAT, the linguist achieves aim A5.

4.6.5 Case Study Summary

By using uVSAT, the researcher has been able to achieve her/his analysis aims, i.e. exploring the data related to the case, analyzing the stance-related phenomena of *anger*, and exporting the analyzed text data. By being able to interpret the regions of interest on the timeline view, the researcher was able to limit a great amount of documents to an amount for a more detailed review. The tool's ability to visualize multiple markers simultaneously in the document overview positively guided the investigation. By viewing the aggregation charts, the researcher's decisions were visually supported, and she/he was able to draw the conclusions about stance phenomena in the data set. The potential for employing these different refinement possibilities lets the researcher review statistical plots that are dynamic and updated as new postings are incorporated into the document view. The analysis features provided by the document view complements the manual stance analysis based on close reading. Overall, the patterns constructed by uVSAT create an ample opportunity for the researcher to employ user-based data *en masse*.

On a final note, the linguist began with one specific study area. After using uVSAT, the researcher concluded that the data has also revealed three other possible areas of interest: (1) directionality and frequency of the ANGER markers, i.e., who the poster intends as the recipients and how often they appear and respond; (2) instances of how posters modify their use of ANGER, i.e., intensifiers or attenuators; and (3) if ANGER is negated so as to create a positive meaning. The tool has provided several new potentials for future lines of research that could have gone unnoticed if traditional linguistic investigations were used.

4.7 Expert Reviews and Discussion

In this section, we present the results of two domain expert reviews as well as a discussion of performance issues. Based on these findings, we discuss some lessons learned during the development and testing phase of uVSAT.

4.7.1 Domain Expert Reviews

For the time being, our research partners at Lund University have been the primary users of uVSAT. They are familiar with standard tools for corpus analysis (e.g. AntConc, BYU-BNC, WORDSMITH, or Google Ngram Viewer) as well as manual text analysis. As a kind of project preparation, we introduced basic visualization concepts and techniques to them at the beginning of our collaboration. Their suggestions and feedback during the design and development stage of uVSAT are summarized in the following with regard to general analysis workflow, visualization and interaction techniques, and possible improvements for the tool.

4.7.1.1 General Analysis Workflow

The experts have been very enthusiastic about the opportunity to analyze a large number of online social media documents in detail with regard to stance and sentiment in an interactive way. They have noted that their usual tools of choice in most cases require text preprocessing and employ static or rarely updated corpora, as opposed to our approach:

> *"The uVSAT tool can accommodate the time factor and help the analyst sift through large amounts of data where important chunks could easily*

> *be overlooked. Using the uVSAT tool, which is visually driven to reveal patterns, the researcher can track these and follow how language is being shaped by current digital communications."*

The experts have also appreciated the fact that uVSAT is implemented as a web application which does not require a specific OS or installation/update procedures.

4.7.1.2 Interactive Visualization Approach

The feedback on the design of both timeline and document views has been positive. The experts have approved of the features facilitating the time series analysis, in particular, they have liked that ROI highlighting is turned on by default. The experts have commended the usage of color coding to highlight the regions of interest as well as the markers/terms. They also have approved our decision to convert HTML documents into plain text in order to concentrate on the text content in the document view tabs. The experts have also been very positive about the aggregation charts as a means of overview, pattern detection, and navigation:

> *"Aggregation charts give extremely comprehensive views that are easily understood by this user. These images result in giving the researcher a direct visual confirmation of the number of markers, which then can be scrolled through, chosen and loaded."*

The ability to export stance markers and the content for further manual investigation was also commented on:

> *"This gives the user a pro-active involvement in the ongoing improvement of the tool that is neither confusing nor time-consuming."*

4.7.1.3 Possible Improvements

One of the experts' suggestions during the development was related to the comparison of several timeline plots. We have addressed it by providing an ability to control the layout of the timeline view and to disable the automatic vertical scaling which allows the user to compare the plots situated side by side. The feedback also included some complaints related to the tool performance (see below in the next subsection) as well as a wish for additional functionality related to document set overview (e.g. clustering the documents in aggregation charts by the URL domain). We have also learned that the trend analysis feature is only rarely used since it currently focuses on already available time series data—therefore, we are planning to extend this feature by supporting predictive trend analysis to increase its level of utility.

4.7.1.4 Expert Review Summary

The experts have stated that uVSAT is a useful addition into their arsenal of stance analysis techniques. They are using it to explore and analyze the social media data, and complement it with manual stance analysis as well as by processing the exported data with other software tools, e.g., for concordance analysis. They have also started to collect the ML training data set, thus achieving the general design goals. In general, the domain experts have concluded the following:

> *"For a linguist, uVSAT is a viable tool for working with stance analysis."*

4.7.2 Performance and Scalability

In this subsection, we discuss certain aspects that affect the user experience when trying to apply uVSAT for the analysis of rather large data sets: data transmission delays, data processing delays, and user interface responsiveness.

The original version of uVSAT described in our article [243] stored neither time series data nor document text data on our visualization server. Hence, uVSAT issued requests for time series data, URIs, and HTML content from external servers on demand. This led to delays while retrieving the source data. Additional delays occurred while transmitting the data between the front-end and back-end components and, finally, while processing the data at the server side.

We addressed the networking delay by conducting some types of analyses (such as ROI highlighting or trend computations) on the client side. It seemed, though, that the performance bottleneck was the step of fetching the HTML content from numerous external servers which may have varying connection speed, performance, access frequency limitations, and even availability. Since that version, we have implemented support for caching the external data and some processing results in a local database (more specifically, MongoDB [298]) on our visualization server.

As for the UI responsiveness: D3 and Rickshaw use SVG for rendering which may require significant computational resources (and leads to UI lags). On a 2013 MacBook Pro computer with Intel Core i7 processor (2.3 GHz), sensible UI delays start to occur when re-rendering plots with a total of about 3,000 points. This is partially addressed with a style of workflow involving preliminary analysis of time series overview and focusing on selected time intervals.

4.7.3 Lessons Learned

Our visualization approach in uVSAT involves multiple coordinated views [343] based on standard representations. Its main advantage (as opposed to a more complex integrated view) is the ease of user adoption: the primary users of our tool are researchers in linguistics who do not tolerate abundant details or unintuitive visual representations. The corresponding disadvantage, however, is the necessity of large display area to lay out all the views in sufficient size. We plan to address this issue in the future by developing novel visual representations for stance-related and time-dependent text data, having the domain particularities in mind.

The fact that our source data originates in online social media also has certain consequences: the text documents may be edited or deleted at any time. This presents us with a trade-off between data validity and performance. By fetching online data on user's demand, every document is analyzed in its up-to-date state (or it is marked as unavailable), but it requires computational resources (and it is also related to inevitable networking delays). Otherwise, if the data is cached while the original data is modified, it would actually invalidate the detailed analysis of document contents. The current version of uVSAT seeks the trade-off between these cases by storing the precomputed time series in the database and caching the contents of a number of recently accessed documents. Other possible strategies for addressing these issues would be to involve uncertainty tackling techniques at the visualization stage or to store multiple versions of document data.

4.8 Summary

In this chapter, we have introduced an approach for stance analysis based on sentiment and CERTAINTY/UNCERTAINTY considerations. We have presented our visual stance analysis tool uVSAT that supports interactive exploration of time series data associated with online social media documents, including the text content of such documents. While uVSAT does not provide completely automatic stance analysis, it assists linguists by complementing manual stance analysis of text documents based on close reading with a visual analytics approach that allows the researchers to make use of massive data sets originating from social media.

The contributions of the chapter include the description of a VA tool that contains multiple approaches for analyzing temporal and textual data as well as exporting stance markers in order to prepare a stance-oriented training data set. We have also presented special visualization techniques developed for our tool: the history diagram (for document set query analysis provenance) and the aggregation charts (for document set overview, navigation, and comparison).

We used uVSAT for the purposes of the StaViCTA project, and we provided feedback from the linguistics experts in this chapter. By using uVSAT, our researchers in linguistics were able to collect stance markers and utterances that were later used to define stance categories other than sentiment and CERTAINTY/UNCERTAINTY (e.g. CONCESSION, DISAGREEMENT, etc.). The tool was actively used for collecting documents that formed the training data set for our researchers in NLP as well as for actual stance analysis conducted by the linguists. The process of applying this collected data for training a machine learning classifier is discussed in the next chapter.

At the time of finishing our original article on uVSAT [243], some of the future work plans included support for a local database in order to improve the performance and implementation of a lightweight lexical matching-based NLP engine independent from Gavagai API in order to support additional data sources. The former of these tasks was completed with MongoDB being supported on our visualization server. Implementation of an NLP engine and support for consuming streaming data from Twitter [436] at the back end was also carried out for the latter task, and the tool will eventually switch to this data pipeline instead of Gavagai API. Other future work plans for uVSAT include implementation of additional overview and navigation techniques for document sets, support for uncertainty tackling (with regard to missing time series data as well as unavailable web documents), and user studies to evaluate the effectiveness of single techniques such as the history diagram and the aggregation charts.

Chapter 5

Active Learning and Visual Analytics for Stance Classification with ALVA

Contents

5.1 Background 109
5.2 Related Work 110
5.2.1 Annotation Tools for Textual Data 111
5.2.2 Visual Support for Annotation and Classification 112
5.2.3 Stance Visualization 113
5.3 System Overview and Visualization Tasks 115
5.3.1 System Architecture 115
5.3.2 Annotation Process 116
5.3.3 User Tasks for Visualization 118
5.4 Visualization Methodology 119
5.4.1 CatCombos Representation Design 120
5.4.2 Visualization Interface in ALVA 124
5.4.3 VSM Visualization Interface in ALVA 129
5.5 Case Study: Annotation Data Exploration 130
5.5.1 Initial Findings 130
5.5.2 Combination of 5 Categories 131
5.5.3 Combinations of 4 Categories 132
5.5.4 Combinations of 3 Categories 132
5.5.5 Utterances Table Exploration 133
5.5.6 Summary 134
5.6 Discussion 135
5.6.1 Active Learning 135
5.6.2 Annotation Visualization 136
5.6.3 VSM Visualization 136
5.6.4 Overall System Utility and Generalizability 137
5.7 Summary 139

As we have established in the previous chapters, research on stance [121] is actively ongoing in linguistics and computational linguistics (CL) / natural language processing (NLP) [296, 297, 389–391]. Stance-taking ranges from general agreement/disagreement to fine-grained indications of wishes and clearly expressed emotions. These may be expressed through a wide variety of constructions in human communication such as morphemes ("think*able*", "do*able*"), words ("perhaps", "believe"), or larger units ("I rest my case").

The potential applications for such an analysis technique include social media monitoring, analysis of political debates, literature studies, integration in intelligent user interfaces (e.g., in order to adapt the interface based on the stance expressed by the user), etc. However, the existing efforts to build a fine-grained stance classifier are facing multiple challenges related to the collection of training data as well as the actual design and training of a machine learning (ML) classifier. The existing annotation and classifier training approaches have only limited support for the tasks related to stance classification. Furthermore, they do not provide a sufficient level of support for exploratory visual data analysis of annotated data. The researchers and analysts working with such annotation environments could benefit from data visualization by gaining insights about various aspects of stance phenomena in the collected data. They could apply this knowledge to improve both the theoretical framework and practical applications of stance analysis. Therefore, there is a need for an approach combining computational and visual analyses to facilitate such users.

In this chapter [242][1], we discuss our system, called ALVA (Active Learning & Visual Analytics), which was introduced in our previous poster abstract [238]. ALVA is an integrated visual analytics solution designed as part of a collaboration with researchers in linguistics and CL in our project StaViCTA to support all stages of the annotation and classifier training process (see Figure 5.1). Since some of the stance categories are very sparse in the source data, our solution follows the *active learning* approach [370, 422] to select candidate utterances for further annotation. Our collaborators have used it for annotation and analysis of stance in social media texts in English collected from blogs and forums. In fact, they have compiled an annotated corpus of text on Brexit[2], named *Brexit Blog Corpus (BBC)* [383].

We deliberately constrain the scope of this chapter to supporting the data annotation and classifier training stages with interactive user interfaces and visualizations. The detailed description of the annotation protocol motivated by linguistic research on stance is discussed in a separate article on the BBC

[1]This chapter is based on the following publication: Kostiantyn Kucher, Carita Paradis, Magnus Sahlgren, and Andreas Kerren. Active learning and visual analytics for stance classification with ALVA. ACM Transactions on Interactive Intelligent Systems, 7(3):14:1–14:31, October 2017. doi:10.1145/3132169 © 2017 ACM.

[2]`http://en.wikipedia.org/wiki/Brexit` (last accessed in February 2019)

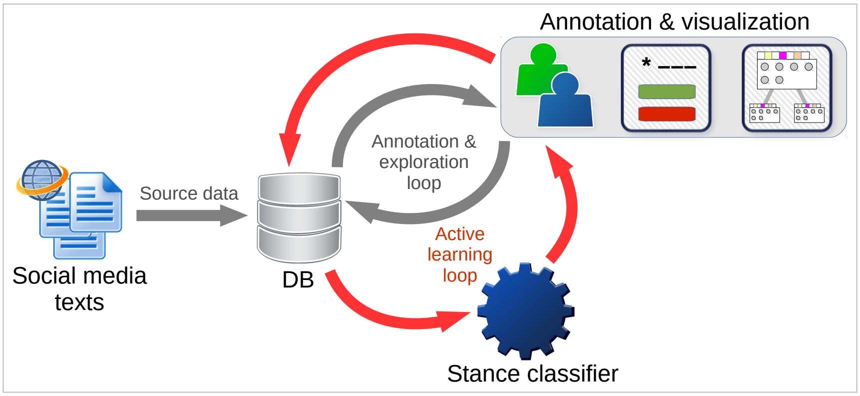


Figure 5.1: The overview of the main aspects of the proposed approach. The overall process involves the annotation and visual exploration loop initiated by the users (in gray), and the active learning loop initiated by the stance classifier (in red). *Reprinted from [242] © 2017 ACM.*

corpus [383]. The description of the stance classifier from a CL/NLP perspective is discussed by Skeppstedt et al. [390].

The rest of this chapter is organized as follows. We start by providing the background information on stance analysis and stance annotation concerns in Section 5.1. In Section 5.2, we discuss related work on annotation tools for textual data, visual support of annotation and classification, and the previous work on stance visualization. An overview of the system and the concrete user tasks for visual analysis based on the available data structure and classification schema are discussed in Section 5.3. Our resulting visualization methodology is described in Section 5.4. We illustrate the features of ALVA with a case study in Section 5.5. Finally, we discuss the insights and notes from our domain experts in Section 5.6 and conclude this chapter with Section 5.7.

5.1 Background

As described in previous chapters, our StaViCTA project concerned stance analysis involving written texts in English originating in social media. The approach described in this chapter aimed to facilitate the project stage dedicated to data annotation and classifier training. The researchers in linguistics defined a set of notional stance categories that could occur in utterances (text chunks) in a non-exclusive manner, i.e., a single utterance could in general be associated with several categories. The source data was presented to the annotators in batches in a specific order, thus forming multiple annotation rounds. Further details about the annotation process are discussed in Section 5.3.2. Table 5.1 presents the list of stance categories used during the annotation process alongside the statistics of their occurrence in the collected data set.

Table 5.1: The stance categories used in our work with ALVA

Category	Occurrences in data
AGREEMENT AND DISAGREEMENT	3.6%
CERTAINTY	6.8%
CONCESSION AND CONTRARINESS	16.3%
HYPOTHETICALS	7.8%
NEED/REQUIREMENT	8.3%
PREDICTION	9.3%
SOURCE OF KNOWLEDGE	11.1%
TACT AND RUDENESS	3.9%
UNCERTAINTY	8.1%
VOLITION	2.4%
IRRELEVANT	4.3%
NEUTRAL	39.9%

Note: IRRELEVANT and NEUTRAL represent invalid utterances (e.g., only numbers or URLs) and lack of any other categories, respectively. They are treated as additional categories for convenience throughout this chapter. Color coding used for categories in ALVA is discussed below in Section 5.4.1. *Reprinted from [242] © 2017 ACM.*

Since our annotation process involves multiple non-exclusive categories and varying sets of annotators assigned to annotation rounds, it also presents certain difficulties for the analysis and visualization of process statistics. Annotation studies in CL usually involve computations of *inter-annotator* agreements to evaluate the concord between several annotators, and sometimes *intra-annotator* agreements for the same annotator over several annotation rounds. Artstein and Poesio [18] provide a comprehensive overview of methods for computing the inter-annotator agreement in several common settings. We apply some methods discussed in their survey and calculate agreement values separately for each stance category. *Observed* agreement is calculated as the proportion of agreeing annotations for two sets of annotations. *Chance-corrected* agreement adjusts the observed value to account for random agreement occurring by chance. We calculate chance-corrected agreement as Cohen's *kappa* [82]. Individual agreement values and summaries for stance categories are represented in our visualization interface described in Section 5.4.2.

5.2 Related Work

In the following, we distinguish between related work on (1) annotation tools for NLP tasks, (2) visualization of annotation and classification processes, and (3) visualization of stance phenomena.

5.2.1 Annotation Tools for Textual Data

The existing annotation tools for textual data can be classified into two classes for our purposes: (1) regular general-purpose tools and (2) tools which support active learning.

5.2.1.1 Regular Annotation Tools

There exist a number of text annotation tools for CL/NLP tasks discussed in the literature. Some of the earlier examples include systems such as WordFreak by Morton and LaCivita [302], MMAX by Müller and Strube [303], and Knowtator by Ogren [313]. These tools are typically designed to annotate data for well-known tasks such as part-of-speech (POS) tagging, dependency parsing, and named entity recognition (NER). While they can provide a rich set of customizations with regard to the annotation schemata, their weaknesses include deployment, management, and usability issues for annotation projects involving multiple annotators with varying level of technical skills.

More recent contributions address most of such issues by providing web-based annotation environments which separate user roles and provide annotators with less cluttered and more intuitive user interfaces. Stenetorp et al. [402] discuss a web-based system, called BRAT, which supports annotation for multiple NLP tasks and uses colored text spans and labeled edges in the annotation interface. Yimam et al. [486] extend the ideas of BRAT in their system, called WebAnno, which provides better support for multiple annotation layers and tag sets configuration. These environments are limited with regard to the exploratory visual analysis of both the annotated data and the annotation process, though. A better level of analytical support for administrators and analysts is provided in the Marky system by Pérez-Pérez et al. [323], which includes a bar charts-based overview of inter-annotator agreement as well as a confusion matrix for a given annotation round. Still, it does not support the visual analysis of the resulting annotated data set.

5.2.1.2 Active Learning Tools

Several existing tools provide some forms of integration with a classifier, which is also relevant to our work. For instance, WordFreak and BRAT discussed above support variations of the active learning approach. JANE by Tomanek et al. [420] is an annotation environment based on the MMAX editor discussed above, which provides support for active learning. The system includes a line chart for monitoring the active learning process, but does not support further visual analysis of the annotated data and the annotation process. Settles [369] introduces an active learning system DUALIST that can be used for sentiment classification, among other options. It provides a web-based annotation interface that focuses on annotating both documents and specific features (in this case, salient terms). DUALIST lacks of support for visualization of the collected data, though, which makes it difficult to get an overview during/after the training. Huang et al. [185] propose a user interface for customer reviews editing, which interacts with an aspect-based sentiment classifier and collects the training data at the same time—however, their interface does not support either active learning or visual analysis of the collected data. Poursabzi-Sangdeh et al. [328] introduce a system called ALTO designed for interactive annotation of texts with topic labels. ALTO uses active learning and topic modeling to suggest candidate labels. In contrast to our approach, it uses a list-based interface to represent the data set which does not support exploratory visual analysis. Finally, the work of Magnini et al. [276] discusses an active learning system TextPro-AL that provides a customizable pipeline including NLP components, annotation editor, and a visualization component for monitoring the active learning process with a line chart. Compared to TextPro-AL, our approach supports a much wider variety of visual analysis tasks and representations. Our system ALVA is also focused (at least currently) on annotation and visualization tasks related to stance classification, which is not addressed by the tools discussed above, even though some of them could potentially be adapted/applied to this problem.

5.2.2 Visual Support for Annotation and Classification

Our work is relevant to the visual support of data annotation and classifier training. Thus, previous work includes contributions from the visualization community as well as CL and ML communities.

5.2.2.1 Existing Approaches in CL/ML

Ware et al. [462] discuss an interactive visual interface for creating decision trees using node-link diagrams, scatterplots, and bar charts. In contrast to our work, their approach is not applicable to high-dimensional data, and it also requires the data to be labeled, i.e., it does not provide the annotation functionality. Kranjc et al. [228] describe how active learning can be achieved for sentiment analysis of data streams with a data mining platform ClowdFlows. The data annotation is performed with a text-based interface, and the results of sentiment analysis are visualized with line/area plots and word clouds. Compared to our approach, ClowdFlows does not provide means for visual exploration of the collected annotated data. Li et al. [253] describe a visual interface for annotating text data. They support two separate tasks: clustering text chunks relevant to the same named entities by dragging nodes in a force-directed graph, and constructing a parse tree by using a tree visualization and additional text interface. Our approach differs in the granularity of annotations, the set of categories, and the user tasks for visual analyses.

5.2.2.2 Existing Approaches in InfoVis/VA

Related work in visualization includes the approach described by Brooks et al. [52, 245] which includes a collaborative annotation environment supporting multiple non-exclusive affect categories. Their environment provides visual representations of annotations over time by using a calendar view and a timeline view. Our approach is different with regard to the set of used categories and our focus on combinations of categories. The follow-up work of Torkildson [424] presents a novel representation of confusion data to facilitate ML classifier training. In contrast to this approach, our work mostly concerns the annotated data exploration. Lu et al. [271] describe an approach that focuses on feature selection and training of various ML models for predictive tasks involving social media data. While the overall goal of our approach is also to facilitate an ML classifier training, we currently focus on different tasks such as visual analysis of annotated data. The work of Makki et al. [278] focuses on sentiment lexicon expansion/annotation with a visual interface. The users can validate and edit predicted sentiment values of individual words by using several visual representations: a tree cloud and a scatterplot with embedded word clouds. In contrast, our work focuses on different categories (stance rather than sentiment) and a different data granularity (utterances rather than words). Another approach is discussed by Heimerl et al. [172] who provide an interactive interface for training a text classifier for two mutually exclusive categories: “Relevant” and “Irrelevant”. Their approach includes support for data annotation and several visualizations based on dimensionality reduction of feature vectors and classification confidence values. The main difference of our approach lies in the choice of categories and corresponding tasks. Our stance annotations are effectively vectors with eleven binary dimensions (ten stance categories + IRRELEVANT). The usual representation of dimensionality reduction results with scatterplots, as used by Heimerl et al., does not support the tasks important for our users. Similar considerations also apply when comparing our work to the *inter-active* learning approach by Höferlin et al. [178], who integrate active learning for video data ad-hoc classification into a visual analytics system. Besides the differences in data types, categories, and visual analysis tasks, our approach in general separates the roles of annotators and analysts to provide the former with a simpler and non-overwhelming environment.

5.2.3 Stance Visualization

The specific application in the focus of our work is stance visualization, which has not been described well in existing work in InfoVis or VA, as discussed above in Section 3.3. In this section, we focus on the few existing visualization techniques and systems that support categories beyond mere polarity or emotions, which are related to stance.

For instance, Small [394] analyses the maps of science based on sentiment analysis of citation contexts in scientific literature. Several categories used for the analysis are clearly related to stance, for instance, UNCERTAINTY and DIFFERENTIATION (contrast). Compared to our contribution, the proposed approach does not support the tasks related to annotation and active learning.

Almutairi [10] discusses the application of the *appraisal* framework to text analysis and visualization. The resulting system AppAnn supports data annotation, classification of text fragments with regard to dimensions of AFFECT, JUDGMENT, and APPRECIATION, and several visualizations. In contrast, our work uses a different and larger set of categories which requires a different approach to annotated data representation, and it also supports active learning.

Our previous work [243] described in Chapter 4 introduces the problem of stance visualization and proposes a VA approach, called uVSAT. Based on sentiment analysis, uVSAT allows the users to analyze the occurrences of stance markers in social media texts using lexicons for six emotion categories as well as CERTAINTY and UNCERTAINTY. However, uVSAT does not support the stance annotation process, visual analysis of annotated utterances, and integration with an active learning classifier. Our domain experts used uVSAT to detect the promising social media texts, which were imported to ALVA as the source data for the annotation stage.

Mohammad et al. [296] provide a dashboard visualization of the annotated stance data set used for the SemEval-2016 contest, which includes annotations for sentiment and FOR/AGAINST stance. The visualization provides an overview of the data set with regard to class distributions by using several bar charts, a tree map, and confusion matrices. The annotated tweets are available in a text table. While this visualization provides a certain degree of interactivity with filtering, brushing, and linking [219], it does not allow the users to spot the patterns occurring in the annotated data as our visualization approach does. Furthermore, ALVA does serve not only as the starting point for the classifier training, but rather as an environment for annotators, linguists, and CL experts, which is used during the whole training period.

Finally, ConToVi by El-Assady et al. [115] provides an animation-based visualization of conversation transcripts, e.g., political debate transcripts, which allows the users to monitor the stance of individual speakers with regard to specific topics. ConToVi focuses on stance categories related to the argumentation, such as COMMON GROUND and MINIMAL CONSENSUS. It also employs categories such as SENTIMENT, POLITENESS, CERTAINTY, and ELOQUENCE to describe the speaker's behavior. Our approach differs from ConToVi in the set of categories, and more importantly, in the data origin and corresponding tasks: ALVA is designed to support the annotation and classifier training stages, rather than to be used with the final classifier for monitoring and analysis of incoming source data.

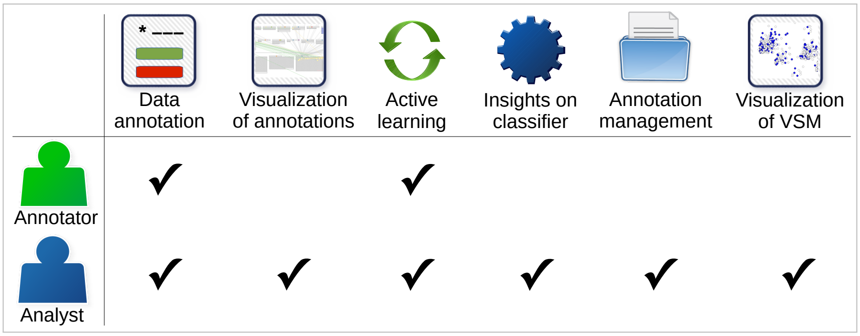


Figure 5.2: The overview of supported tasks from the user perspective. *Reprinted from [242] © 2017 ACM.*

5.3 System Overview and Visualization Tasks

Our approach has been designed to address some of the challenges of stance analysis discussed in Section 2.6 and Section 5.1. Figure 5.2 illustrates a range of tasks supported by ALVA, which take two different user roles into account: *annotators* who work only with the annotation interface, and *analysts*, our domain experts in linguistics and CL who oversee the whole process. The design choices for ALVA were initially motivated by the need to collect a training data set of utterances (text chunks) annotated with our set of stance categories. Since it was evident that we were facing data sparsity problems for some categories, we decided to follow the *active learning* approach [370, 422]. It meant that the classifier would have to suggest candidate utterances to be annotated next based on certain metrics, which would make the training process more efficient than selecting candidates randomly would. This approach requires the annotation interface to be integrated not only with a database of utterances, but also with the classifier. The natural next step was to introduce several visualizations to monitor and analyze the collected annotation data, the annotation process itself, and the classifier performance after the corresponding active learning rounds. Furthermore, additional visualizations can be added to analyze some specific aspects of the data, such as vector space models (VSMs) computed for the utterances. As demonstrated in Figure 5.1, the workflow in ALVA follows two main loops: (1) a human-computer interaction loop, and (2) the active learning loop triggered by the computational analyses.

5.3.1 System Architecture

The architecture of ALVA is affected by some loosely coupled external components, as depicted in Figure 5.3. The source data for our approach comprises social media texts collected with our previous tool uVSAT [243], which are converted to

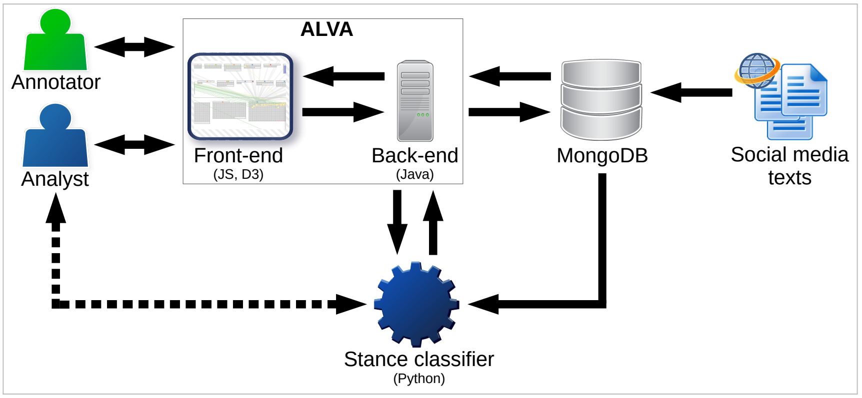


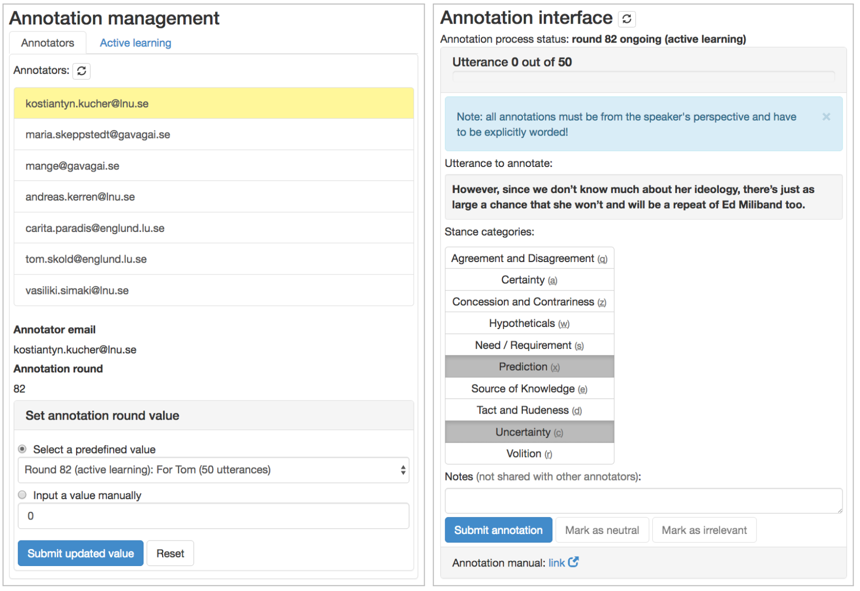
Figure 5.3: Architecture of ALVA. The dashed line indicates interaction between CL experts and the classifier, which is outside of the scope of our system presented in this chapter. *Reprinted from [242] © 2017 ACM.*

plain text, divided into utterances (in most cases, corresponding to sentences), and saved into the database. We are using MongoDB [298] to store these utterances as well as annotations and other data in a flexible way. The stance classifier is developed in Python with Scikit-learn [321] by our CL experts, and it provides a RESTful API [131] with Flask [132] to request text classification, active learning candidates, or historical performance results. The classifier itself consists of multiple support vector machines (SVMs) [326] for individual stance categories. Finally, the core components of ALVA comprise a back end developed in Java with Spark Web Framework [398] and a front end developed in HTML/JavaScript with Bootstrap [38], jQuery [205], D3 [94], and Rickshaw [338]. All the components use the JSON [206] format to transfer data. The front end[3] provides a tab-based web interface with a set of tabs corresponding to the respective user's privileges. The tabs provide interfaces for user details configuration, annotation, annotation management, table-based annotation summaries, annotated data visualization, and VSM visualization. We discuss some of these tabs below, starting with the interfaces related to the annotation process.

5.3.2 Annotation Process

The annotation process started with a number of annotation rounds which did not involve active learning. Candidate utterances were selected randomly from a pool of available utterances. To collect the initial data, multiple project members acted as annotators. Using the annotation management interface depicted in Figure 5.4(a), the annotators were assigned with certain annotation rounds.

[3] A demo video for ALVA is available at `https://dl.acm.org/citation.cfm?id=3132169` or `https://vimeo.com/230645678` (last accessed in February 2019).



(a) **(b)**

Figure 5.4: Screenshots of the web-based interfaces of ALVA. (a) The annotation management interface for assigning annotation rounds to individual annotators. (b) The annotation interface: here, the annotator has selected the categories PREDICTION and UNCERTAINTY for the utterance presented. *Reprinted from [242] © 2017 ACM.*

Several annotators were assigned with the same round to test the *inter*-annotator agreement (i.e., annotators A and B worked on the same set of utterances S during round X). Also, some annotators were assigned with a round with the same data after some time to test the *intra*-annotator agreement (i.e., annotator A was assigned with the same set S during rounds X and Y). This resulted in a data set where an utterance corresponds to an arbitrary number of annotations. After collecting the initial data and preparing the classifier, we have started to follow the active learning approach. The candidate utterances for an annotation round are now selected by the classifier based on distance to the SVM hyperplane.

An annotator is presented with the interface displayed in Figure 5.4(b). Each utterance can be either marked as IRRELEVANT (for invalid utterances such as text in a different language, code, URLs, etc.), or labeled—in our case—with up to ten stance categories such as AGREEMENT AND DISAGREEMENT, VOLITION, etc. The annotator may select several categories at a time. If no category is selected, i.e., the utterance is tagged as NEUTRAL, the annotator must use the corresponding button to confirm this explicitly. The annotator can also add some private notes about his/her choices, which are later available to analysts. Finally, the interface includes a link to the annotation manual prepared by our experts in linguistics.

The current data set comprises about 10,000 annotations of utterances in English (in most cases, individual sentences) collected from social media (blogs and forums) on political topics such as the US presidential election 2016 and Brexit. The statistics for individual categories are presented in Table 5.1. While about 4.3% of all annotations are marked as IRRELEVANT and the majority of annotations are marked as NEUTRAL (about 39.9%), the rest of the values clearly illustrate our biggest issue—the data is very sparse with regard to meaningful stance categories. For instance, AGREEMENT AND DISAGREEMENT is currently present only in 3.6% of all annotations, making it unlikely to get good classification results. The need for further analysis of the available data leads us to the discussion of visualization tasks.

5.3.3 User Tasks for Visualization

The main users of visualization components in ALVA are the members of our research project who are interested in linguistic and computational aspects of stance analysis. These analysts are (1) a professor, (2) a postdoctoral researcher in linguistics, (3) a senior scientist, and (4) a postdoctoral researcher in computational linguistics. As part of the iterative design process, we have identified a number of user tasks based on the discussions and requests from these users. These tasks for visualization components reflect some of the general challenges of stance analysis and visualization outlined in Section 2.6 and Section 5.1. Some of the tasks are in principle not exclusive for the problem of stance analysis, e.g., the tasks related to individual category distributions or annotator agreement could be generalized for the problems of supporting active learning or text classification with non-exclusive categories in general. However, our design study has been focused on the concrete problem with the design choices motivated by the annotation schema and the data at hand.

Most of the user tasks for visualization in ALVA are related to the exploratory analysis of individual annotations and distributions of stance categories used in the annotations. Our analysts want to be provided with an overview of the overall data set, analyze the co-occurrence of categories, identify interesting cases related to multiple categories, and compare annotations made for the same utterances. Furthermore, some of their requests go beyond the scope of the annotations data per se. For instance, they would like to monitor the annotation process and the active learning process. Some of the concrete tasks presented by our analysts are as follows:

T1 Are there many annotations marked as NEUTRAL or IRRELEVANT? This is important in order to estimate the complementary proportion of annotations with meaningful stance categories.

T2 What is the distribution of individual stance categories in the annotation data?

T3 Are there many annotations labeled with multiple stance categories, rather than a single one?

T4 Which stance categories tend to co-occur in annotations? This task together with T3 are of particular interest to our experts in linguistics, since the discovery of stable patterns can lead to advances in the understanding of stance.

T5 Is it possible to compare annotations made for the same utterances?

T6 What is the overall status of the annotation process with regard to annotators, annotation rounds, and relations between them?

T7 What are the inter- and intra-annotator agreement values for comparable annotator/round pairs?

T8 What is the overall average agreement for each stance category in the annotation data?

T9 How has the classification performance changed over time during the active learning process?

T10 Is there any observable correlation between the annotated stance categories and lexical content of utterances?

Some of these tasks on their own could be solved without employing visual analysis, e.g., T1 and T2 could be solved with a spreadsheet editor, and T3 could be solved with a Python script. However, even in these cases a significant effort would be required from the users to export/convert the data annotated with one of the existing tools (if supported at all), or even programming skills, which constitutes a problem for users such as linguists. Moreover, the tasks such as T4–T6 presuppose the analysis of rather large sets of annotations with non-trivial combinations of categories, which in turn requires non-trivial representation and interaction techniques [219]. In order to support these tasks, we have designed and developed several visualizations using multiple linked representations that are discussed in the following section.

5.4 Visualization Methodology

We start this section with a discussion of the novel visual representation used for the annotation data. Then, we describe the complete annotation visualization interface of ALVA which includes multiple coordinated representations. Finally, a separate interface for VSM visualization is briefly discussed in relation to the corresponding user task.

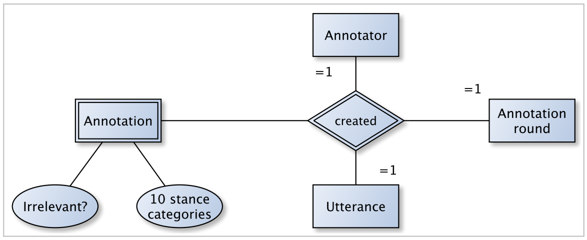


Figure 5.5: Entity–relationship diagram for text annotations in ALVA. *Reprinted from [242] © 2017 ACM.*

5.4.1 CatCombos Representation Design

As discussed in the previous section, text annotations were performed by several annotators during multiple annotation rounds. Figure 5.5 demonstrates how a single annotation corresponds to a combination of annotator, annotation round value, and actual utterance. Each annotation can be labeled with up to ten stance categories or marked as IRRELEVANT in our concrete use case, which can be treated as a vector of 11 bits for convenience. This data schema results in a multidimensional data set presenting multiple visualization challenges. For instance, color coding is not feasible for individual annotation items since the total number of possible category combinations is $2^{10}+1$. Our main motivation for designing the appropriate representation was to provide the users with an overview of the whole data set with regard to categories, and provide details on demand for individual annotations at the same time.

During the initial prototyping stage, we have considered a scatterplot representation based on a dimensionality reduction (DR) technique [267]. Our basic prototype used a PCA projection [204,367] for the bit vectors representing categories selected for each category. It was easy to identify a number of issues with that approach. First and foremost, the color coding problem mentioned above meant that we could not simply map the choice of categories to the color attribute. Instead of a direct mapping, we considered introducing a star plot or another glyph [40] for each annotation item, but this approach was not capable of scaling to thousands of visual items, i.e., the users would not get an overview. Such glyphs would need to have a rather large size to represent 11 dimensions, which leads us to the problem of occlusion encountered with the prototype. The input data for the DR technique in our case has low variance due to its binary and unbalanced nature, which resulted in severe occlusion problems. We used *jittering* as clutter reduction strategy [118,460], but the overall result was still not satisfactory, especially due to the color coding issue.

Instead of experimenting with other DR techniques for this data set, we considered the user tasks concerned with an overview. It was evident that

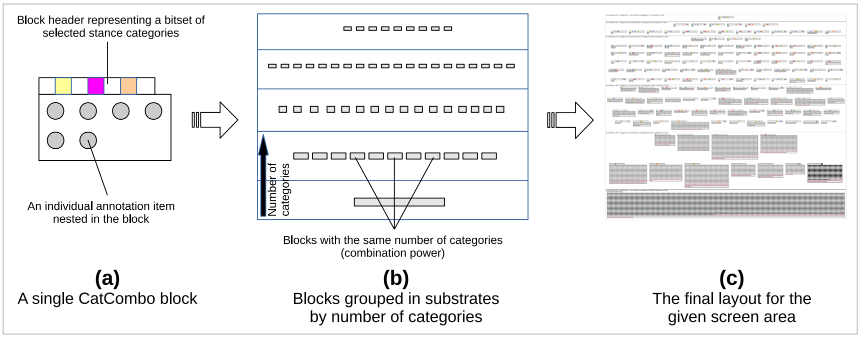


Figure 5.6: The design of the CatCombos representation. *Reprinted from [242] © 2017 ACM.*

tasks T1–T4 paid special attention to the combinations of categories selected by annotators or, in special cases, individual categories rather than combinations. Therefore, we have decided to ignore the spatial positions of individual annotation items, and focus instead on groups of items. The details about individual items and the links to relevant items could be displayed on demand, resembling the semantic substrates technique by Shneiderman and Aris [375].

Figure 5.6 displays the main steps of our new visual approach called *CatCombos* ("Category Combinations"). Individual annotations (represented by dots or, in case of active learning annotations, diamonds) are grouped together into rectangular blocks by the combination of categories which occur in the data set. A single CatCombo block depicted in Figure 5.6(a), therefore, corresponds to a certain combination of categories. Color-coded rectangle labels in the block headers represent the corresponding sets of categories, using a modified color map from ColorBrewer [85]. The design of blocks/headers was inspired by the set visualization technique proposed by Sadana et al. [352].

Individual blocks are, in turn, grouped and laid out by the number of corresponding categories. Thus, the groups of blocks form substrates, cf. Figure 5.6(b): the top substrate contains blocks labeled with multiple categories simultaneously, and the bottom substrate with a single block contains only NEUTRAL annotations corresponding to zero selected categories. The left-to-right order of blocks in substrates is based on the bit set interpretation of sets of categories, e.g., presence of AGREEMENT AND DISAGREEMENT is treated as the most significant bit, and IRRELEVANT—as the least significant bit. This ordering is stable with regard to concrete category combinations present in the data.

The final layout takes the available area dimensions into account to specify individual block ratios and distribute blocks into layers in each substrate. It can be recalculated on events such as the window resize to use the available space in an efficient way, for instance, to fit the overall layout in a square

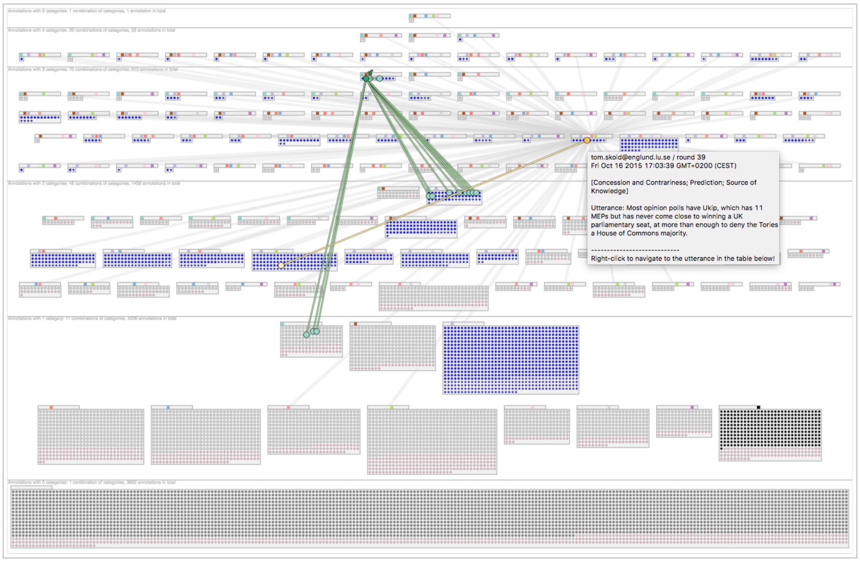


Figure 5.7: Example of a CatCombos visualization for the complete data set of about 10,000 text annotations. Here, the user has highlighted all annotations related to the category CONCESSION AND CONTRARINESS represented by blue dots using the filtering panel (see Figure 5.8 and Section 5.4.2). Yellow and green links connect the currently highlighted (by hovering) and pinned (by clicking) annotation items to other annotations made for the same utterances, respectively. *Reprinted from [242] © 2017 ACM.*

shape (see Figure 5.6(c)). Investigation of alternative layouts which could reveal groups/clusters of CatCombo blocks in each substrate is considered by us to be part of future work, although such layouts could lead to higher visual complexity and higher space consumption.

Color coding is used sparingly for individual visual items. Most annotations are colored gray, while dark gray and black are used for NEUTRAL and IRRELEVANT annotations, respectively. Other color coding choices include red color for active learning annotations and several colors used for dynamic highlighting (see below).

Figure 5.7 shows a CatCombos view of the complete annotations data set. The resulting visualization gives an overview of the stance category distribution and the characteristics of annotations important for our analysts. For instance, it is easy to estimate that the majority of annotations are contained in blocks at the bottom which correspond to a single selected stance category or no category at all, i.e., NEUTRAL, thus supporting user tasks T1 and T2 (partially). The interesting cases where an annotator has used multiple categories simultaneously are located

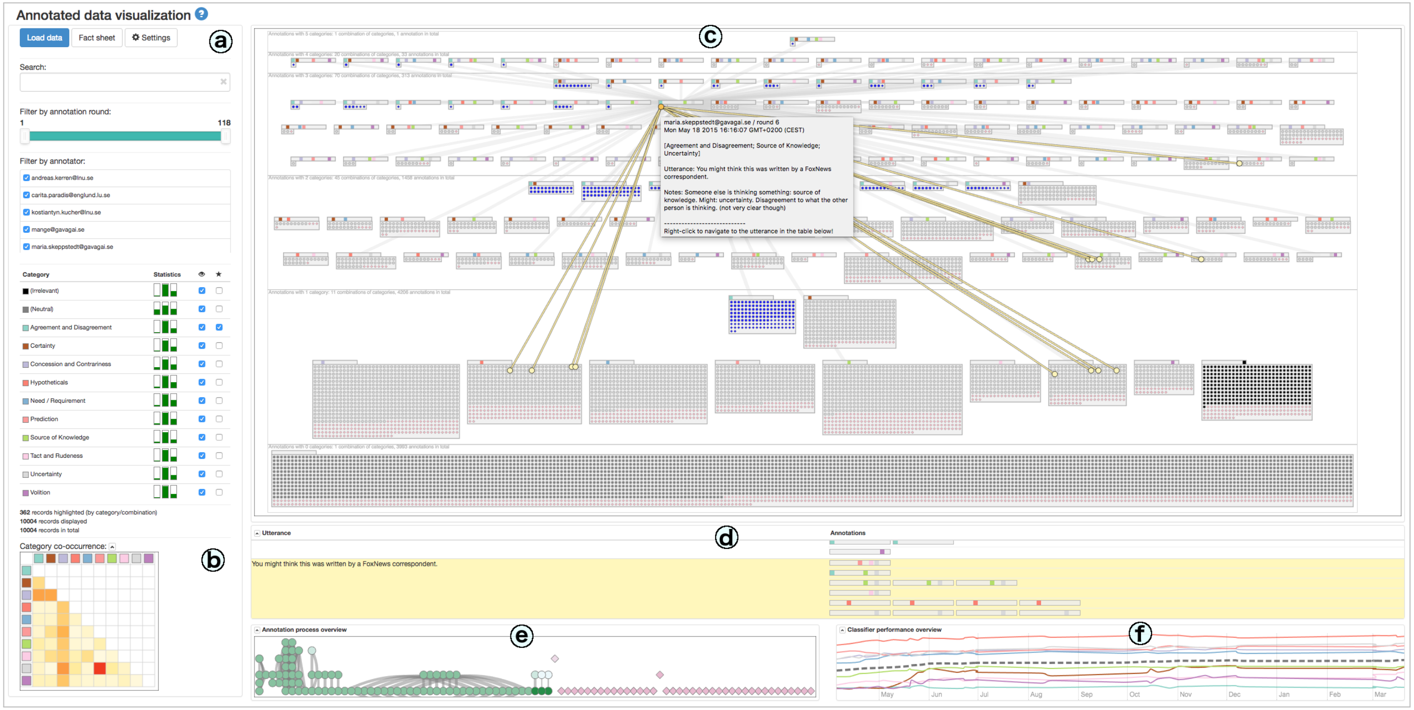


Figure 5.8: The annotated data visualization screen in our VA system ALVA: (a) the filters panel; (b) the category co-occurrence matrix; (c) the main canvas with a CatCombos visualization in which all annotations related to the category AGREEMENT AND DISAGREEMENT are highlighted in blue; (d) the utterances and annotations table; (e) the annotation process overview graph; and (f) the classifier performance plot. *Reprinted from [242] © 2017 ACM.*

in the top blocks. By investigating such cases, users can carry out tasks T3 and T4.

The visualization supports pan & zoom, dynamic queries, highlighting, and details on demand for individual blocks and annotations (some of these features are discussed in more detail below). For instance, while task T2 is not directly supported for all categories, it is possible to use color coding to highlight annotations with a certain category. In this case, the visual items are displayed in blue.

By hovering over an annotation item (highlighted with orange), the users can view its details and see links to other annotations for the same utterance (highlighted with yellow). Simultaneously, semi-transparent gray links to blocks with related categories of combinations are displayed for the enclosing block (their opacity can be adjusted by the user in the settings dialog). It is also possible to preserve the selection of individual items by clicking. In this case, green color is used for the items and links, and a pin icon is displayed near the selected item. These features facilitate the exploration related to user task T5.

5.4.2 Visualization Interface in ALVA

Figure 5.8 displays the overall visualization interface of ALVA for the annotation data. The users initiate visualization by using the "Load data" button in the left panel (Figure 5.8(a)). This panel also provides the means to open a dialog box with a text summary of the annotation process and collected annotations (see Section 5.5). The users are also provided with a number of filters affecting the visualization, including a text search field, annotator filters, and stance category filters. It is also possible to highlight annotations by category, as discussed above: for instance, annotations with the category AGREEMENT AND DISAGREEMENT are highlighted in blue in the main CatCombos visualization view (Figure 5.8(c)).

5.4.2.1 Statistical Charts

Besides filtering and highlighting controls, there are also several sparkline-style [432] bar charts (colored in green) providing basic statistics for each stance category. The left chart displays the proportion of annotations with the respective category in the overall data set (see Table 5.1) to support user task T2. The other two charts convey average agreement values for annotations (see Section 5.1). The middle chart displays the average observed agreement for the respective category. This value is calculated as average over observed agreement values for comparable pairs of records associated with the same utterance, i.e., pairs of various annotators and/or various rounds. The right-hand chart displays the average chance-corrected agreement calculated for the same pairs of records. More specifically, it is calculated as average over Cohen's kappa values. These statistics facilitate the user in carrying out task T8.

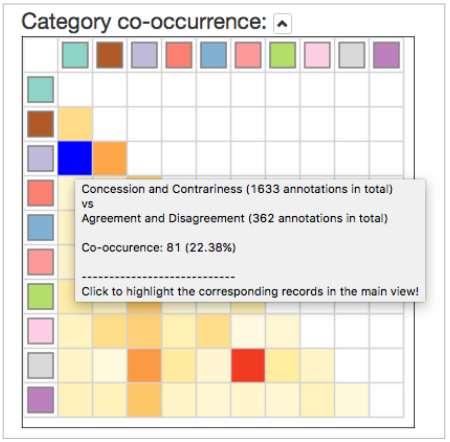


Figure 5.9: The category co-occurrence matrix. Cell color encodes the proportion of annotations with the corresponding pair of categories. Here, the user has clicked the cell representing the combination of AGREEMENT AND DISAGREEMENT and CONCESSION AND CONTRARINESS (colored in blue) to highlight the corresponding items in the main CatCombos view. *Reprinted from [242] © 2017 ACM.*

5.4.2.2 Category Co-occurrence Matrix

The left panel also contains an auxiliary view, the category co-occurrence matrix, displayed in Figure 5.8(b) and Figure 5.9. This matrix is designed to facilitate the task T4 by visualizing the proportion of annotations with a certain pair of categories with respect to the less frequent category of the pair. Row and column headers use the same color labels as the filters table and CatCombo headers to represent stance categories. Matrix cells use a color map ranging from white to red to encode co-occurrence proportion values. For instance, the largest value currently observed in our data set is represented by a bright red cell (see Figure 5.9). It corresponds to the categories of PREDICTION (933 annotations) and UNCERTAINTY (811 annotations) with 322 co-occurrence cases (39.7%). The users can explore the details on hover and click the cells to highlight the corresponding annotations in the main CatCombos visualization view for detailed exploration. Figure 5.9 displays how the user has hovered and clicked on a cell corresponding to the combination of AGREEMENT AND DISAGREEMENT and CONCESSION AND CONTRARINESS, which is colored in blue.

5.4.2.3 Utterances Table

While the main visualization view provides the actual utterance texts for annotation items on hover, our analysts requested an additional view focusing on

Figure 5.10: The utterances table. By looking at each row, it is possible to compare the corresponding stance annotations made for the same utterance. The annotations are grouped by the set of categories and ordered by the number of categories. *Reprinted from [242] © 2017 ACM.*

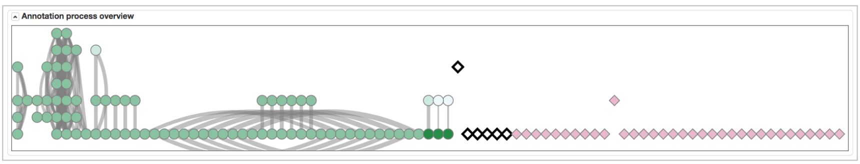


Figure 5.11: The annotation process overview graph. Edges reveal the usage of the same utterances data for regular annotation rounds (marked by green). At the same time, the active learning rounds (marked by red) involve candidate utterances without prior annotations selected by the classifier, hence the lack of edges. *Reprinted from [242] © 2017 ACM.*

the textual data in order to carry out task T5. The utterances table depicted in Figure 5.8(d) and Figure 5.10 contains the list of utterances and the corresponding annotations, represented in the same way as CatCombos headers to preserve the users' mental map. The specific details about annotations are available on hover. The annotations for each utterance are grouped by the set of categories and ordered by the number of categories. The grouping can be adjusted by the user in the settings dialog to achieve a more compact layout. The utterances are themselves ordered by the total combined number of categories used in the annotations. This way, the utterances initially displayed at the table top correspond to interesting cases worth exploring. The table is affected by global filtering, and it is related to the main view by linking & brushing. It is also possible to navigate to the corresponding table row by right-clicking an annotation item in the main view.

5.4.2.4 Annotation Process Graph

The representations described above focus mostly on the annotation *data*. At the same time, our analysts were interested in the status of the annotation *process* (task T6), i.e., how many annotation rounds were created, which annotators were assigned to which rounds, and so on. The annotation process graph depicted in

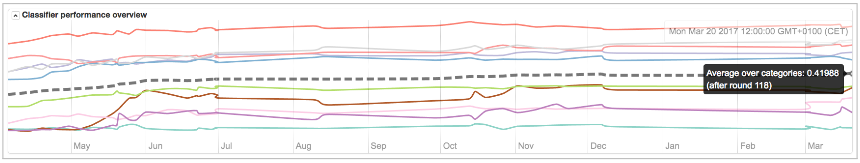


Figure 5.12: The historical classifier performance plot. The tooltip for the thick dashed line reveals the average value over all categories after the latest active learning round. *Reprinted from [242] © 2017 ACM.*

Figure 5.8(e) and Figure 5.11 provides an overview of this process. Each node corresponds to a single annotator / annotation round, for instance, "Maria / annotation round 42". The node shape and hue encode the annotation round type: green circles and red diamonds encode regular and active learning rounds, respectively. The saturation of a node encodes the corresponding number of annotations. The nodes are laid out with regard to the annotation round number (left to right) and the annotator (top to bottom). The links between pairs of nodes denote an overlap in the set of annotated utterances. For such cases, both observed and chance-corrected agreement (Cohen's kappa, see Section 5.1) are calculated to support task T7, and the values are available on hover. The width of a link is proportional to average Cohen's kappa over all categories. In order to address the cluttering issue without allocating additional area to this graph, a fadeout effect is used when hovering over nodes and links. The graph is also affected by global filtering, and its individual nodes can be used to filter out the corresponding annotations by clicking. The disabled nodes are colored white and marked with a thick stroke.

5.4.2.5 Classifier Performance Plot

Finally, the visualization interface provides one more view not directly related to the annotation data. After each active learning round is finished and new annotations are submitted to retrain the classifier, the resulting performance values are stored in the database. The plot in the bottom right side of Figure 5.8(f) (also see an enlarged cutout in Figure 5.12) provides an overview of these values with line plots implemented with Rickshaw [338]. Each data point represents the F1 score for the corresponding category's classifier after the corresponding active learning round. The color coding for stance categories used here is the same as in the CatCombos headers. The average values over all categories are presented with a thick dashed line colored in dark gray to show a quick summary of classifier training progress, thus completing user task T9.

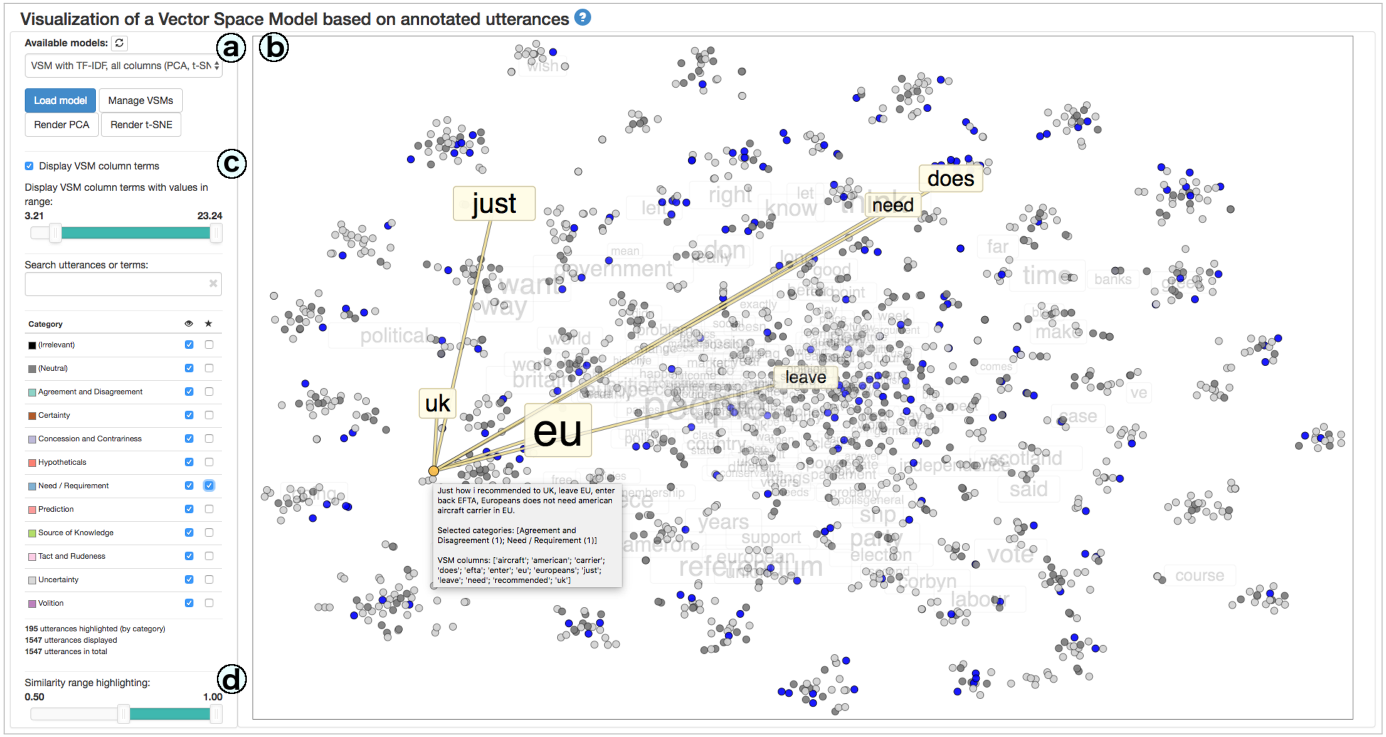


Figure 5.13: The VSM visualization screen in our VA system ALVA: (a) the filters panel; (b) the main canvas with a t-SNE visualization of utterances (dots) and terms (here, all utterances related to the category NEED/REQUIREMENT are highlighted in blue); (c) the controls for terms display; and (d) the similarity range slider for highlighting control. *Reprinted from [242] © 2017 ACM.*

5.4.3 VSM Visualization Interface in ALVA

One of our analysts' requests for ALVA (namely, user task T10) was related to the assumption about relation between the lexical content of utterances and stance categories used in the corresponding annotations—after all, lexical features play an important role for our classifier. Our CL experts have therefore implemented functionality for computing a vector space model (VSM) [279,435] based on word counts in annotated utterances. The model is essentially a document-term matrix with rows corresponding to individual utterances, and columns corresponding to terms (words) detected in the utterances. Initially, raw term counts were stored in the matrix, but later TF-IDF weighting [354] was implemented to decrease the effect of common words. Such a matrix can be used for analysis of similarity between utterances. It can also be visualized using one of the DR techniques: the number of columns, i.e., dimensions, in our model is around 5,200.

Based on the recommendations by Sedlmair et al. [367], we decided to support two DR techniques to offer the users several exploration options, namely, PCA [204] and t-SNE [275]. While the projection data can be used to visualize utterances as points in a scatterplot, we also made use of terms present in the matrix. By calculating weighted average positions over the related utterances, it became possible to place the actual terms into the same projection space.

Figure 5.13 displays the resulting interface of the prototype VSM visualization in ALVA. It follows the same design as the annotation data interface with a control panel on the left. Since VSM and DR computations take a considerable amount of time, they are not calculated on the fly. Instead, the users request the model creation and specific DR projection calculations explicitly using the buttons visible in Figure 5.13(a). The computational results are stored into the database and can later be loaded for visualization.

The main visualization view in Figure 5.13(b) demonstrates a t-SNE projection of a data subset consisting of about 1,500 utterances. Each utterance is represented by a dot which is colored similar to the CatCombos view: dark gray or black for utterances with only NEUTRAL or IRRELEVANT annotations, or gray in other cases. Highlighting by category is also similar to the CatCombos view: the utterances relevant to the category NEED/REQUIREMENT are highlighted in blue in the screenshot. Beside the utterances, the visualization includes the terms represented by text labels. The position of each term is computed as weighted average over the corresponding utterance positions (see above), using the values from the VSM as weights. The sum of such weighted counts from the VSM is used as the combined value for a term, and it is mapped to the font size. Since terms can easily occlude the view, it is possible to control the range of displayed values or hide them altogether using controls depicted in Figure 5.13(c). Additionally, the visual clutter is reduced by fading out unrelated terms on hover as part of the highlighting interaction (see below).

The VSM visualization supports highlighting (on hover) and selection pinning (on click) similar to the CatCombos visualization. However, it is easy to define whether an utterance and a term are related, but it is not so obvious for pairs of utterances. The users can therefore control the similarity range using the slider displayed in Figure 5.13(d). The similarity between two utterances is calculated using the Jaccard index [279] for the corresponding sets of column terms.

5.5 Case Study: Annotation Data Exploration

As previously mentioned in Section 5.3.3, our analysts are mainly interested in the analysis of annotation data which could reveal patterns about stance categories distribution formulated as user tasks T3 and T4. As discussed in Section 2.6 and Section 5.1, the research of notional stance categories in linguistics is one of the general challenges of stance analysis. Therefore, the evidence about co-occurrence of stance categories in the data is important for our analysts for refining the theoretical stance framework and improving the classification model. These analysts are the members of our research project specializing in linguistics and CL. While all the project members, including the researchers in InfoVis/VA, have participated in several initial rounds of annotation for training and data collection purposes (see Section 5.3.2), the majority of annotations were later carried out by a postdoctoral researcher in CL and an external annotator with background in linguistics (see the bottom row of Figure 5.11) who are currently responsible for about 19% and 64% of the overall data set, respectively. Therefore, our project members interested in linguistic and computational aspects of stance analysis have to rely on ALVA in order to gain insight about the collected data.

In this section, we describe the results of an exploratory visual analysis of the complete annotated data set as of March 21, 2017 with interesting stance category combinations in mind. We shall try to understand "the big picture" and identify specific combinations in the context of substrates with multiple categories, thus illustrating ALVA's support for user tasks T1–T5.

5.5.1 Initial Findings

After loading the data set, the analysts are presented with the CatCombos visualization interface (see Figure 5.8). Besides the visual representations, some of the basic facts, data statistics, and tips about the visualization are presented in the fact sheet dialog. The data set currently comprises about 10,000 annotation records based on about 5,340 utterances. As previously noted in Table 5.1, IRRELEVANT and NEUTRAL annotations constitute about 4.29% and 39.91% of the data set, respectively. Most of the utterances have been annotated only once or twice (about 2,600 and 2,300 cases, correspondingly). The total number of unique combinations of categories occurring in the data is 148. More detailed statistics

Table 5.2: Statistics on simultaneous occurrence of multiple stance categories in the annotation data set

Combination power	**Number of category combinations**	**Number of annotations**
0	1	3,993
1	11	4,206
2	45	1,458
3	70	313
4	20	33
5	1	1

Note: The total size of the data set is 10,004 annotations, and there are 148 unique category combinations in total. *Reprinted from [242] © 2017 ACM.*

about category combinations in Table 5.2 reveal that the main body of annotations have only 0–2 categories associated with them. This is a predictable result, but it still implies that two possible cases can be interesting for the analysts: first, those few cases that have the largest number of associated categories, and second, the cases of category combinations that have much larger numbers of annotations compared to other combinations with the same power, i.e., number of categories.

5.5.2 Combination of 5 Categories

Figure 5.14 shows the actual CatCombos view of the data set that conveys the same message about the distribution of category combinations. The analysts can quickly estimate the number of annotations in the two bottom substrates, but it is the CatCombos blocks at the top which are interesting. The areas marked with ellipses correspond to interesting cases with regard to category combination power and number of annotations (also provided in Table 5.3). First and foremost, the topmost block with a single item catches the eye (Figure 5.14(a)). This is the only annotation in the data set labeled with 5 categories simultaneously, namely, AGREEMENT AND DISAGREEMENT, CERTAINTY, NEED/REQUIREMENT, SOURCE OF KNOWLEDGE, and TACT AND RUDENESS. The actual utterance text (available by hovering or by looking into the utterance table) is the following:

> *"- might want to avoid numbers and scottish politics altogether since you quite obviously don't have a f***ng clue what you are talking about."*[4]

[4]Retrieved August 9, 2015 from `http://scotgoespop.blogspot.com/2015/06/why-adam-ramsay-is-wrong-to-claim-that.html`

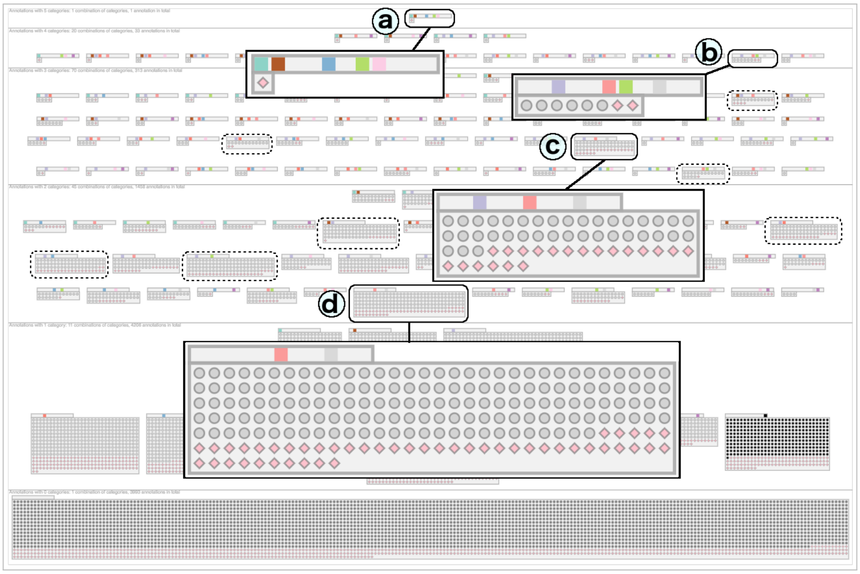


Figure 5.14: The CatCombos visualization of the annotation data set with some particularly interesting cases marked.

5.5.3 Combinations of 4 Categories

Next, the examination of the substrate with combinations of 4 categories reveals that most of such combinations occurred only in one or two annotations. One notable exception is the CatCombo block in Figure 5.14(b) with seven annotations, which corresponds to the following categories: CONCESSION AND CONTRARINESS, PREDICTION, SOURCE OF KNOWLEDGE, and UNCERTAINTY. Despite the small sample size, the analysts can make a tentative assumption about existence of a pattern related to these categories.

5.5.4 Combinations of 3 Categories

The third substrate from the top contains combinations of 3 categories, and these combinations are currently the most numerous in our data set. There are 70 combinations containing the total of 313 annotations. The visual exploration of the corresponding CatCombos reveals, however, that most combinations have only several nested annotations. At the same time, there are several interesting cases with numerous annotations (marked with dashed lines). The most prominent is the CatCombo block in Figure 5.14(c) with 57 annotations. The corresponding categories combination is CONCESSION AND CONTRARINESS, PREDICTION, and UNCERTAINTY. According to Table 5.3, the average number of annotations in the corresponding substrate is only 4.47. The relation between the categories

Table 5.3: Interesting stance category combinations discovered in the annotation data set

Category combination	Number of annotations	Average in substrate
{AGREEMENT AND DISAGREEMENT, CERTAINTY, NEED/REQUIREMENT, SOURCE OF KNOWLEDGE, TACT AND RUDENESS}	1	1
{CONCESSION AND CONTRARINESS, PREDICTION, SOURCE OF KNOWLEDGE, UNCERTAINTY}	7	1.65
{CERTAINTY, CONCESSION AND CONTRARINESS, PREDICTION}	16	4.47
{CONCESSION AND CONTRARINESS, HYPOTHETICALS, UNCERTAINTY}	14	4.47
{CONCESSION AND CONTRARINESS, PREDICTION, UNCERTAINTY}	57	4.47
{PREDICTION, SOURCE OF KNOWLEDGE, UNCERTAINTY}	17	4.47
{CERTAINTY, CONCESSION AND CONTRARINESS}	85	32.4
{CONCESSION AND CONTRARINESS, HYPOTHETICALS}	71	32.4
{CONCESSION AND CONTRARINESS, NEED/REQUIREMENT}	80	32.4
{CONCESSION AND CONTRARINESS, SOURCE OF KNOWLEDGE}	122	32.4
{PREDICTION, UNCERTAINTY}	202	32.4

Note: The right column contains the average number of annotations over all combinations in the corresponding substrate. *Reprinted from [242] © 2017 ACM.*

PREDICTION and UNCERTAINTY in our data is further illustrated by the largest block with 2 categories depicted in Figure 5.14(d), which contains 202 annotations (compared to the substrate average of 32.4). This relation is also suggested by the category co-occurrence matrix discussed in Section 5.4.2. These cases definitely provide our analysts with evidence of specific patterns or stance constructions, which they can investigate further with a focus on utterance texts.

5.5.5 Utterances Table Exploration

In fact, the analysts may reverse the flow of analysis completely by looking at the utterances table first instead of the annotations (see Figure 5.8(c)). We provide some examples of utterances with multiple selected categories in Table 5.4, which were collected by exploring the utterances table from the top.

Table 5.4: Examples of interesting utterances with corresponding categories discovered in the annotation data set

Utterance	Annotations
"I am convinced that SYRIZA would have been able to achieve MUCH better results by implementing the things we talked about and not be completely idiotic about their approach." [a]	{certainty, hypotheticals, tact and rudeness} {certainty, hypotheticals, tact and rudeness} {certainty, concession and contrariness, tact and rudeness} {certainty, concession and contrariness, hypotheticals} {certainty}
"But if you put words into my mouth to the extent of the above again, rest assured I'm not adverse to pressing that button." [b]	{concession and contrariness, need/requirement, source of knowledge, volition} {certainty, concession and contrariness, hypotheticals} {hypotheticals, tact and rudeness} {hypotheticals} {hypotheticals}
"Unlike staunch Labourites or dyed-in the-wool Conservatives, if they don't like what they see, they take their vote elsewhere." [c]	{concession and contrariness, hypotheticals} {concession and contrariness, hypotheticals} {concession and contrariness, hypotheticals} {concession and contrariness, hypotheticals} {concession and contrariness, tact and rudeness}
"In sum, epistemic democrats like you should not conflate your own attraction to cognitive diversity with the interests-based sociology of post-Marxist Rousseauians like Yoram." [d]	{certainty, concession and contrariness, need/requirement, tact and rudeness} {certainty, need/requirement, volition} {need/requirement} {need/requirement} {need/requirement}
"I actually disagree, but that is not really the point." [e]	{agreement and disagreement, concession and contrariness} {agreement and disagreement, concession and contrariness} {agreement and disagreement, concession and contrariness} {agreement and disagreement, concession and contrariness} {agreement and disagreement, certainty}

Reprinted from [242] © 2017 ACM.

[a] Retrieved June 18, 2015 from `http://forums.somethingawful.com/showthread.php?pagenumber=262&perpage=40&threadid=3624340&userid=0`

[b] Retrieved June 18, 2015 from `http://www.sheffieldforum.co.uk/showthread.php?page=19&t=1405170`

[c] Retrieved June 18, 2015 from `http://www.telegraph.co.uk/news/general-election-2015/politics-blog/11680714/Heres-why-the-Tories-should-want-Labour-to-pick-a-strong-leader.html`

[d] Retrieved June 18, 2015 from `https://equalitybylot.wordpress.com/2015/06/05/short-refutations-of-common-objections-to-sortition-part-3/`

[e] Retrieved June 18, 2015 from `https://equalitybylot.wordpress.com/2015/06/05/short-refutations-of-common-objections-to-sortition-part-3/`

5.5.6 Summary

In conclusion, this section has demonstrated a typical case study of visual analysis of annotation data in ALVA. The users were able to gain an overview about the current status of the annotated data set with regard to individual stance categories as well as the special IRRELEVANT and NEUTRAL cases, thus illustrating the support for user tasks T1 and T2. These results provide the analysts with insights about the presence of stance phenomena in the social media data and allow them to identify relatively frequent and rare stance categories, which has implications for the classifier training process. The users were also able to identify frequent combinations of several stance categories (user tasks T3 and T4). On the one hand, it provides opportunities for in-depth linguistic investigation, and on the other hand, it can be used in the future to improve the classifier, e.g., by implementing hierarchical features as opposed to the current naïve classification approach which treats the categories as independent. Finally, the users were able to access individual utterances with multiple annotations (user task T5), which provide good examples of various stance categories and could be used in further linguistic analyses. Our analysts have made use of these findings. For example, the study by Simaki et al. [381] has focused on comparing frequently co-occurring category pairs such as PREDICTION and UNCERTAINTY and the corresponding improvements for the classifier. Skeppstedt et al. [393] have discussed co-occurring categories such as HYPOTHETICALS and UNCERTAINTY as part of their analysis of local cue words in the annotated data set. While we have focused on user tasks T1–T5 related to the annotation data exploration in this case study, ALVA's support for other tasks is outlined in Sections 5.4.2–5.4.3 and discussed below.

5.6 Discussion

In this section, we discuss the utility of the specific components of ALVA and the system in general. As mentioned in Section 5.3.3 and Section 5.5, so far the main users of the visual analysis aspects of ALVA have been our research project members with background in linguistics and CL. Therefore, the results below are mainly based on discussions with these users.

5.6.1 Active Learning

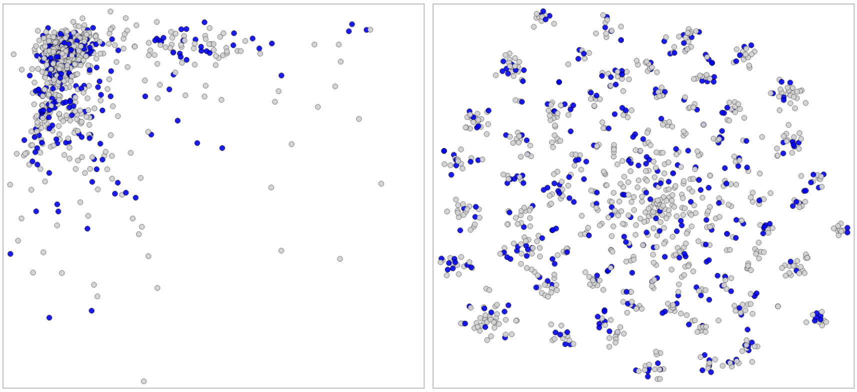
As described in Section 5.4.2, our analysts can use the performance plot to monitor the classifier training process while more training samples are collected after each active learning round in order to support user task T9. Examination of the plot in Figure 5.12 reveals a trend for overall performance improvement. The current average F1 score over all categories is 0.419 after collecting approximately 2,400 training samples with active learning. Performance scores for individual categories tend to fluctuate over time, but it is possible to identify certain trends. The category HYPOTHETICALS has got the best results during the complete training process, currently with an F1 score of 0.668. At the same time, AGREEMENT AND DISAGREEMENT, VOLITION, and TACT AND RUDENESS have got the worst results. The corresponding F1 scores are currently 0.134, 0.214, and 0.292. This can be explained by the sparsity of corresponding training data (cf. Table 5.1).

5.6.2 Annotation Visualization

According to our experts in linguistics, visualization of data is very useful for linguists in general. One important reason is that many linguists today make use of large numbers of texts in corpora containing several billion words. In order to be able to explore, observe, and determine usage patterns in large data sets of natural language, visualization techniques are key. In this particular study, we have examined how speakers take stance in text and for that purpose we set up a framework of ten stance categories. The categories were not defined on the basis of a preconceived list of words assumed to express stance, but the annotators were instructed to identify different types of stance-taking *on the basis of meaning* rather than form. Our particular focus of attention in this study was to find out if the notional categories appear together or not (think of CatCombos), and if they do, which of the categories tend to co-occur. Since the categorization in this case is not made on the basis of words, it is very practical that the visualization tool also includes a module where the user can retrieve the actual utterances for further analysis of what the expressions are like that convey the function of these various stances in discourse. Through visualizing these relations it becomes very clear that some stances combine more often with other stances. We can easily observe patterns such as CERTAINTY is primarily co-occurring with PREDICTION, CONCESSION AND CONTRARINESS, SOURCE OF KNOWLEDGE, and NEED/REQUIREMENT. Such information is easily available through our visualization for observation and for further analysis and explanation [381].

5.6.3 VSM Visualization

The experts have used the prototype VSM visualization to investigate the user task T10. We must note that this visualization has been subject to the same issues as our original prototype for annotation data, namely, occlusion and infeasibility of color coding. The more important outcome, though, is the lack of evidence supporting our analysts' assumption about the direct correspondence between lexical contents and annotated categories, at least for the currently used data. Figure 5.15 demonstrates an example of both PCA and t-SNE projections. Here, the terms display is turned off, utterances with NEUTRAL and IRRELEVANT annotations are filtered out, and the category CONCESSION AND CONTRARINESS is highlighted. There are no observable patterns related to the positions of highlighted items, and this also holds for the rest of stance categories. Our experts in linguistics have concluded that this fact may have to do with the stop-word filtering and weighting. In many stance categories, function words are the structure builders that indicate, for instance, CONCESSION AND CONTRARINESS ("while", "whereas", "not", "but") or UNCERTAINTY ("may", "might", "can", "could"), while in other cases more contentful lexical constructions carry the meanings (such as "I don't know", "apparently", "I am convinced", "regrettably").



(a) **(b)**

Figure 5.15: VSM visualizations in ALVA using (a) PCA (the first and second principal components) and (b) t-SNE projections. While the positions of visual items are based on the VSM model, the highlighting with blue color is applied to the stance category CONCESSION AND CONTRARINESS in the corresponding annotations. *Reprinted from [242] © 2017 ACM.*

Additionally, our experts in CL think that lack of patterns can be explained by sparsity of the underlying terms-by-utterance matrix.

Nevertheless, our experts have agreed that this VSM visualization can be useful for exploration with particular stance words and constructions in mind. Also, displaying the terms provides a good overview of the topic structure of the annotation data: as seen in Figure 5.13(b), the most prominent terms in this data subset are related to Brexit.

5.6.4 Overall System Utility and Generalizability

Table 5.5 summarizes the comparison of ALVA with several classes of the existing approaches discussed in Section 5.2. Based on typical representatives of such classes, we have estimated their potential applicability to the tasks related to text annotation, active learning, and analysis (visual or computational) of stance annotation data, which reflect the general challenges of stance analysis discussed in Section 2.6 and Section 5.1. ALVA's support for user tasks T1–T10 has been illustrated in Sections 5.4, 5.5, and 5.6. We can argue that none of the previous approaches support the complete range of tasks. Therefore, the utility of ALVA as a single integrated solution becomes apparent. Our research project members have continuously used it for several years (see Figure 5.12 for the overview of the active learning stage), thus providing additional evidence of its utility.

Table 5.5: Comparison of ALVA with previous annotation and visualization approaches

Task/feature vs **Approach**	Ann. tools[a]	AL tools[b]	Vis. tools for ML[c]	Stance vis. tools[d]	**ALVA**
Text annotation	●	●	●		●
• Utterance-level	○	○	○		●
• Multiple categories	○	○	○		●
Classifier training support		●	●		●
• Active learning		●	○		●
Stance visualization				●	●
Analysis of annotated data	○	○	○	○	●
• T1	○	●	●	●	●
• T2	○	●	●	●	●
• T3	○	○	○	○	●
• T4		○		○	●
• T5	○	○	○	○	●
• T6	○	○			●
• T7	○	○			●
• T8	○	○			●
• T9		●	○		●
• T10			○		●

Note: The contents of this table represent potential applicability/support for some general tasks and their more specific subtasks by typical tools from several classes in comparison to ALVA: ● denotes full support, ○ denotes partial support. *Reprinted from [242] © 2017 ACM.*

[a] Regular annotation tools discussed in Section 5.2.1 such as BRAT [402] or Marky [323].

[b] Active learning tools discussed in Section 5.2.1 such as ALTO [328] or TextPRO-AL [276].

[c] Visualization tools for classifier training support discussed in Section 5.2.2 such as the approaches by Heimerl et al. [172] or Makki et al. [278].

[d] Stance visualization tools discussed in Section 5.2.3 such as the approaches by Mohammad et al. [296] or El-Assady et al. [115].

Even though our work so far has focused on stance classification in social media texts in English (also, in the particular genre of political texts), our approach in general is applicable to other text annotation and classification problems (e.g., for sentiment analysis). It can also be applied, for instance, to train a stance classifier for text in a different genre or language. The CatCombos representation is also applicable to other data types with multiple nominal/ordinal attributes or labels.

5.7 Summary

In this chapter, we have introduced our approach for visual support of text annotation and classification in the context of stance analysis in written language. Our system, called ALVA, was developed to support annotation of text utterances with multiple stance categories, machine learning classifier training with the active learning approach, and exploratory visual analysis of the annotated data.

The contributions of this chapter include the analysis of user tasks related to visual analysis of text annotation data and text annotation process, as requested by our experts in linguistics and computational linguistics, and the description of our VA approach. We have also introduced a novel visual representation, called CatCombos, that was designed to represent the individual annotation records and stance category combinations.

While our work is currently focused on visual stance analysis, in this chapter we have discussed how ALVA could be applied to more general tasks related to annotation and classification of data. Such a generalization is a part of future work on ALVA alongside the following plans:

- improvements for the CatCombos representation, such as edge bundling and more efficient layout algorithms to reduce the amount of white space and represent grouping/clustering of blocks with similar sets of categories;
- trend analyses for historical classifier performance data to predict how much more annotated data is required to achieve a certain performance level;
- implementation of additional token-level annotation interface for ALVA (similar to the BRAT tool [402]) and integration with the pre-annotation tool PAL [388];
- support for interactive labeling involving visual representations, as suggested in the recent VIAL approach by Bernard et al. [32];
- improvements and additions for the VSM visualization, such as clustering of terms and involvement of topic models; and
- further analysis of relationships between distributional semantic models [435] and stance categories.

ALVA was used by the members of the StaViCTA project for several years with three main results. First, a subset of the data annotated by several annotators was collected and extracted into a separate data set to serve as a gold standard for future stance classification models. This data set was named *Brexit Blog Corpus (BBC)* and described in detail by Simaki et al. [383]. This resource provides an opportunity for further theoretical research on stance for linguists (as described in the subsequent article by Simaki et al. [382]), which was problematic at the beginning of the project due to the lack of resources explicitly labeled for multiple fine-grained stance categories. Second, the annotated data collected with ALVA during regular and active learning rounds was used to train the SVM-based stance classifier used by subsequent visualization applications. Finally, the insights obtained while conducting visual analyses with ALVA had implications for further research in the project with regard to linguistics [381] and computational linguistics [393]. Our experts reached the conclusion that usage of local cue words could improve the performance of the utterance-based classifier, and then they implemented a logistic regression-based classifier [386] using additional token-level annotations that were carried out outside of ALVA.

Both the SVM and LR-based stance classifiers developed using the data collected with ALVA were used in multiple visualization applications, which are described in the next chapters.

Chapter 6

Visual Analysis of Sentiment and Stance in Social Media Texts with StanceVis Prime

Contents

6.1 Background and Related Work 143
6.1.1 Sentiment and Stance Analysis 145
6.1.2 Sentiment and Stance Visualization 145
6.1.3 Visualization of Time-Varying Data 146
6.2 Requirements Analysis 147
6.3 Architecture 149
6.4 Visualization Methodology 150
6.4.1 General Considerations 150
6.4.2 Representation of Data Series Similarity 153
6.4.3 Representation of Data Series Values 155
6.4.4 Representation of Document Lists 157
6.5 Case Studies 159
6.5.1 Case Study A: Brexit and European Politics in Twitter Data 159
6.5.2 Case Study B: European Politics in Twitter Data vs Reddit Comments 162
6.6 Discussion 165
6.6.1 Preliminary User Feedback 165
6.6.2 Scalability 166
6.6.3 Overall Utility and Generalizability 167
6.7 Summary 168

The recent years have demonstrated how massively available digital communication channels, such as social media, affect the world politics and shape the agenda in multiple spheres of life. The understanding of phenomena occurring in the corresponding data is therefore interesting and important for decision makers, researchers, and the general public. Some of the most interesting aspects of human communication to analyze in such data are related to various expressions of subjectivity in social media document texts, such as sentiments, opinions, and emotions [295, 317]. The analysis of stance-taking in texts [121, 381, 393] can provide even further insights about the subjective position of the speaker, for instance, agreement or disagreement with a certain topic [296, 297], or expression of certainty and prediction [383].

However, the manual analysis of texts and the manual examination of raw output of computational text analyses do not scale up to the amount of data produced by social media, which can range from hundreds to millions of messages per day, depending on the topic or target of interest. Besides the traditional close reading task, support for distant reading [197] is required to make sense of such data. Information visualization and visual analytics approaches have been therefore applied successfully to address this challenge for social media data [74]. In particular, text visualization [236] and visual text analytics [264] methods can support various tasks related to the analysis of individual text documents and large document collections such as summarization of main topics or identification of events in discourse [105], using the techniques developed for time-oriented data visualization [4], when necessary.

Visualization of sentiments and emotions detected in textual data has also become an important topic of interest in the visualization community [240]. Multiple existing sentiment visualization techniques address the tasks of visualizing polarity [59] and emotions [492] detected in temporal text data. However, the related task of stance visualization has not enjoyed the same level of support by the existing approaches so far. One of the main challenges of stance visualization compared to sentiment visualization is related to the need to accommodate the visual design to a specific definition of stance and a data format produced by a specific computational method, which are selected by the users (for instance, domain experts in computational linguistics). Several existing techniques support stance visualization for temporal text data to a certain extent; however, they either make use of a very limited set of stance categories/aspects [104, 115, 243], or interpret such categories as mutually exclusive [286]. Visualization of multiple non-exclusive stance categories at the same time (e.g., produced as an output of a multi-label rather than multi-class classification task) for social media texts is, thus, still an open challenge.

We present our work on a visual analytics platform, called StanceVis Prime, which is designed to support visual analysis of sentiment and stance in temporal text data from social media in this chapter[1] (based on our previous poster abstract [241]). Our approach was designed in collaboration with domain experts in linguistics as part of a larger research project on stance analysis. It follows the definition of stance categories defined by the linguists and uses a custom classifier for 12 stance categories developed by the collaborating experts in computational linguistics / natural language processing. Compared to the existing works related to stance visualization, our approach aims to support the following (see Figure 6.1):

- consumption of data from several social media sources;
- classification of both sentiment polarity and multiple non-exclusive stance categories at utterance/sentence level;
- visual analysis of data series for various sentiment and stance categories at multiple levels of granularity, including both values and similarity between the series; and
- support for distant and close reading of corresponding text document sets augmented with multi-label classification results, including export of document lists processed and annotated by the users.

The rest of this chapter is organized as follows. In the next section, we briefly introduce the necessary background information on sentiment and stance analysis, and then we discuss the related work relevant to our text and time data visualization approaches. Section 6.2 presents the analysis of the workflow and user tasks guiding our design. We briefly describe the overall architecture of our implementation in Section 6.3 and then discuss our design decisions for the visualization components in Section 6.4. We demonstrate our approach with several case studies in Section 6.5 and discuss preliminary user feedback as well as some other aspects of our work in Section 6.6. Finally, we conclude this chapter and outline the directions for future work in Section 6.7.

6.1 Background and Related Work

While the background information and the related work on analysis and visualization of sentiment and stance are discussed in detail in Chapters 2 and 3, respectively, we start this section with a brief introduction of the classification approaches taken in our work. Additionally, we compare StanceVis Prime to the existing stance visualization techniques to highlight its novel aspects. We also outline existing techniques for visualization of temporal text data and even more general time-varying data visualization techniques, as those are relevant to some of the parts of our visual analysis workflow.

[1]By the time of submission of this dissertation in February 2019, the materials of this chapter had been used to prepare a full paper manuscript.

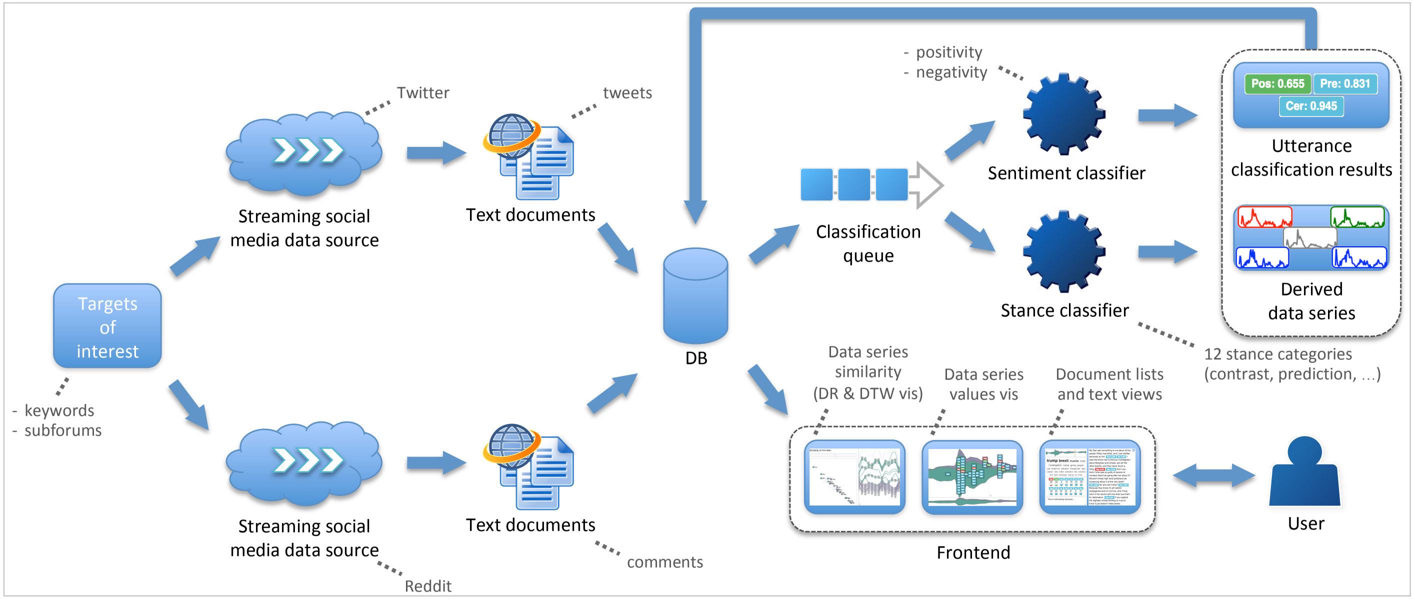


Figure 6.1: Overview of our approach: (1) the data from several streaming text sources is tracked according to the list of targets of interest and saved into a database; (2) the retrieved documents are processed with sentiment and stance classifiers at utterance/sentence level; (3) the classification result counts are used to produce multiple data series; and (4) the user is provided with means for visual analysis of text and time data via interactive visual representations.

6.1.1 Sentiment and Stance Analysis

The existing surveys [295, 317] describe a multitude of existing approaches for sentiment classification at various levels of granularity, but for our purposes, detection of POSITIVE and NEGATIVE sentiment at the level of utterances/sentences in written text data is sufficient. We have used an existing rule-based classifier called VADER [191] designed for short social media texts in our work. VADER reports normalized scores for POSITIVITY, NEUTRALITY, and NEGATIVITY for each input utterance, so that the sum is equal to 1.0. We use the information about utterances with POSITIVITY and/or NEGATIVITY scores above a certain threshold (currently set to 0.3), and consider the rest as NEUTRAL with regard to sentiment.

In contrast to sentiment analysis, the existing works on automatic analysis of stance are not so numerous. Stance-taking has been studied in linguistics [37, 121, 149] with regard to the subjective position of the speaker, which might not necessarily imply POSITIVE/NEGATIVE polarity or some other emotions. The existing computational stance classification approaches usually focus on AGREEMENT OR DISAGREEMENT on a certain topic [296, 297], and only a few works take a wider view of stance aspects/categories into account, such as NECESSITY and VOLITION [383]. In this work, we follow the approach to stance analysis taken in our interdisciplinary project, where researchers in linguistics defined a list of stance categories of interest (presented below in Table 6.1) and experts in computational linguistics implemented a custom stance classifier [390, 393]. Classification is carried out at utterance level in multi-label fashion, i.e., one utterance can be labeled with multiple stance categories simultaneously. We use the information about detected categories for further analysis and visualization.

6.1.2 Sentiment and Stance Visualization

Sentiment and stance visualization problems can be treated as part of a more general text visualization field [236], as discussed in Chapter 3. The existing techniques addressing its various tasks and aspects are covered by several surveys, including the works on topic- and time-oriented visual text analytics by Dou and Liu [105], techniques supporting close and distant reading by Jänicke et al. [197], social media visual analytics by Chen et al. [74], and the recent work on visual text mining by Liu et al. [264]. More specifically, the existing sentiment visualization techniques are discussed in our survey [240] (see Section 3.2). The comprehensive list provided there includes multiple approaches that support sentiment visualization in temporal text data: for example, Whisper by Cao et al. [59] visualizes polarity of tweets as part of a real-time Twitter monitoring task, and PEARL by Zhao et al. [492] represents the emotions detected in tweets by an individual user over time with a stacked graph.

Several existing approaches are also relevant to the stance visualization task: for instance, Lingoscope by Diakopoulos et al. [104] visualizes the language use of *acceptors* and *sceptics* of a certain topic in blogs. ConToVi by El-Assady et al. [115] supports visualization of debate transcripts using categories beyond sentiment, such as CERTAINTY, ELOQUENCE, and POLITENESS. uVSAT [243] (see Chapter 4) focuses on data series based on markers of sentiment, emotions, and stance categories such as CERTAINTY and UNCERTAINTY in blogs and forums. All of these works make use only of a limited number of stance-related categories. Our own work on ALVA [242] discussed in Chapter 5 overcomes this limitation, but the main aim of ALVA lies in supporting data annotation and classifier training stages rather than analyses of data sets collected from social media over time. Another approach relevant to our work is StanceXplore by Martins et al. [286] (discussed below in Section 7.1), which supports visualization of multiple stance categories detected in social media data, however, (1) it treats such categories in a mutually exclusive way for visualization purposes, (2) it does not support sentiment analysis and visualization, and (3) it is limited to data from a single data source (Twitter). In contrast to the existing works, our contribution discussed in this chapter is designed specifically for visual analysis of multiple non-exclusive stance categories as well as sentiment categories in the data from several social media sources.

6.1.3 Visualization of Time-Varying Data

Time is one important aspect of our application domain that must also be considered for the design of efficient visual analysis tools [4]. Data is generated and distributed through social media constantly and in a fast pace, in reaction to events, news articles, or other external trends [273]. This means that, besides the challenge of extracting sentiment and stance from large amounts of unstructured data, we must also consider how these sentiments and stances change through time, and how these changes are related to each other and to other social events.

One example of a classic visualization abstraction for time-varying textual data is ThemeRiver [167]. Themes detected in a collection of time-varying textual documents are depicted as colored "currents", which together form a "river" that flows from left to right (following a timeline). The width of each current reflects the changes in the theme's strength as time passes. This abstraction (i.e., a stacked area graph [54]) has been used, adapted, and re-purposed in many different forms throughout the years for both textual and non-textual data. Recently, for example, MultiStream [92] adapted ThemeRiver to support the interactive exploration of hierarchies of multiple time series using a non-linear time axis. RankExplorer [372] uses a stacked graph to support exploration of rank changes over time in sets of time series. In a recent related work, Lu et al. [273] use a similar visualization to allow users to explore social media data and link it to other secondary data sources, helping with the identification of external events that influence the social streams. We also use a stacked graph abstraction for one of our views and augment it with additional cue labels to represent sentiment and stance classification results, as discussed in Section 6.4.3.

The analysis of multiple data streams may focus on the task of finding similar time series, i.e., streams that show similarities in how they develop through time. In such cases, similarity computations themselves generate a stream of data that must be analyzed, and the main patterns of interest are how similarities between pairs or groups of time series change and how their relationships evolve through time. Storyline techniques, such as the ones discussed by Tanahashi and Ma [410], Liu et al. [265], and Silvia et al. [380] visualize each entity in the data as a timeline that converges (and diverges) with other timelines during periods of more (or less) interaction. Storygraph by Shrestha et al. [376] also represents the actors' dynamics as timelines in a horizontal time axis, but it focuses on actual spatial distances and movement instead of abstract similarities.

Dimensionality reduction (DR) is an effective technique for the visualization of similarities between groups of entities, and it has also been applied in the context of time-varying data sets. In the more usual scenario, where time-varying multidimensional data is projected into 2D and visualized with interactive scatterplots, possible approaches are to project all time steps at once, then visualize time with trajectories (as in the work by Bernard et al. [33]); or to generate one such 2D projection per time step, then present them in a sequence, as done by Alencar et al. [6] and Rauber et al. [334]. Another solution is to project time steps into 1D, and use one axis (usually the horizontal one) for time. With temporal multidimensional scaling (TMDS), Jäckle et al. [194] achieve this by arranging (but not explicitly connecting) sequences of 1D MDS projections in a horizontal axis. Crnovrsanin et al. [91] propose a similar approach, this time with line segments connecting the time steps of the same series, but, as with Storygraph [376], this technique focuses on spatial movement of actors.

Two of our views discussed in Section 6.4.2 are related to these works, with further customizations both in the underlying algorithm and the presentation. In contrast to the previous work, we use an efficient implementation [285] of dynamic time warping (DTW) [34,122] in order to compute meaningful distances between the data series, which are then used as input for a DR technique that generates 1D and 2D projections at each time step.

6.2 Requirements Analysis

The work described in this chapter was carried out as the last stage of an interdisciplinary research project dedicated to stance analysis of written text data. Besides the researchers in visualization, project members included researchers in linguistics and computational linguistics. These domain experts provided the definition of stance categories used in our work and developed a multi-label stance classifier for 12 categories listed in Table 6.1. Multiple discussions with our project members and our earlier experiences in the project have laid the

Table 6.1: Data categories used in our work

Type	Categories	Description
General	*Document Count*	Number of documents detected in the source data for a target of interest
Sentiment	POSITIVITY, NEGATIVITY	Number of occurrences of sentiment in utterances detected with VADER [191]
Stance	AGREEMENT, CERTAINTY, CONCESSION AND CONTRARINESS, CONTRAST, DISAGREEMENT, HYPOTHETICALS, NEED AND REQUIREMENT, PREDICTION, RUDENESS, SOURCE OF KNOWLEDGE, TACT, UNCERTAINTY	Number of occurrences of stance in utterances detected with a custom stance classifier [390,393]

foundation for the overall design of our approach, which is aimed to support exploratory analysis [433] of sentiment and stance in temporal text data.

Based on interviews with data analysis practitioners, Alspaugh et al. [12] describe some of the core exploratory activities as (1) searching for new interesting phenomena, (2) comparing the data to the existing understanding, and (3) generating new analysis questions or hypotheses. These activities in general fit the overall requirements provided by our domain experts, who wanted to use an interactive visual analytics tool to access and explore large amounts of data retrieved from social media. The overview of the workflow expected from our implementation is illustrated in Figure 6.2: users would (1) start from the aggregated temporal data, (2) identify and select interesting time ranges, (3) retrieve and study the corresponding sets of documents, and then (4) focus on individual documents. Interested users should also be able to (5) export filtered and annotated sets of documents for further offline investigation with a focus on close reading [197] and for presentation/dissemination purposes [325]. This workflow allows the users interested mostly in textual data reach it in a rather straightforward way ("simple is good" [348]). At the same time, other users who are interested in analyzing trends in social media could focus mostly on temporal data exploration using coordinated multiple views [343].

The concrete list of user tasks corresponding to this workflow is as follows:

T1. Investigate temporal trends in social media data from several data sources/ domains with regard to several targets of interest.

T2. Investigate temporal trends with regard to sentiment and stance data series.

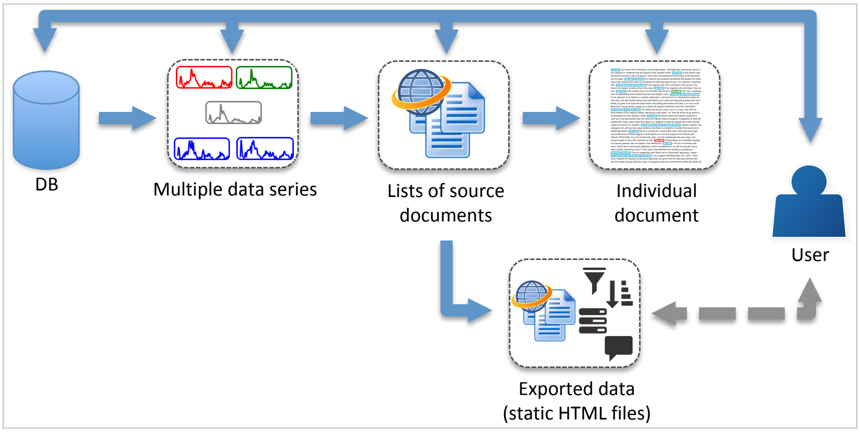


Figure 6.2: The visual analysis workflow for StanceVis Prime according to the requirements discussed with the domain experts. The user should be able to conduct interactive analysis of temporal and textual data using the system and export certain textual data for further offline investigation outside of the scope of the system (denoted with a gray dashed edge).

T3. Investigate similarities of multiple data series over time.

T4. Retrieve underlying documents for specified time ranges for sets of targets/domains.

T5. Summarize document sets with regard to text contents and sentiment & stance.

T6. Engage in close reading of classified text documents.

T7. Export document lists for further offline investigation.

6.3 Architecture

The backend of StanceVis Prime is designed around a data collection service implemented in Python which consumes text data streams from several data sources (see Figure 6.1). Our implementation currently supports Twitter and Reddit, however, it could be easily extended to other data sources in the future. There are multiple targets of interest to be tracked, each defined by a list of key terms and phrases. In addition, the Reddit data stream is configured to include only comments from subforums relevant to our targets of interest. The current list of tracked targets focuses mostly on several key political actors, movements, and events, as these targets are interesting for our domain experts, for instance, European Politics and Brexit[2]. The retrieved text documents are saved into MongoDB [298] and put on the queue for classification.

Our implementation uses the VADER sentiment classifier [191] and a custom logistic regression-based [183] stance classifier [390, 393] developed with scikit-learn [321] for 12 stance categories (see Table 6.1). The classification occurs at the level of individual utterances/sentences, and the information about each detected sentiment and stance category is saved into the database alongside the documents. The stance classifier also reports classification decision confidence, and VADER reports normalized valence values for POSITIVITY and NEGATIVITY, which are also saved into the database. The utterance classification results for each combination of target, domain, and category (e.g., Brexit/*Reddit*/PREDICTION) are used to create the corresponding data series at the granularity of one second and several levels of aggregation (minute, hour, day). The document count for each target/domain is also saved as data series to allow the users investigate all of the retrieved text documents, even including the ones with no sentiment or stance detected.

Our visualization frontend[3] comprises a web-based application served with Flask, and it is implemented in JavaScript with D3 [94] and Rickshaw [338]. The details about its components are discussed in the next section.

6.4 Visualization Methodology

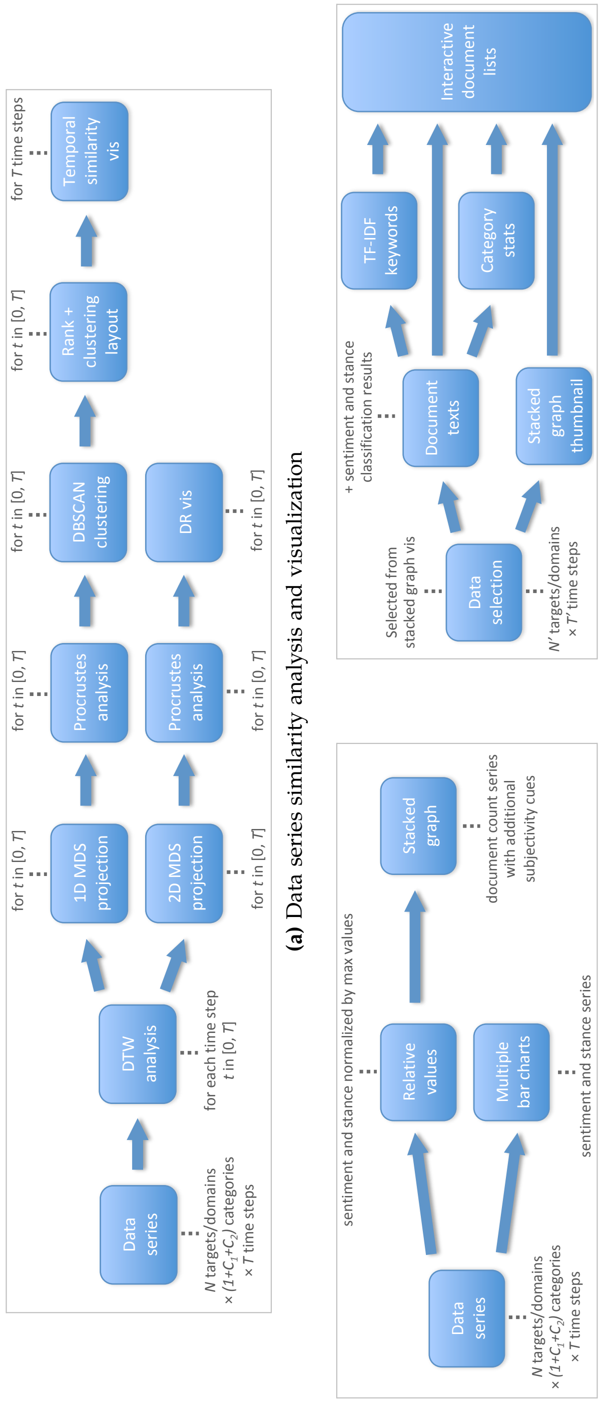
In this section, we are going to discuss the design of visual representations used in StanceVis Prime for various parts of the workflow depicted in Figure 6.2. We start with some general considerations affecting the overall design, and then discuss the specific data processing and representation concerns in detail (see Figure 6.3).

6.4.1 General Considerations

Our visualization approach was required to support multiple data series and text document representations associated with classification results for sentiment and stance, as discussed in Section 6.3. This presented us with challenges related to color coding consistency across multiple coordinated views [343]. More specifically, the user could desire to investigate the data for N combinations of targets of interests and data domains, for instance, Brexit/*Reddit*, Brexit/*Twitter*, Vaccination/*Reddit*, etc. Each of such combinations is associated with one data series based on retrieved document count, two series for sentiment categories, and twelve series for stance categories (see Table 6.1), which means that the user would typically work with dozens of data series simultaneously, and it would not

[2]Here and below in this chapter, sans serif font is used to indicate targets of interest.

[3]A demo video for StanceVis Prime is available at `http://bit.ly/stance-vis-prime` (last accessed in February 2019).



(a) Data series similarity analysis and visualization

(b) Data series values visualization

(c) Document lists and text views

Figure 6.3: The pipelines for data processing and visualization in StanceVis Prime.

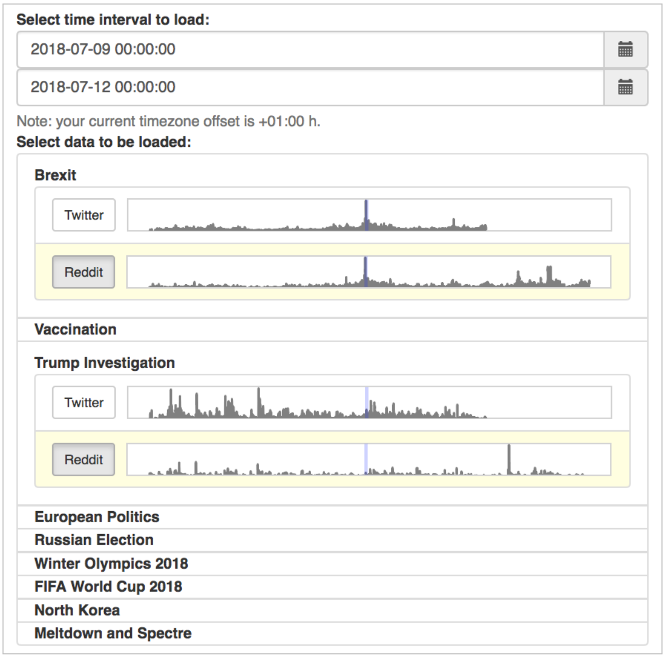


Figure 6.4: The data loading dialog in StanceVis Prime. The user is presented with date & time input fields and a collapsible list of tracked target/domain combinations with sparkline plot previews for the corresponding document count data series. The currently selected time range is highlighted in blue in the plots. The target/domain combinations selected for loading are highlighted in yellow.

be possible to encode each series with a unique hue. The sentiment categories of POSITIVITY and NEGATIVITY are usually encoded by green (or blue) and red hues in the existing sentiment visualization techniques [240], and our design had to incorporate this fact, too.

The current color coding approach in our implementation handles targets/domains and data categories in an orthogonal fashion. First of all, each target is assigned with a unique hue using a qualitative scheme from ColorBrewer [85], and the concrete target/domain combinations (see Figure 6.6(a)) are then encoded with darker or brighter variations of that color. These colors are used in the visual representations of data series to facilitate the comparison tasks T1–T3 between various targets/domains (see Figures 6.6 and 6.7). As for the data categories, we have assigned green and red colors to the respective sentiment categories, blue to all stance categories, and gray to the document count series. These encodings are used in the colored labels in the targets table (Figure 6.6(a)), subjectivity cues in the document count representation (Figure 6.7(d)), sentiment and stance chart headers (Figure 6.7(g+h)), and document views (Figure 6.8(d+f+g)).

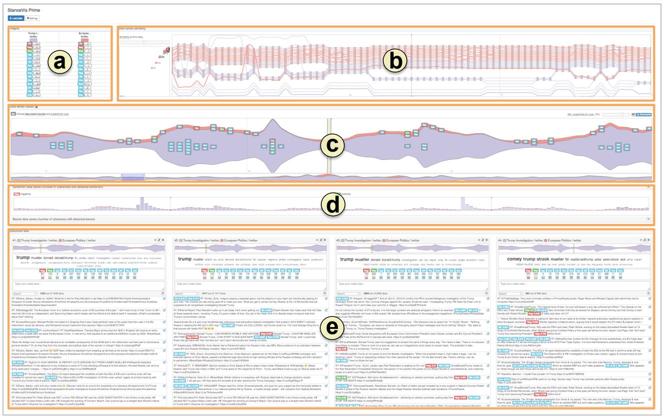


Figure 6.5: The overview of the user interface in StanceVis Prime: (a) the loaded data series table; (b) data series similarity views; (c) the document count graph; (d) bar charts with data series values; and (e) document list views.

A typical use case scenario for StanceVis Prime starts by logging in and accessing the data loading dialog displayed in Figure 6.4, which contains the list of available targets of interest and their respective data domains (currently, Twitter and Reddit). For each target/domain combination, the dialog also provides a sparkline [432] plot preview of the document count data series. After selecting the list of interesting targets and domains and specifying the overall time interval, the user is presented with a visualization interface displayed in Figure 6.5, with the panels (a–d) visible initially. The table displayed in Figure 6.5(a) and Figure 6.6(a+b) acts as a legend for colors associated with each target/domain and enumerates the corresponding data series with their abbreviated titles and sparkline plots. The category labels are also used in other parts of the interface to avoid the introduction of complicated glyphs for the corresponding C data series (currently, 1 document count series + 2 sentiment series + 12 stance series). After loading the data, the user can proceed to data series exploration using the similarity views displayed in Figure 6.5(b) and the detailed value charts displayed in Figure 6.5(c+d). The user can then perform multiple document list queries and explore their results, as displayed in Figure 6.5(e).

6.4.2 Representation of Data Series Similarity

As described above, the data initially loaded by the user in StanceVis Prime consists of C data series (currently, 1+2+12) for each of N target/domain combinations. While the investigation of separate groups of such series is feasible, the number of data series becomes overwhelming for task T3 if all of the values are visualized at once. This was our motivation for introducing separate views with a focus on *similarity* rather than *values*.

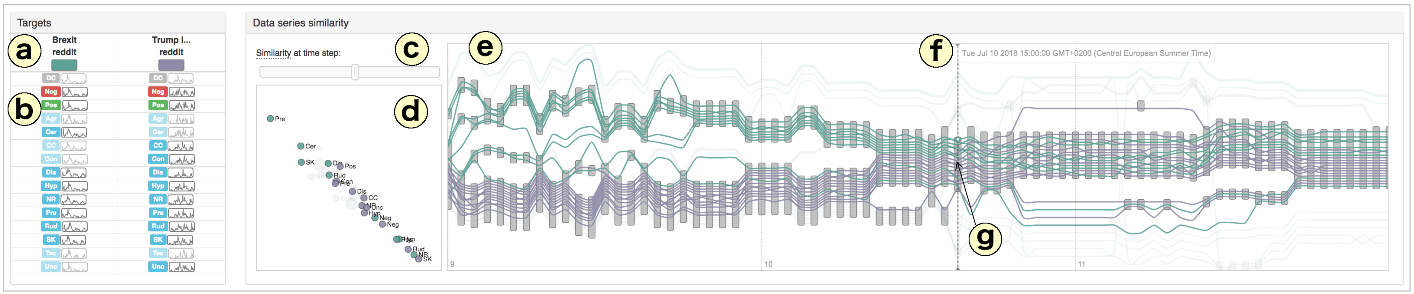


Figure 6.6: Visualization of data series similarity in StanceVis Prime: (a) a table with all loaded target / data domain combinations, which is also used for filtering and brushing purposes; (b) a list of all loaded data series corresponding to targets/domains with sparkline plot previews; (c) a temporal slider for selecting a time step to be displayed in the 2D projection view below; (d) a 2D projection view with representation of data series at the current time step; (e) a temporal similarity view with line chart representation of data series' projections over time; (f) an indicator of the time step selected with the slider on the left; and (g) a representation of a data series cluster in the 1D projection at the corresponding time step. Here, the user has hovered over the cluster (g), which caused all the data series not included in this cluster to be faded out.

As the corresponding pipeline in Figure 6.3(a) displays, the processing starts by computing dynamic time warping (DTW) [34, 122] distances between the pairs of data series over the available time range. We use a custom DTW implementation with a sliding window discussed by Martins and Kerren [285], and the size of the window is currently set to 30% of the data series length. It is then possible to analyze the distances between the series at each time step (temporal slice). Currently, we employ the multidimensional scaling (MDS) [39] method to compute 1D and 2D projections. We have chosen MDS as a standard and reliable technique for this task, but it would be possible to extend our approach with other techniques. Since MDS is invariant to transformations such as rotation, our next step is to align the projections from consecutive time steps. We have applied Procrustes analysis [232] to align subsequent temporal slices of both 1D and 2D projections in a consistent way.

The resulting 2D projection can be visualized with a standard scatterplot representation one temporal step at a time, as displayed in Figure 6.6(d), controlled by a slider displayed in Figure 6.6(c). However, an attempt to directly visualize the 1D projection data with a representation such as a line chart results in a cluttered view that does not really facilitate the user task. Therefore, we have decided to forgo the idea of using the exact projected values and instead drew inspiration from rank-based [372] and storyline-based [380, 410] representations. For each temporal step, we have computed clustering of the 1D projections using DBSCAN [123]. The algorithm is configured to detect clusters containing at least three items; the items not belonging to any cluster are labeled as outliers. To compute the resulting layout for this temporal step, we (1) order the 1D projection items by ascending value and (2) assign increasing y coordinates while (3) ensuring vertical space proportional to intra-cluster, inter-cluster, and intra-outlier cases. The resulting layout is presented in Figure 6.6(e+g): lines represent individual data series, and dark rectangular blocks represent clusters. By glancing at this representation, the user can identify the major groups of data series and time steps where the behavior of such groups and series changes drastically. This view is coordinated with the others by brushing and linking, so the user can explore the similarity between data series to fulfill task T3 and then switch to investigation of specific data series values, as discussed below.

6.4.3 Representation of Data Series Values

The main steps related to visualization of actual values for all the loaded data series are illustrated in Figure 6.3(b). A stacked graph [54, 167] is used in the central part of the interface (see Figure 6.7(a)) to represent the overall counts of processed documents for each target/domain over time. The user is initially presented with an overview of the complete loaded data set and can then focus on a specific time interval using the range slider depicted in Figure 6.7(b), which might cause the change of data granularity (e.g., from days to hours). The main

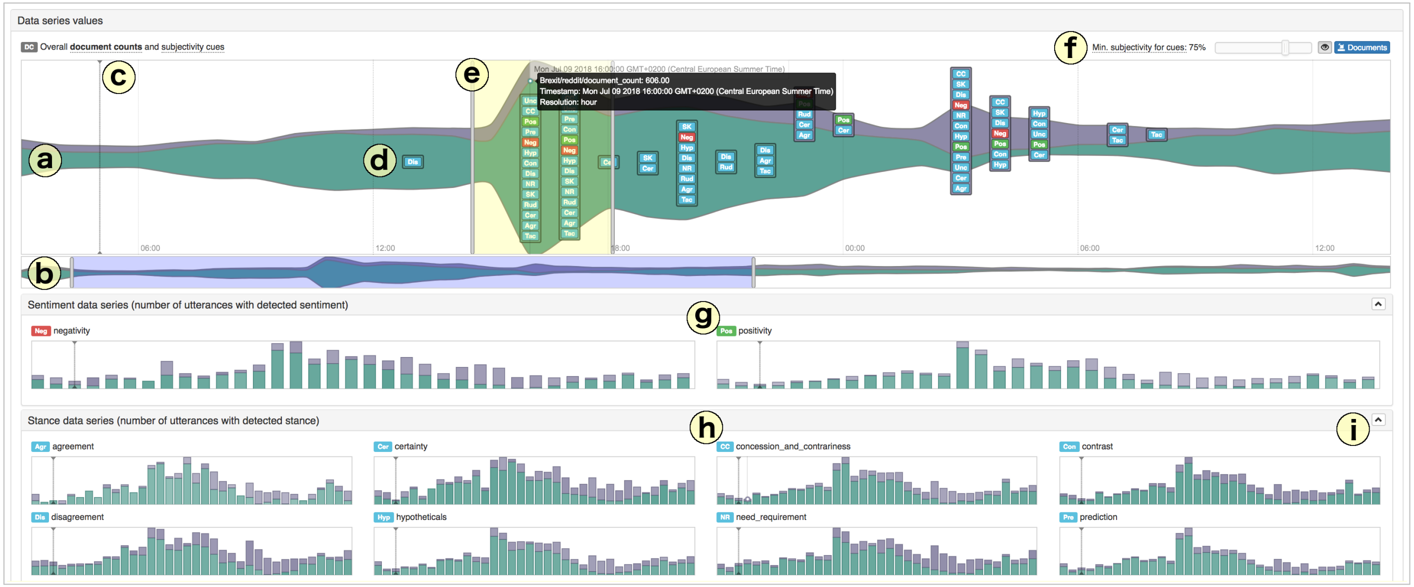


Figure 6.7: Visualization of data series values in StanceVis Prime: (a) a stacked graph representing document counts for target/domain combinations over time; (b) a range slider providing an overview about document counts for the complete loaded data set; (c) an indicator of the time step selected with the slider in Figure 6.6(c); (d) a subjectivity cue label indicating a relatively large value of the corresponding sentiment/stance data series at the same time step; (e) a selection made by the user to retrieve the corresponding set of documents; (f) controls including a slider and a toggle button for the subjectivity cues representation, as well as a button for retrieving the documents for the selection in (e); (g) bar chart representations for sentiment data series; (h) bar chart representations for stance data series; and (i) a button for collapsing a bar chart panel.

graph focuses mostly on the overall document counts rather than subjectivity categories, since it would require up to 2+12 additional plots per target/domain. However, the representation includes visual cues about the temporal points with relatively high amounts of detected subjectivity. For instance, a "*Dis*" label is displayed in Figure 6.7(d) over the graph for Brexit/*Reddit*. It means that at the corresponding time step the value of the Brexit/*Reddit*/DISAGREEMENT data series was relatively high compared to the maximum value in this loaded series. By using the controls depicted in Figure 6.7(f), the user can adjust the minimal threshold for the relative level of subjectivity or hide such subjectivity cues altogether (as they could lead to visual clutter in some cases). This visual representation supports tasks T1 and T2.

To support T2 further, our implementation also provides multiple small bar charts for separate sentiment and stance data series displayed in Figure 6.7(g+h). If the user is not interested in such detailed information, it is possible to collapse the corresponding panels to save screen space (see Figure 6.7(i)).

Finally, the main stacked graph representation also supports task T4: the user can select a time range (cf. Figure 6.7(e)), click the button depicted in Figure 6.7(f), and request the documents corresponding to the selected target/domain combinations for this time range.

6.4.4 Representation of Document Lists

When the user requests documents by performing a selection in the stacked graph depicted in Figure 6.7(e), the document texts with sentiment and stance classification results are retrieved from the database (see Figure 6.3(c)), and a new document list is displayed in the user interface, as displayed in Figure 6.8. Additionally, a static thumbnail of the graph selection is created to serve as a snapshot (cf. Figure 6.8(b)).

One part of task T5 is concerned with summarization of text contents of such document lists. To address this, we have decided to provide the users with a list of key terms. Our implementation computes the counts of uni- and bigrams [279] present in the document texts and then processes them using TF-IDF weighting [354]. Currently, up to 25 top key terms are returned for each document list and represented as an interactive list displayed in Figure 6.8(c). We used a simple list representation (with eventual wrapped lines/rows) instead of a word cloud with spatial layout, as a recent study demonstrated similar effectiveness of these approaches for topic discovery tasks [130].

To represent the summary of category classification results for T5, our implementation includes an interactive list of statistics for sentiment and stance categories displayed in Figure 6.8(d). The total number of utterances with a specific category detected is computed over all document list texts and displayed. Additionally, the average classification confidence reported by the stance classifier

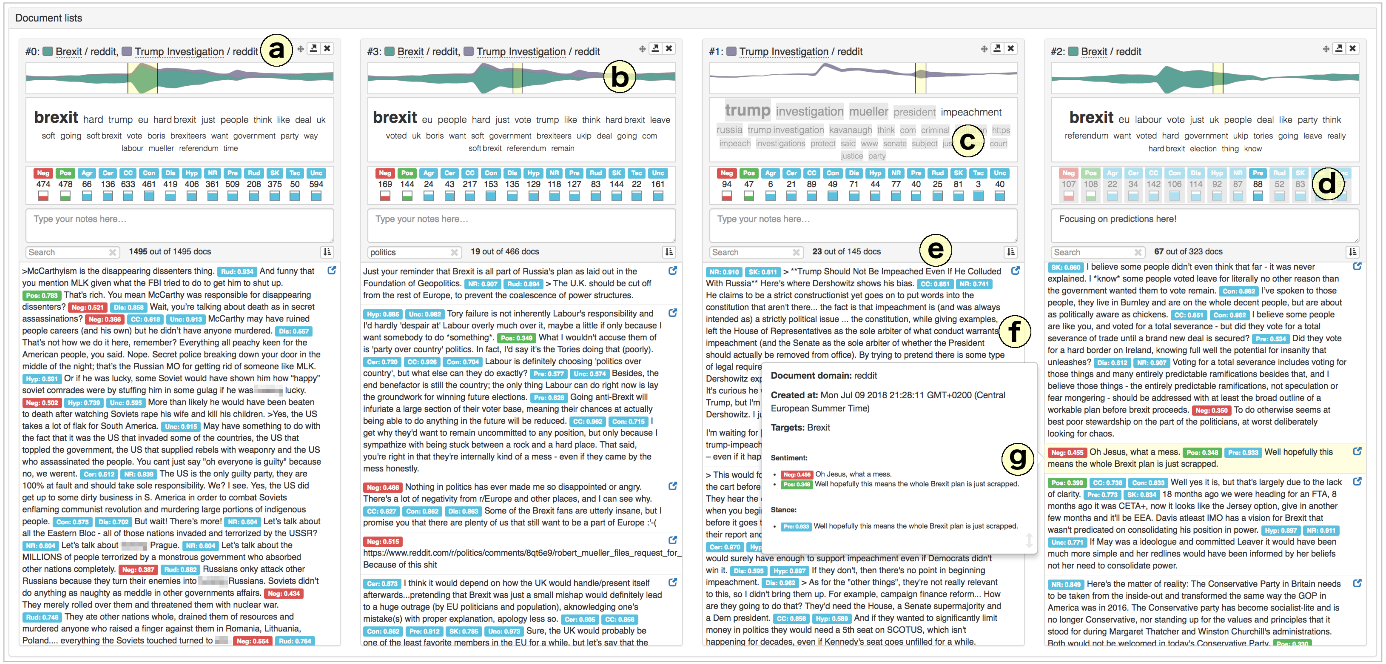


Figure 6.8: Document list representations in StanceVis Prime: (a) a header including the list description, close and export buttons, and a drag'n'drop indicator icon; (b) a static thumbnail copied from the representation displayed in Figure 6.7(a); (c) a list of salient uni- and bigrams discovered in the document texts; (d) statistics for sentiment and stance classification results in the document texts; (e) additional controls including an area for user notes, a text search field, and a sort button; (f) a document view with injected labels with utterance classification results; and (g) a tooltip with details for a selected document. Here, the user has changed the position of one of the document lists with drag'n'drop and used various interactions in (c), (d), (e), and (g).

and the average valence for POSITIVITY/NEGATIVITY reported by the sentiment classifier are visualized with bar charts.

The list of key terms and the statistics view can be used for filtering the document list alongside the text search field displayed in Figure 6.8(e). The user can add notes, sort the documents in multiple ways (e.g., by document timestamp or by number of detected stance occurrences), and close/remove the document list. It is also possible to change the position of document lists by dragging and dropping, which can be useful for comparison tasks.

To support T6, document lists provide access to the actual texts as seen in Figure 6.8(f), including links to the source social media posts. The sentiment and stance classification results are injected directly into the document view as labels with abbreviated category titles and classification confidence results. By hovering over a document, the user is presented with a tooltip displayed in Figure 6.8(g) which includes additional details about the document timestamp, data domain, associated target(s) of interest, and detailed classification results. It is also possible to click a document to make the tooltip persistent.

Finally, the user can use the export button located in the document list header (see Figure 6.8(a)). In this case, a static version of the document list is exported as an HTML file, including its current state with regard to user notes, filtering, and sorting (cf. Figure 6.9(d+g)). Exported lists can be used for further offline investigation, thus supporting the final user task T7.

6.5 Case Studies

In this section, we demonstrate the usage of StanceVis Prime with two case studies. In the first study, the information about public sentiment and stance on two different targets of interest is compared for Twitter data. In the second study, we focus on a single target of interest in two different social media data sources, Twitter and Reddit.

6.5.1 Case Study A: Brexit and European Politics in Twitter Data

In this case study, we have explored the Twitter data for summer 2018 available in StanceVis Prime. We have focused on two targets of interest tracked in our system: Brexit and European Politics. The initial time range of interest was the complete summer; the resulting stacked graph, initially loaded at the granularity of one day, is displayed in Figure 6.9(a).

We can see that the volume of tweets for Brexit (encoded with green color) dominates this data set, and the maximum number of posts is created at a time step highlighted in Figure 6.9(b). This time step corresponds to July 9, 2018, and there are 742,139 tweets retrieved in our system for this data point. According to the subjectivity cues, multiple categories of sentiment and stance reach large

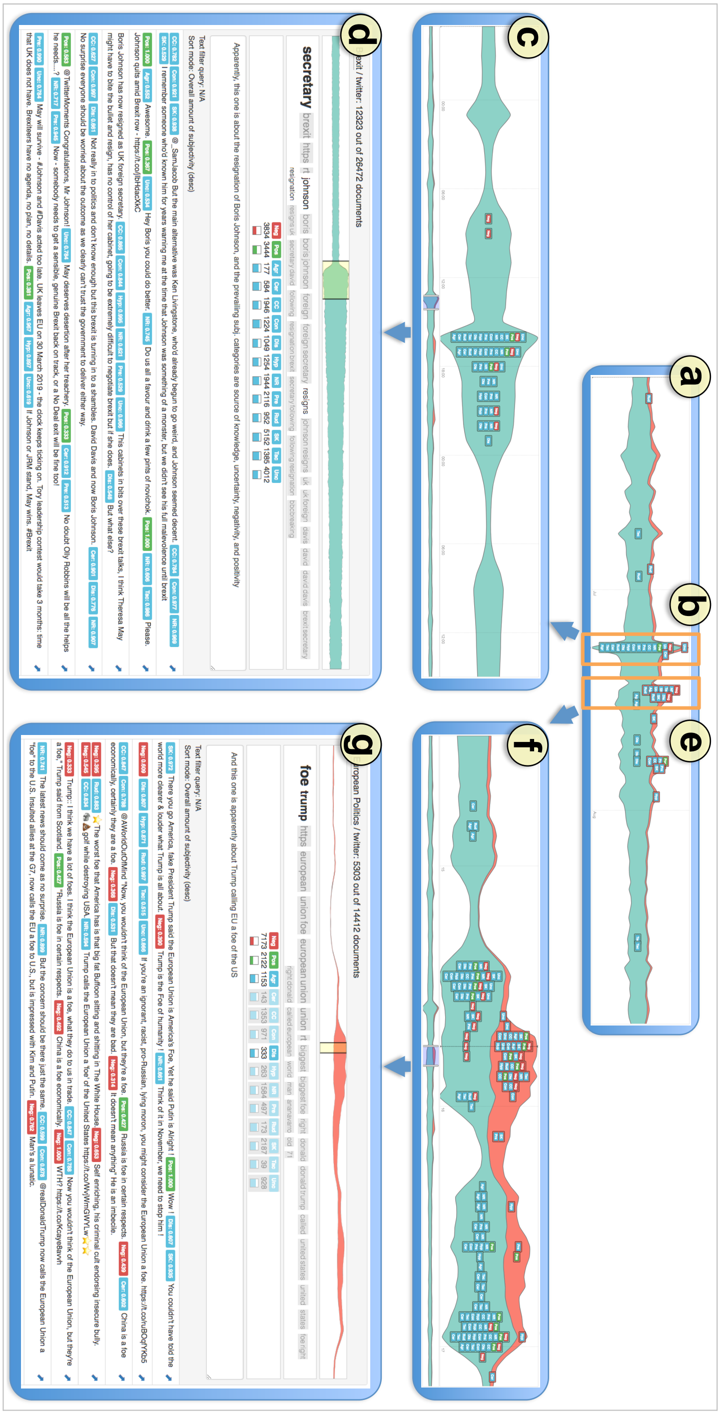


Figure 6.9: The data series and exported document lists in case study A

Table 6.2: Examples of interesting utterances with corresponding categories discovered in the data for case study A

Utterance	**Detected categories**
A. Brexit/*Twitter*, July 9, 2018	
"Chaos ensues" [a]	NEGATIVITY
"If true it's effectively the final nail in the coffin for May" [b]	POSITIVITY, HYPOTHETICALS, UNCERTAINTY
"Could be good for Ireland if it leads to Theresa May having a stronger hand to push through a soft Brexit" [c]	POSITIVITY, HYPOTHETICALS, NEED AND REQUIREMENT, UNCERTAINTY
B. European Politics/*Twitter*, July 15, 2018	
"If you look at the thing they quoted, Trump clearly said that the European Union was an economic foe" [d]	CERTAINTY, HYPOTHETICALS, SOURCE OF KNOWLEDGE
"Man's a lunatic" [e]	NEGATIVITY
"But the concern should be there just the same" [f]	NEED AND REQUIREMENT

[a] Retrieved July 9, 2018 from `https://twitter.com/statuses/1016325066756448256`
[b] Retrieved July 9, 2018 from `https://twitter.com/statuses/1016321947045646336`
[c] Retrieved July 9, 2018 from `https://twitter.com/statuses/1016322079958892544`
[d] Retrieved July 15, 2018 from `https://twitter.com/statuses/1018518507502350339`
[e] Retrieved July 15, 2018 from `https://twitter.com/statuses/1018528605515730945`
[f] Retrieved July 15, 2018 from `https://twitter.com/statuses/1018528605515730945`

values at this day, with NEGATIVITY, UNCERTAINTY, and POSITIVITY being the top ones.

As we focus on a shorter time range corresponding to this day and filter out the data series related to the other target, the representation changes to hourly values, as displayed in Figure 6.9(c). We can now see that a large number of tweets were produced starting around 16:00 (GMT+2). Exploration reveals that there are 79,711 tweets in total for this hour, containing 14,706 utterances with UNCERTAINTY, 12,992 with NEGATIVITY, and 11,279 with POSITIVITY.

After focusing on a shorter period (16:00–16:20) and retrieving the documents, we can find out that the main key terms for the corresponding document list were "secretary", "johnson", "resigns", and so on. Apparently, this spike in the discussion of Brexit on Twitter was related to the resignation of the UK Foreign Secretary Boris Johnson[4]. After interacting with the document list, we can export it for further offline exploration, with the result displayed in Figure 6.9(d) and examples listed in Table 6.2 (group A).

[4] `https://www.bbc.com/news/av/uk-politics-44771278/boris-johnson-resigns-as-foreign-secretary` (last accessed in February 2019)

Returning to the complete loaded data set, we can now focus on the data region where European Politics gained prominence (see the red band of the stacked graph in Figure 6.9(e)), which corresponds to July 15, 2018 with 122,298 tweets in our retrieved data. Focusing further and switching to the granularity of one hour, we can see in Figure 6.9(f) that sudden growth of interest for this target started around 18:00 (GMT+2). There are 13,945 tweets corresponding to this hour, with 7,990 utterances classified with NEGATIVITY, 2,177 with POSITIVITY, and 2,070 with the stance of SOURCE OF KNOWLEDGE. After focusing on the period 17:30–18:30 and investigating the document list, we can find out that this case is related to the quote of POTUS Donald Trump naming European Union "a foe" in an interview[5]. As we can see in the exported document list in Figure 6.9(g), a lot of expressions of sentiment were produced in the tweets, followed by several stance categories such as SOURCE OF KNOWLEDGE, NEED AND REQUIREMENT, and CONCESSION AND CONTRARINESS. Examples are provided in Table 6.2 (group B).

By using StanceVis Prime in this case study, we were able to understand the temporal trends in public sentiment and stance on several targets of interest, retrieve the underlying text data for specific time ranges, get an overview of the corresponding texts, and use the exported versions of document lists for close reading offline (cf. user tasks T1–T2 and T4–T7).

6.5.2 Case Study B: European Politics in Twitter Data vs Reddit Comments

In this second case study, we have compared the data on the same target of interest (European Politics) in two different data sources, Twitter and Reddit. We have selected a time range of several days around September 9, 2018, when a recent general election took place in Sweden[6]. After exploring the temporal data and downloading the document lists, we can confirm the expectation that the number of documents from Twitter for the same time range would be much larger than the number of Reddit comments on this subject, as seen in Figure 6.10(a+b) with approximately 5,000 tweets and 50 Reddit comments. Filtering the document lists on key terms related to the Swedish election reduces these numbers even further. The examples of relevant utterances found in tweets and Reddit comments are provided in Table 6.3 (groups A and B, respectively).

A more lively discussion of the Swedish elections on Reddit does not start until the morning of the next day, September 10, as seen in Figure 6.10(c). It is interesting to compare the documents between the data sources, though, as Reddit comments tend to contain much longer statements and arguments rather than short tweets, which might be interesting for further detailed exploration with close reading. Several examples of relevant utterances are listed in Table 6.3 (group C).

[5] `https://www.bbc.com/news/world-us-canada-44837311` (last accessed in February 2019)
[6] `https://www.bbc.com/news/world-europe-45466174` (last accessed in February 2019)

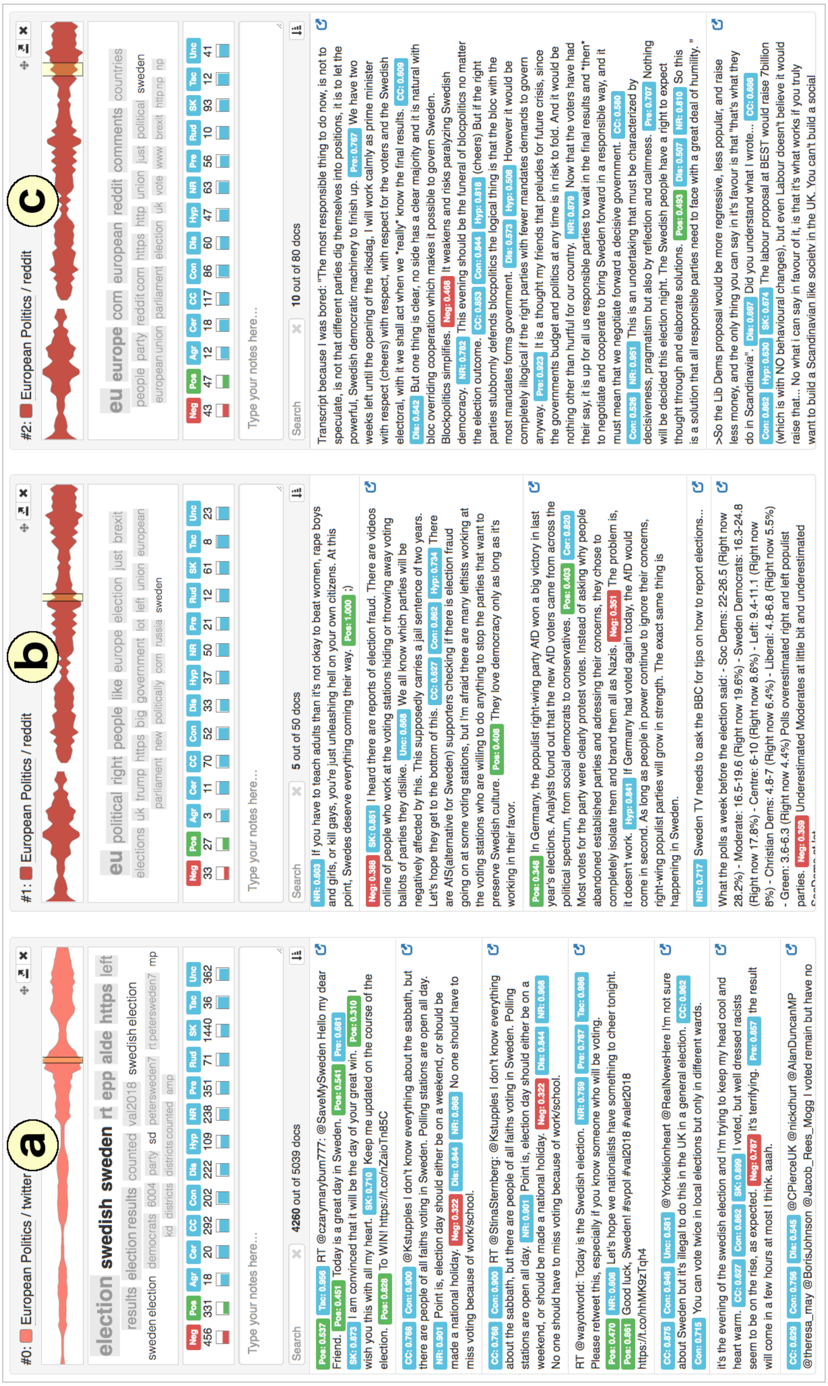


Figure 6.10: The document lists in case study B

Table 6.3: Examples of interesting utterances with corresponding categories discovered in the data for case study B

Utterance	Detected categories
A. European Politics/*Twitter*, September 9, 2018	
"Today is a great day in Sweden" [a]	POSITIVITY
"I guess that will do Sweden" [b]	PREDICTION, UNCERTAINTY
"I understand none of it but hey I might live there some day so I should probably watch their election process" [c]	CONCESSION AND CONTRARINESS, CONTRAST, NEED AND REQUIREMENT, UNCERTAINTY
B. European Politics/*Reddit*, September 9, 2018	
"I heard there are reports of election fraud" [d]	NEGATIVITY, SOURCE OF KNOWLEDGE
"Most votes for the party were clearly protest votes" [e]	POSITIVITY, CERTAINTY
"Sweden TV needs to ask the BBC for tips on how to report elections. . . " [f]	NEED AND REQUIREMENT
C. European Politics/*Reddit*, September 10, 2018	
"However it would be completely illogical if the right parties with fewer mandates demands to govern anyway" [g]	DISAGREEMENT, HYPOTHETICALS
"Sweden is extremely progressive, most parties would be seen as progressive parties in the US except the nationalist Sweden Democrats and the Christian Democrats" [h]	CONCESSION AND CONTRARINESS, CONTRAST
"As the statistics showed, most SD voters come from the Socialdemocrats which have been the most progressive party in Sweden" [i]	SOURCE OF KNOWLEDGE

[a] Retrieved September 9, 2018 from `https://twitter.com/statuses/1038889571000430592`

[b] Retrieved September 9, 2018 from `https://twitter.com/statuses/1038885246299721733`

[c] Retrieved September 9, 2018 from `https://twitter.com/statuses/1038881881566191616`

[d] Retrieved September 9, 2018 from `https://www.reddit.com/r/The_Donald/comments/9ed8px/the_media_is_propaganda_101_how_about_we_change/e5omh5x/`

[e] Retrieved September 9, 2018 from `https://www.reddit.com/r/worldnews/comments/9ec9ga/sweden_votes_amid_nationalist_surge/e5olcou/`

[f] Retrieved September 9, 2018 from `https://www.reddit.com/r/europe/comments/9e38aq/swedish_general_election/e5oha38/`

[g] Retrieved September 10, 2018 from `https://www.reddit.com/r/europe/comments/9e38aq/swedish_general_election/e5pu6bg/`

[h] Retrieved September 10, 2018 from `https://www.reddit.com/r/worldnews/comments/9ejj2s/sweden_liberals_seek_alliance_govt_wont_work_with/e5pvy4r/`

[i] Retrieved September 10, 2018 from `https://www.reddit.com/r/worldnews/comments/9ejj2s/sweden_liberals_seek_alliance_govt_wont_work_with/e5pvy4r/`

In summary, this case study demonstrated support for comparisons between data from several sources in StanceVis Prime, mostly with a focus on retrieval, overview, and close reading of multiple document lists (cf. user tasks T4–T7).

6.6 Discussion

In this section, we report the preliminary user feedback received for StanceVis Prime and reflect upon various aspects of our implementation, including its scalability and generalizability with regard to underlying data sources and analyses.

6.6.1 Preliminary User Feedback

Due to the time and expert user availability constraints, we have not evaluated our approach with a larger user study yet. However, we have gathered some preliminary user feedback in December 2018 during a session with one of our project collaborators, a postdoctoral researcher in computational linguistics. This expert user had contributed to the development of the stance classifier used by our system's backend and has prior knowledge about information visualization and visual analytics, but she had not been involved in the design and development of StanceVis Prime. During the session which lasted several hours, we (1) explained the underlying design choices and data considerations, (2) demonstrated and explained most features of the current implementation, (3) allowed the user to engage in free data exploration with a think-aloud protocol, and finally, (4) conducted a semi-structured interview. For the interview, we used questions from the ICE-T questionnaire by Wall et al. [446] focusing on the value of visualization.

During the free exploration stage, the expert focused on the target of Vaccination rather than political targets, as she had previously worked on this topic in her own research and was still interested in it. The expert spent quite a lot of time investigating sparkline charts in the data loading dialog to find interesting time ranges and settled on the data from mid-August to early September 2018 from both Twitter and Reddit. After loading the data, the user studied the visualization of document counts and subjectivity cues; the cues about the categories of NEGATIVITY and NEED AND REQUIREMENT in the Twitter data for a specific day (August 26, 2018) caught the eye of the expert. The expert then used the range slider to focus on this date, studied the detailed sentiment and stance category charts, and then used the button for loading the corresponding tweets. The resulting document list contained approximately 33,000 entries. The expert used the category bar charts to focus on the documents with NEED AND REQUIREMENT and/or NEGATIVITY and engaged in close reading of the resulting texts. After studying a number of documents, she concluded that the sentiment and stance classification results seemed reasonable and there were a lot of urges and requests both to vaccinate and not to vaccinate in the tweets. The expert also got an insight that NEGATIVITY did not co-occur with NEED AND REQUIREMENT in this data.

The expert also loaded and investigated other time ranges with a focus on stance categories such as TACT and SOURCE OF KNOWLEDGE. Some of her insights included the discoveries that (1) many occurrences of TACT in this social media data are in fact explained by sarcasm; and (2) many references to sources of knowledge include vague statements (*"people are saying that..."*) and other people's opinions rather than facts and authorities (however, several references to organizations such as CDC and WHO were also found).

In general, the expert concluded that StanceVis Prime seemed useful for the tasks of investigating public sentiment and stance. As her expertise is in computational linguistics, she was particularly interested in exploring the classification results in documents corresponding to peaks in the data series. The expert also commended the possibility of exporting the document lists annotated with user notes for further analyses and close reading.

The expert user also had several comments about shortcomings of the current implementation and potential features, mostly related to the document list view and related interactions. First of all, she noted a number of duplicate documents (retweets) in the Twitter data and suggested hiding/collapsing them in this view in order to focus only on unique texts. Second, she noted that category filters in document list views are currently based on the "OR" operator, but it would be useful to also support "AND"-based filters in order to focus on documents with particular combinations of categories (cf. Chapter 5), such as NEGATIVITY and NEED AND REQUIREMENT mentioned above. Taking such selected combinations of categories into account when sorting the document lists was also potentially useful in the expert's opinion. Finally, the delays associated with the document loading stage were also perceived negatively. These comments and suggestions will be taken into account as part of the future work to improve StanceVis Prime.

6.6.2 Scalability

The scalability of the backend components of our approach is subject to storage space availability similar to other systems consuming data from social media. For example, at this point we have collected approximately 168 million documents in our database, and the storage size required for this source data, classification results, and backups grows every day. Besides disabling tracking of certain targets of interest, one possible solution for this issue is to remove old documents from the database while preserving the corresponding document IDs/URLs as well as the aggregated data series.

With regard to the scalability of the visualization and user interface at the frontend, we should note that our chosen strategy of using coordinated multiple views has a drawback related to the overall screen area usage. The visual interface of StanceVis Prime supports window resizing with regard to width, but it is designed with a vertical workflow in mind, from the source data selection and listing to the data series charts and the document lists, which results in vertical scrolling on smaller monitors. One possible alternative solution would be to introduce separate windows or tabs for separate views, similar to document list tabs in uVSAT [243] (see Section 4.5.2). However, it could arguably disrupt the user's mental map and would also make comparison between multiple document lists more difficult. For the time being, we intend our implementation to be used on desktop monitors rather than screens of laptops or handheld devices.

Additionally, the visualization of data series similarity discussed in Section 6.4.2 can be subject to visual clutter depending on the number of displayed data series and the output of MDS and DBSCAN algorithms. This visual representation could be improved in the future using the methods introduced for storyline visualization techniques [380, 410].

Finally, the performance of the visual interface depends on the size of the data subset loaded by the user. For example, if all the documents for some popular target of interest are requested for a range of days, weeks, or months, the web browser will struggle to render dozens or hundreds of thousands of document text representations. Thus, the workflow currently recommended to the users is to focus on moderate time ranges when requesting document lists.

6.6.3 Overall Utility and Generalizability

As discussed in Section 6.2, StanceVis Prime is designed to support user tasks related to various data types and granularities, from individual utterances and documents to data series with aggregated daily values spanning multiple months. Therefore, we expect its components and views to have different value to different users: the preliminary user feedback discussed in Section 6.6.1 confirms our expectation that experts in linguistics and computational linguistics would ultimately be interested in accessing and studying text data when using our tool. The visual representations of temporal data, including the data series similarity view, could be more interesting to the users interested in overall trends in public sentiment and stance, such as brand managers or political scientists, and it would be a part of our future work to collaborate with such users.

We foresee several ways to generalize our approach. First of all, StanceVis Prime was initially designed to support multiple data sources: it is already possible to access and compare the data from two currently supported social media platforms. Our approach can also be generalized to other temporal text data sources and, with some modifications, to static corpora. Second, we are currently using a specific set of sentiment and stance categories which could be extended or replaced as long as text analyses are carried out at utterance/sentence level: for instance, our stance classification pipeline could be extended with another classifier which achieves better classification results for a limited set of categories such as AGREEMENT and DISAGREEMENT [297]. Finally, our research so far has focused only on texts in English, and it would be interesting to apply StanceVis Prime to text data in other languages once the respective sentiment and stance classifiers are available.

6.7 Summary

In this chapter, we have discussed our work on StanceVis Prime, a visual analytics platform for social media texts supporting sentiment and stance analysis. Our approach is based on the definition of stance categories and a classifier developed as part of an interdisciplinary project on stance analysis. This chapter presents the analysis of user tasks for visual analysis of sentiment and stance in social media texts based on the requirements of our project collaborators. We describe our design choices for visual analysis of the textual data and the corresponding data series based on retrieved document counts and utterance classification results. StanceVis Prime allows the users to investigate temporal trends, get an overview about the contents of document sets, study individual document texts, and export processed document sets for further offline exploration. We demonstrate our implementation with case studies focusing on comparison of several targets of interest and several data domains/sources. Preliminary user feedback from one of our project collaborators is also promising.

The backend component of our system has been collecting the data from Twitter and Reddit for more than a year, and further visual analysis of this data using the frontend components in collaboration with experts in sociolinguistics is one of the next steps in our work. Another step is an evaluation of the system with a full-fledged user study.

This contribution opens up additional opportunities for future work on visual analysis of sentiment and stance as well as other aspects of temporal text data. Our approach could be enriched with additional types of analyses and visualizations, e.g., by using geospatial information, named entity recognition, and topic modeling [105]. Visual analysis of relations between the retrieved documents and reconstruction of the conversation structure (e.g., as discussed by El-Assady et al. [117]) enriched by sentiment and stance data is also an interesting prospect. Finally, the components related to data series analysis could be extended with the methods of predictive analytics [269] in order to forecast public sentiment and stance on specific targets of interest in social media.

Chapter 7

Further Applications of Stance Visualization

Contents

7.1 Interactive Exploration of Temporal, Thematic, and Geographic Aspects of Stance in Social Media Texts with StanceXplore 170
7.1.1 Background and Related Work 171
7.1.2 Visualization Methodology 173
7.1.3 Case Study 176
7.1.4 Summary 179
7.2 Visualization of Stance Categories in Longer Individual Texts with DoSVis 179
7.2.1 Related Work 181
7.2.2 Visualization Methodology 182
7.2.3 Use Cases 185
7.2.4 Discussion 187
7.2.5 Summary 188
7.3 Visualization of Sentiment and Stance for Supporting Argument Extraction with Topics2Themes 189
7.4 Summary 192

The previous chapters have demonstrated the progression in the methodology and applications of visual stance analysis from using a lexical approach and supporting only blog and forum data to using an ML-based stance classifier and supporting massive data from popular social media platforms, thus achieving the aims of the StaViCTA project. At the same time, we have developed additional visualization and visual analytics approaches to fill various gaps in the design space discussed in Chapter 3. These approaches demonstrate further opportunities to apply stance visualization for various data types and user tasks.

7.1 Interactive Exploration of Temporal, Thematic, and Geographic Aspects of Stance in Social Media Texts with StanceXplore

In digital humanities (DH) [362], the combination of text mining and visualization methods has resulted in tools that exploit modern or contemporary text corpora, extract linguistic patterns from various language resources, and provide the scholars with new and enriched digitalized educational material (e.g. [19,72,384]). The availability of large-scale, user-generated textual content from social media, such as reviews, opinions, and comments on politics and news, raised interest to the areas of sentiment analysis and opinion mining [240]. Techniques from these areas approach text analysis by extracting opinions and sentiments with the goal of aiding in the comprehension of how people feel about something, how these feelings are expressed, and how they spread [317].

Among the many related fields, one that has attracted attention lately is stance identification in discourse [243,296,381]. Stance taking is the way speakers position themselves in relation to their own or other people's beliefs, opinions, and statements in ongoing communicative interaction with others. Interesting findings about the attitude of people can be derived by looking at their stance regarding cultural, educational, social, and political events [166,445].

In this section [286][1] we present StanceXplore, a visualization approach for interactive exploration of stance-taking in social media. Stance analysis of content from social media is usually met with unique challenges due to the highly dynamic and heterogeneous language forms and constructional patterns in discourse, which can vary considerably depending on geography, time, and user identities/roles. All of these factors (or dimensions) of the data are relevant and must be considered together when exploring trends within a corpus, as such trends may be spread over different dimensions due to, e.g., specific reactions to relevant events (time), the effect of different cultural backgrounds (space), and previously unknown similarities between the writing of different groups. Our proposed visualization aids the exploration of stance in social media with a coordinated multiple views approach, where each of these dimensions can be explored separately, while, at the same time, all views react to brushing and filtering. We aim to help DH researchers discover stance-taking patterns in social media corpora by moving interactively from a general overview of the data's features into subsets defined by different combinations of filters for each dimension.

[1] This section is based on the following publication: Rafael M. Martins, Vasiliki Simaki, Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. StanceXplore: Visualization for the interactive exploration of stance in social media. In Proceedings of the 2nd Workshop on Visualization for the Digital Humanities, VIS4DH '17, 2017. © 2017 The Authors.

We demonstrate our visualization with a case study on the use of the English language by Twitter users from Sweden. By exploring Twitter's hashtag functionality, which allows users to specify topics that thematically orient their tweets, we show how our tool can support tasks such as: (1) identifying the stance distribution on the most frequent hashtags, (2) grouping these hashtags into broad thematic fields by similarity of content, (3) understanding the geographical distribution of stance-taking trends in the corpus, and (4) finding important events during a certain time period and checking how Twitter users have positioned themselves in relation to these events. We conclude that StanceXplore offers DH researchers the opportunity to obtain insights into the corpus that are not readily available without interactive exploration, are multidimensional by nature (i.e. are simultaneously based on independent aspects such as time, space, and language use), and are relevant to the comprehension of the dynamics of stance-taking in this specific Twitter user base.

7.1.1 Background and Related Work

While no universally accepted definition of DH exists, Schreibman et al. state that the discipline of DH "includes not only the computational modeling and analysis of humanities information, but also the cultural study of digital technologies, their creative possibilities, and their social impact" [362, p. XVII]. DH research on literary studies commonly use techniques developed under the umbrella of information visualization (InfoVis) and visual analytics (VA), more specifically, text visualization [236]. Important examples include the *literature fingerprinting* approach by Keim and Oelke [214] and VarifocalReader by Koch et al. [224]. As an attempt to introduce and popularize such methods and techniques among DH researchers, Drucker [108], for instance, analyzes some popular representations such as bar charts and bubble charts and hints at their shortcomings from the point of view of a DH scholar. Sínclair and Rockwell [384] introduce computational methods for text analysis to the DH audience and discuss their software suite called Voyant Tools, which includes several visual representations of text analysis results. The authors argue that such tools facilitate the exploration of the data and can lead to interesting discoveries. In general, these techniques focus on close and distant reading tasks, as described by Jänicke et al. [197] in a systematic overview of text visualization techniques for DH studies. Other recent examples are related to the analysis of text variants [19], named entities such as fictional characters [403], or arbitrary concepts and relationships within a large text document [71].

One common feature for most of the techniques is their orientation towards works of literature as the input data. In this work, however, we focus on data originating in social media rather than literary fiction. Chen et al. [76] provide an overview of the existing analysis and visualization methods for social media data, concluding that the most popular analytical approaches for such texts include extraction of keywords, detection of topics, and sentiment analysis.

Sentiment analysis usually involves automatic detection of positive, neutral, and negative content in texts [317]. Our recent survey [240] (see Chapter 3) discussed the corresponding sentiment visualization techniques developed both inside and outside the InfoVis/VA community, concluding that the majority of such techniques use social media data rather than customer reviews, editorial media data (e.g., news reports), or literature. Such techniques have been used to provide an overview of a Twitter corpus or a monitoring interface for a stream of text posts (tweets), usually with an option to drill down to the underlying texts on demand—which arguably also mirrors the distant and close reading tasks in DH discussed above. With regard to application scenarios, Diakopoulos et al. [103] and Marcus et al. [281] use their respective systems Vox Civitas and TwitInfo for digital journalism; Cao et al. [59] apply their system Whisper for the analysis of emergency events; and Humayoun et al. [186] analyze the public response to Brexit using their recent system TExVis.

Besides the analysis and visualization of positive and negative sentiments, emotion, or similar affective categories, social media data also provides interesting opportunities for the analysis and visualization of *stance*. Stance classification studies usually address stance-taking as a binary issue of the PRO or CON positioning of the speaker towards a fact/event/idea. In most cases, the data is extracted from online debates, where controversial opinions and stance-taking are observed, and they are automatically annotated [166, 445]. The classification accuracy achieved in these studies varied from 69 to 88%, and various different feature sets were used: lexicons, n-grams, cue words, post information, punctuation, and POS tags. More recent studies include other categories of subjectivity such as AGREEMENT AND DISAGREEMENT [391], CONDITION AND CONTRAST [390], or PREDICTION and UNCERTAINTY [381, 393].

The existing work in stance visualization includes the works by Almutairi [10] and El-Assady et al. [116], which focus on works of literature and transcripts of debates, respectively. Textual data from social media has been used for stance visualization in our system called uVSAT [243] (see Chapter 4); however, Twitter is not supported as a data source, and typical input documents are much larger/longer than tweets. Mohammad et al. [296] provide a dashboard visualization of a stance-annotated Twitter corpus, and ALVA [238] (see Chapter 5) supports visual analysis of the stance annotation process for utterances (sentences). In contrast to these approaches, the focus of this work is to provide an interactive stance visualization of a Twitter corpus with support for the temporal, geospatial, and topic perspectives—similar to TwitInfo [281], Whisper [59], or TExVis [186], but supporting stance analysis rather than the usual task of sentiment analysis.

7.1.2 Visualization Methodology

We propose to approach the challenge of interactively exploring stances in social media by using coordinated multiple views, where each view shows a different perspective of the data, i.e., a window into a specific aspect of the corpus under analysis. The focus of StanceXplore is the interactive brushing and filtering supported by visualization in such a way that each view can be explored independently, but, at the same time, the whole set of views adapts to users' actions. This design is inspired by Shneiderman's well-known visual information-seeking mantra [374]—"Overview first, zoom and filter, then details-on-demand"—and the implementation of the *distant reading* concept in visualization tools, as described in the survey by Jänicke et al. [197]. These related concepts can, when combined, be used effectively to direct readers to specific subsets of text that are relevant to the task at hand.

In order to be used with StanceXplore, a corpus must contain the full text of all tweets, be geolocalized, and be timestamped (these are related specifically to views (e), (c), and (d) in Figure 7.1, respectively). User information is not necessary, as the tweets are anonymized (every reference using @ is changed to @*User*). Each tweet is classified according to its stance using a Support Vector Machine (SVM) [422] classifier, previously trained on data extracted from political blogs and manually annotated by two linguistic experts. The ten stance categories are based in a cognitive-functional approach introduced recently [381,383,386]. AGREEMENT/DISAGREEMENT expresses a similar or different opinion (e.g., *OK then, I'll do that*); CERTAINTY expresses the speaker's confidence to its sayings (e.g., *Of course it is true*); CONTRARIETY expresses a compromising or contrastive opinion (e.g., *The result is fairly good, but it could be better*); HYPOTHETICALITY expresses a potential consequence of a condition (e.g., *If it's nice tomorrow, we will go*); NECESSITY expresses a request, recommendation, instruction, or obligation (e.g., *I must hand back all the books by tomorrow*); PREDICTION expresses a guess/conjecture about a future event (e.g., *I believe that he will do it for you*); SOURCE OF KNOWLEDGE expresses the origin of the speaker's sayings (e.g., *I saw Mary talking to Elena yesterday*); TACT/RUDENESS expresses pleasantries/unpleasantries (e.g., *You lazy bastard. Get lost*); UNCERTAINTY expresses doubt towards the speaker's sayings (e.g., *I don't know if that is the case, actually*); and VOLITION expresses wishes or refusals (e.g., *I wish I could join you next summer*). If no stance is detected, the tweet is NEUTRAL.

The total number of tweets per stance can be seen in the *Stances* view displayed in Figure 7.1(a), also encoded in the lengths of the bars. This view also functions as a color legend; the color assigned to each stance category in this view is used in most other views during the interactive exploration process. By clicking on the stances in this view, the user can choose to filter all the other views to include only tweets classified with the selected stances (in the example of Figure 7.1(a), all stances are active except NEUTRAL).

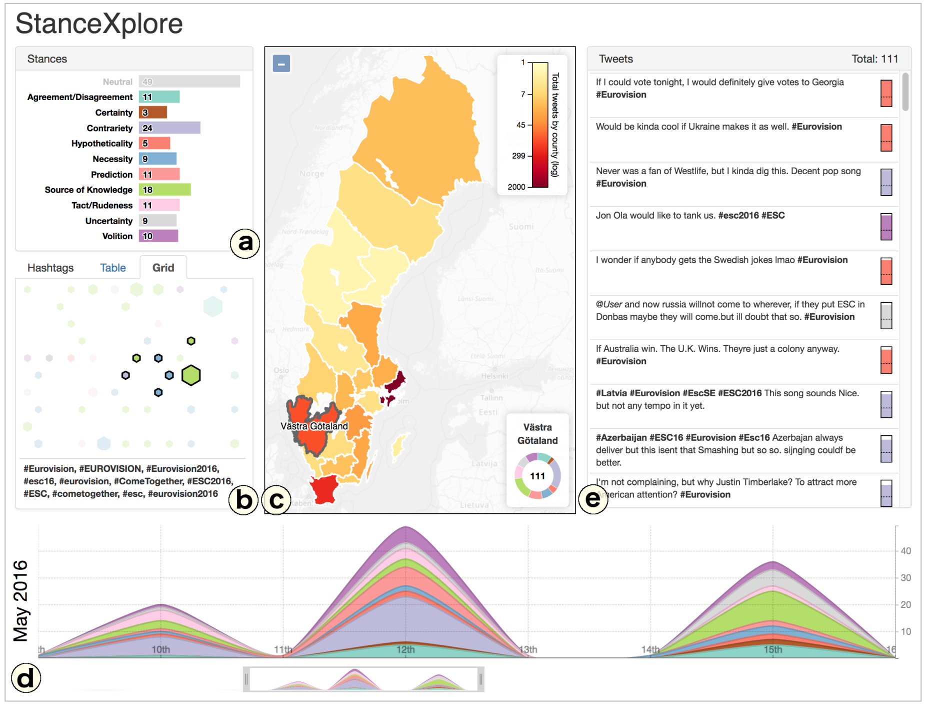


Figure 7.1: Overview of StanceXplore, showing English tweets from Sweden during May 2016 (case study from Section 7.1.3). The following filters are applied according to each view: (a) non-NEUTRAL tweets; (b) hashtags related to `#Eurovision`, manually selected by the user using the hexagon grid; (c) tweets originating from the county of Västra Götaland; and (d) from 9th to 16th of May 2016. Close reading of the tweets (e) shows the users' diverse opinions, their stances, and the classifier's confidence (using Platt scaling [326]). *Reprinted from [286] © 2017 The Authors.*

The *Hashtags* view displayed in Figure 7.1(b) shows the hashtags of the corpus in two interchangeable panels: the *Table*, in descending order of frequency, and the *Grid*, where they are grouped and distributed according to content similarity. These two panels offer two distinct but complementary views, and can be switched by the user as desired. When a hashtag is selected it is always shown on top of the table, while the rest of the hashtags are sorted in descending order by their *string* similarity to the selected one, as computed with the Sørensen-Dice coefficient [195]. This sorting highlights similar hashtags only by their name, e.g. `#Eurovision` and `#Eurovision2016`. The distribution of hashtags in the 2D hexagon grid is obtained with a Self-Organizing Map (SOM) [225] by extracting the best-matching units for each hashtag. In order to train the SOM, features are extracted from each hashtag h by first generating a vector space model representation $v(h)$ [355] that includes the content of every tweet $\{t \mid h \in t\}$, and then computing the TF-IDF transformation of $v(h)$ [354]. Essentially, the interpretation of the hexagon grid layout is simple: hashtags that occupy nearby hexagons are similar in *content*, with *content* referring to the aggregation of the text of all the tweets that include those hashtags. The visual encoding of the grid's hexagon units is further augmented with color, representing the single most frequent stance present on the hashtags of the unit, and size, representing the total sum of tweets in the hashtags that are included in the unit. Again, interacting with either of these two views will change the filtering on all the others, which in this case means that only tweets that contain any of the selected hashtags will be visible after a selection.

With the Twitter API's geo-search function [436] it is possible to estimate the location of each tweet within different administrative regions such as cities, counties, or states. This information is shown in the *Map* view displayed in Figure 7.1(c), along with a color encoding of the total (possibly filtered) number of tweets of each region. In the example from Figure 7.1(c), a log scale is used to improve the visibility of the values, since the difference in the total number of tweets between main and peripheral regions is very large. By interacting with the map, the user can explore the specific stance distribution within each region (see Figure 7.1(c), bottom-right), switch between different administrative levels of granularity (e.g. cities versus counties), and filter the data by limiting tweets to specific regions.

The temporal aspect of the corpus can be seen in the *Timeline* view displayed in Figure 7.1(d) as a stacked area graph that shows the number of tweets per day for each color-coded stance. The used visual encoding is similar to ThemeRiver [167], but we decided to use a fixed time axis as it increased the legibility of the view. An interactive filter is located below the timeline and allows the setting of a specific time range for the analysis (in the example, the time range is between days 8 and 17 of May 2016).

Finally, the *Tweets* view displayed in Figure 7.1(e) shows the full text of every tweet that satisfies all the filters defined interactively. Besides each tweet's text, a small bar shows the stance category assigned to the tweet (color) and the confidence of the classifier (size, computed with Platt scaling [326]), with the minimum size (lowest possible confidence) indicated by a dashed line.

7.1.3 Case Study

In this section we illustrate the features of StanceXplore[2] with a case study on the use of the English language by Twitter users in Sweden. The corpus was extracted using Twitter's REST API [436] with filters by language (English), country (Sweden), and time (May 2016). The aim of this case study is to highlight the ability of StanceXplore to support (1) free exploration of stance-classified data from social media, (2) detection of patterns and trends in stance-taking in social media along temporal and geospatial dimensions, and (3) the iterative and dynamic testing of hypotheses with responsive interaction and feedback from filtering.

We begin with the *Stances* view displayed in Figure 7.1(a). It shows that NEUTRAL is the most frequent result of the classification process. One explanation for this is that the classifier's training set was extracted from political blogs, with no size restrictions. Tweets, on the other hand, can be considered as fragmented discourse because of the limited character size of the text (it can be hard to formulate complete sentences within 140 characters) and the intervention of metacomments. As a result, the classifier sometimes cannot decide with strong confidence for a stance, and when no stances are detected the tweet is classified as NEUTRAL. Another reason is the fact that stance is a very subtle concept that can be difficult to identify, and even in the original manual annotations the NEUTRAL utterances were very frequent. In order to neutralize the effect of NEUTRAL as the most dominant stance and allow for the exploration of different patterns, we disable this category by shift-clicking on it. From now on only non-NEUTRAL tweets will show up in all the coordinated views.

7.1.3.1 Investigation of Cultural Events

We next look at the *Hashtags* grid in Figure 7.1(b) and notice that the largest hexagon unit (with 1,916 tweets) contains only one hashtag: `#Eurovision`. The Eurovision Song Contest (ESC) is a traditional TV song competition that takes place every year between (mainly) European countries. Clicking on this hexagon unit lets us focus solely on tweets that include the hashtag `#Eurovision`. A quick session of close reading of the tweets using the view in Figure 7.1(e) indicates that ESC was held in Stockholm that month, which made it a hot topic of Twitter

[2] A demo video for StanceXplore is available at `https://vimeo.com/230334496` (last accessed in February 2019).

in Sweden. However, maybe the biggest change after this new filtering is in the timeline: the vast majority of tweets were posted between 10 and 15 of May 2016, with almost zero mentions outside that range. Indeed, ESC took place on 10, 12, and 14 of May 2016. But while the dates in which the contest took place may be clear, the stance distribution is not, since too much space is wasted on empty days.

To improve this situation, we zoom into the desired time range using the timeline slider displayed in Figure 7.1(d), which then allows the relevant tweets and stances to occupy most of the space allocated to the timeline. A few trends on the stance-taking regarding ESC are now observable: SOURCE OF KNOWLEDGE (which sets the color of the hexagon), NECESSITY, and CONTRARIETY are regularly strong throughout the period; VOLITION shows a peak of representation in the second day of the contest; and UNCERTAINTY is the strongest stance after the final day.

One natural way to proceed with the exploration is to go back to the *Hashtags* grid and browse through the hexagon units near the selected one; these are the ones that are similar in content to the current focus, so they might be relevant to enrich the results. This leads to an interesting insight into the corpus: many different hashtags were used to refer to the same event. While some might be easy to locate with conventional string-comparison methods, such as those with different capitalizations (`#eurovision`, `#EUROVISION`) and suffixes (`#Eurovision2016`), others might be more challenging to detect without the content-based similarity visualization, such as abbreviations (`#ESC`, `#esc16`) and specific themes (`#ComeTogether`). However, two other nearby hashtag groups prove to be even more interesting and insightful. The first one, `#AUS`, is related to the fact that Australia participated in ESC 2016 even though it is not an European country. By investigating the stances and close reading of the tweets including this hashtag, it is possible to see that the Australian performance was well-liked and received positive feedback, especially on the second day of the contest.

The second interesting nearby hashtag group includes both `#Ukraine`—the winner of ESC 2016—and `#Russia`. Again, by investigating the stances and close reading of the tweets after filtering by this hashtag group, we can infer that a fierce dispute took place between the two countries during the contest, with tweets moving from predominantly PREDICTION in the first days to a small surge of AGREEMENT/DISAGREEMENT and UNCERTAINTY after the final results.

7.1.3.2 Aspects of Geographical Distribution

With all the filters reset, one look at the *Map* view shows clearly that the geographical distribution of English tweets in Sweden is not balanced among all counties. In fact, only three areas contain the vast majority of English tweets: Stockholm, with 55,712 tweets; Västra Götaland, with 16,029 tweets; and Skåne, with 14,295 tweets. Not surprisingly, these counties include, in this same order, the three largest cities in Sweden—Stockholm, Gothenburg, and Malmö. The counties with the most tweets are consistently located in the southern part of Sweden; as we move towards the northern parts of the country, the numbers decrease significantly. This is compatible with the fact that the northern regions of Sweden, known for their increasingly harsh weather, are more sparsely populated than the south. Considering the characteristics of this distribution, an analysis of the busiest areas of the country might be the most common approach. In this section, however, we decided to take a different path and explore a less obvious question: *what are people tweeting about (in English) outside the main areas, and what are their attitudes regarding their chosen topics?*

For this, we first turn to the *Map* view and filter only tweets that come from Norrbotten—the northernmost county in Sweden—totaling 717 tweets distributed in all stances. NEUTRAL is the most frequent stance, representing almost half of the subset of tweets with 353 tweets, followed by SOURCE OF KNOWLEDGE (75 tweets) and NECESSITY (71 tweets). Close inspection of the tweets classified as SOURCE OF KNOWLEDGE shows that average confidence is low, while the opposite is true for NECESSITY. For the rest of the analysis, we again disable NEUTRAL and focus on the rest of the stances.

Looking next at the *Hashtags* grid, we notice that, apart from `#Eurovision`, two other hexagon units are salient (due to their size), including hashtags such as `#Jobs`, `#CareerArc`, and `#Hiring`. The time distribution of these posts shows a periodical pattern throughout the whole month, with tweets being made every few days (with periods of inactivity between them). Close reading of the filtered tweets shows that they are all very similar job advertisements; the strong use of Twitter for job advertisements in this area may be related to a possible difficulty of attracting personnel due to their remote location. However, one interesting observation is that the classifier has achieved low confidence with these tweets, assigning diverse stances such as CONTRARIETY and SOURCE OF KNOWLEDGE.

Repeating the same analysis steps with Västerbotten, a neighboring county immediately to the south of Norrbotten, we notice an especially salient hexagon marked with AGREEMENT/DISAGREEMENT. It contains sports-related tags such as `#Endomondo` and `#endorphins`. Close reading of the tweets after filtering shows that all (but one) have the same structure and almost the same text: a report on the completion of a sports activity. These are known to be generated automatically by health-monitoring applications, and are not supposed to express any specific stance. From this investigation with StanceXplore we can notice, however, that the classifier did not assign the expected NEUTRAL stance, but used AGREEMENT/DISAGREEMENT with low confidence. This insight could be useful to help DH researchers in finding flaws in the stance classification system and improve it with more training data and better examples.

7.1.4 Summary

In this section we proposed StanceXplore, a visualization approach aimed at supporting DH researchers in the interactive exploration of stance-annotated textual content originated from social media. The proposed visualization uses coordinated multiple views to simultaneously show different aspects of the corpus under analysis, in a way that allows the user to explore each view independently and to interactively apply filters that affect the outcome of all the views. With a case study of the use of English by Twitter users from Sweden, we demonstrated how StanceXplore can be used to support a progressive exploration process, starting from a general overview of the data (distant reading) and moving step-by-step into more specific subsets of the corpus (close reading) that exhibit different stance-taking patterns and trends, defined by multiple aspects (or dimensions) of the data such as time, space, and similarities/dissimilarities in the use of the English language.

7.2 Visualization of Stance Categories in Longer Individual Texts with DoSVis

Textual data has been playing an increasingly important role for various analytical tasks in academic research, business intelligence, social media monitoring, journalism, and other areas. In order to explore and make sense of such data, a number of text visualization techniques have emerged during the last 20 years [197,236]. The majority of text visualization techniques rely on methods originating from computational linguistics and natural language processing which analyze the specific aspects of texts, such as topic structure, presence of named entities, or expressions of sentiments and emotions. The latter one, i.e., sentiment analysis / opinion mining, has usually been associated with data domains such as customer reviews, social media, and to a lesser degree, literature and political texts [295,317]. There is also research on sentiment analysis of business reports and CEO letters which studies the relation between the language and financial indicators [211,310]. The existing sentiment visualization techniques for textual data support a variety of data domains, data source types, and user tasks [240].

At the same time, few existing visualization techniques make use of another method related to sentiment analysis—stance analysis [296, 383, 391]. Stance analysis of textual data is concerned with detecting the attitude of the writer ranging from the general AGREEMENT/DISAGREEMENT with a certain utterance or statement (e.g., “I hold the same position as you on this subject”) to the more fine-grained aspects such as CERTAINTY/UNCERTAINTY (e.g., “I am not completely convinced that it really happened”). The StaViCTA project has taken the latter approach in order to develop an automatic stance classifier and visualize stance detected in textual data. The existing stance visualization techniques have usually

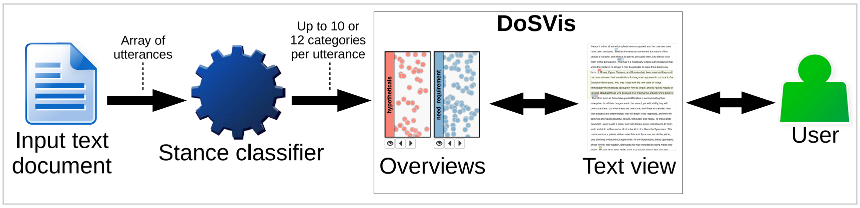


Figure 7.2: The architecture of our approach. DoSVis uses the output of the stance classifier for a text document divided into utterances. Each utterance may be simultaneously labeled with multiple stance categories. *Reprinted from [239] © 2018 SciTePress.*

focused on political text data such as transcripts of debates [115], blog posts and comments [238, 243], and tweets [286, 296].

In this section [239][3], we explore other possible applications of visual stance analysis and focus on data domains and user tasks that are not addressed in the existing literature. In contrast to the techniques which support visual analysis of multiple short documents such as social media posts, we look into scenarios involving exploration of longer documents such as business reports [211] and works of literature [384]. Our visualization approach, called DoSVis (Document Stance Visualization), uses the output of the automatic stance classifier developed as part of the StaViCTA project to provide the users with an environment for exploring the individual documents' contents, annotated with the stance categories detected at the utterance or sentence level (see Figures 7.2 and 7.3). The main contributions of this section are the following:

- a visualization approach for individual text documents that supports visual stance analysis; and
- a demonstration of application scenarios for visual stance analysis in several data domains.

The rest of this section is organized as follows. In the next subsection, we shortly describe the background of stance analysis and existing approaches for stance visualization as well as text document visualization. Afterwards, we discuss our visualization methodology in Section 7.2.2. We illustrate the applicability of our approach with several use cases in Section 7.2.3 and discuss some aspects of our findings in Section 7.2.4. Finally, we provide a summary in Section 7.2.5.

[3]This section is based on the following publication: Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. DoSVis: Document stance visualization. In Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP '18) — Volume 3: IVAPP, short paper track, IVAPP '18, pages 168–175. SciTePress, 2018. doi:10.5220/0006539101680175 © 2018 SciTePress.

7.2.1 Related Work

In this subsection, we establish the proposed approach in the context of the related work on (1) stance analysis & visualization and (2) visualization of individual text documents.

7.2.1.1 Stance Analysis and Visualization

A standard approach to automatic stance analysis of textual data focuses on the detection of AGREEMENT/DISAGREEMENT or PRO/CONTRA positions of the author, typically towards the given topic or target [296, 391]. The latter work describes the results of a shared task (a contest) on stance analysis for a Twitter data set with the majority of submissions using support vector machines (SVM) or neural networks as classifiers and n-grams, word embeddings, and sentiment lexicons as features. The same authors also introduce a dashboard-style visualization of their stance data set that provides a general overview, but does not focus on the contents of individual documents. Another visualization approach for the analysis of speakers' positions towards corresponding topics is ConToVi [115]. This approach is designed for monitoring of political debates, and it also focuses on the overall trends and topics rather than the text content.

There also exist other approaches that focus on a wider set of categories related to stance, such as CERTAINTY/UNCERTAINTY [243] or SPECULATION and CONDITION [390]. Similar to the other stance visualizations discussed above, ALVA [238, 242] (see Chapter 5) focuses on the overview of a data set or corpus consisting of multiple utterances or sentences from blog posts and comments. Finally, StanceXplore [286] discussed in Section 7.1 provides multiple coordinated views for exploratory visual analysis of a corpus of tweets labeled with multiple stance categories by a stance classifier. In contrast to all these works, our contribution proposed in this section is designed for a detailed exploration of individual documents which are much larger/longer than social media posts.

7.2.1.2 Visualization of Individual Text Documents

The existing taxonomies of text visualization techniques recognize individual documents as one of the options of data sources as opposed to corpora [197] or text streams [236, 240], for instance. A typical example of such a document is a work of literature which can be explored by a scholar in digital humanities using a software tool with some form of support for visualization [108]. Providing an overview of the content of individual documents dates back to early techniques, such as SeeSoft [113] and TileBars [170]. Both provide pixel-based summaries for text segments constituting the documents. Affect Color Bar [262] implements a similar idea, but uses categories related to emotions. The resulting visualization allows the user to get an overview of the affective structure of a text, such as a novel, and to navigate to the corresponding segment for close reading. Ink

Blots [1] is a technique based on highlighting regions of text documents with background bubble plots. The resulting bubble plots can be used without the actual text content for overview purposes. Keim and Oelke describe a compact pixel-based technique which can use various text features to represent visual fingerprints of text segments [214]. VarifocalReader [224] supports both distant and close reading (see [197], for example) by using topic segmentation, overview of text structure, and highlighting of automatically annotated words or chunks. Lexical Episode Plots [154] provide an overview of topics recurring throughout a text (more specifically, a transcript of political debates). uVSAT [243] uses scatterplot-like representations for overviews of stance markers detected in a text document (see Chapter 4). Finally, Chandrasegaran et al. implement an interactive interface for visual analysis and open coding annotation of textual data, which includes structural overviews for distant reading and colored text view for close reading [71]. Our approach adopts ideas similar to many of such visualization techniques in order to provide an overview of stance classification results for an individual document at the utterance level. In contrast to some of the techniques discussed above, though, our goal is to preserve the two-way mapping between utterances and visual items used in the overview, so that the users could refer to the overview while performing close reading.

Many existing techniques which provide support for close reading use a certain form of highlighting individual words or chunks of text [405] to represent custom annotations or labels. For example, Ink Blots [1] highlight an approximate region based on the position of certain marker words or features. Serendip [8] highlights words relevant to specific topics. uVSAT [243] highlights words and n-grams from the lists of stance marker words and topic terms. Chandrasegaran et al. provide the user with controls for highlighting specific parts of speech and information content in the detailed text view of their interactive interface [71]. As opposed to these approaches, our goal for representing the textual content of documents is to support the output of a stance classifier with multiple non-exclusive categories. Therefore, we use a strategy relying on non-intrusive glyphs rather than direct highlighting of the text to represent the classification results.

7.2.2 Visualization Methodology

The input data for our tool DoSVis is generated by a stance classifier pipeline developed by our project members [238, 242, 383, 386]. The pipeline (see an illustration in Figure 7.2) divides the input text into utterances and then classifies each utterance with regard to a set of stance categories such as UNCERTAINTY, HYPOTHETICALS, and PREDICTION. The tasks related to the set of stance categories, the data annotation process, and the training of the classifier were carried out in collaboration with our experts in linguistics and computational linguistics. The stance categories used by the classifier are not mutually exclusive, i.e., several categories may be simultaneously detected in any given utterance. Our approach

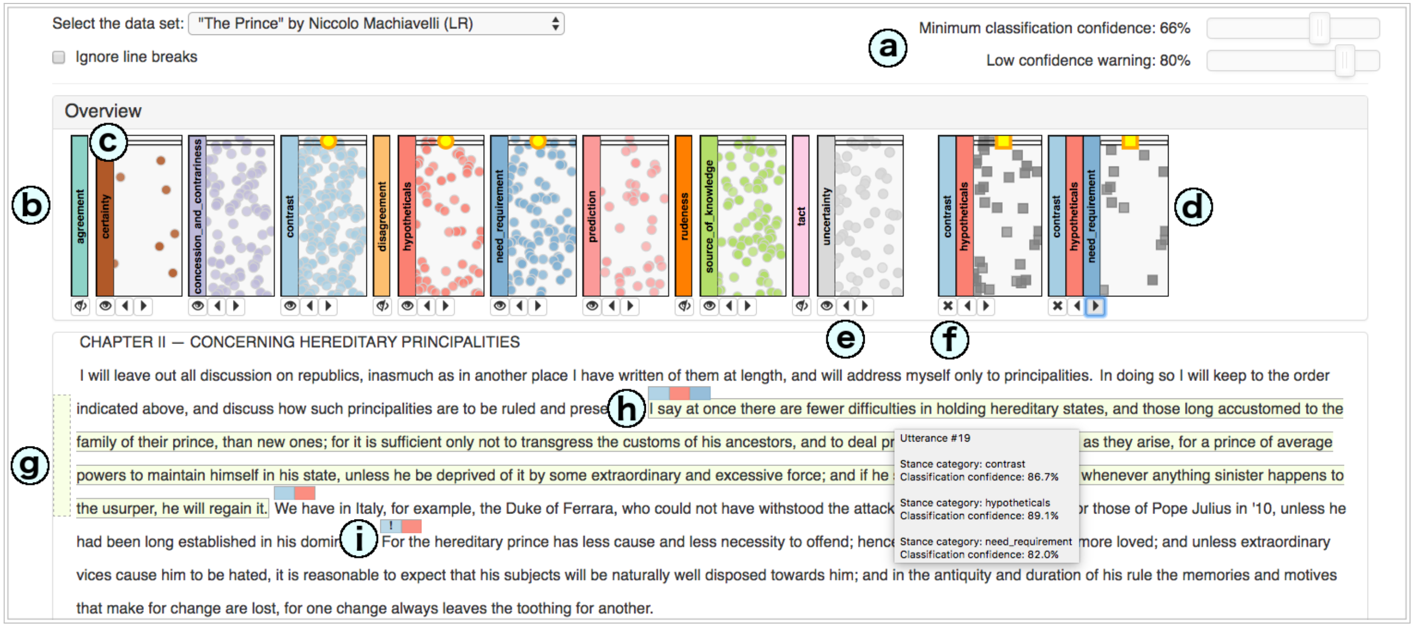


Figure 7.3: Visualization of a 16th century political treatise "The Prince" by Niccolò Machiavelli in our tool DoSVis: (a) sliders used for global filtering and toggling of warning symbols; (b) scatterplot-like overviews based on the detected occurrences of stance categories in the text; (c) a viewport rectangle representing the currently visible area of the document; (d) overviews of detected stance category combinations (created with drag'n'drop); (e) filtering and navigation controls for category overviews; (f) filtering and navigation controls for category combination overviews; (g) the detailed text view including a sidebar mark for the currently highlighted utterance (in yellow); (h) a rectangular glyph representing the stance categories detected in the utterance; and (i) a warning symbol (exclamation mark) representing low classification confidence. *Reprinted from [239] © 2018 SciTePress.*

can actually be generalized to any set of categories or labels associated with utterances. We have tested this by using two versions of the stance classifier: (1) an SVM-based classifier with 10 stance categories [242], and (2) a logistic regression (LR)-based classifier with 12 stance categories [386]. Both of these classifiers also provide a form of confidence estimates for the classification decisions based on (1) Platt scaling [326] and (2) probability estimates [183], respectively. After the initial preprocessing and classification stages, the input data for the visualization module consists of a JSON file with an array of utterances labeled with classification results.

Our approach is based on a rather straightforward visual design in order to be intuitive to the users without prior training in visualization. DoSVis is implemented as a web-based system[4] using JavaScript and D3 [94]. Its user interface depicted in Figure 7.3 provides an overview and a detailed text view for the selected document. The users can control the interpretation of line break symbols to adjust the document layout, which might be preferable for some documents converted from the PDF format (see Section 7.2.4). The sliders located at the top right (see Figure 7.3(a)) specify the classification confidence thresholds for displaying the classification results at all and displaying warning symbols (exclamation marks within the glyphs, see Figure 7.3(i)), respectively, in order to help the users focus on more reliable results.

The overview of stance classification results consists of scatterplot-like representations for individual stance categories displayed in Figure 7.3(b). We have decided to follow this design with separate representations for categories due to the data considerations described above. Any utterance in our data can potentially be labeled with up to 10 or 12 stance categories simultaneously, therefore, alternative designs would have to use overly complex glyphs or ignore the resulting categories to some extent [242, 286]. Each utterance with a detected stance category is represented by a dot marker in the corresponding overview plot. The dot position itself reflects the position of the utterance in the text. More specifically, the position is based on the coordinates of the HTML element representing the utterance relative to the overall text view HTML container. Each stance category is associated with a certain color based on the color maps from ColorBrewer [85]. The opacity of the dot is based on the classification confidence value. Visual items with confidence values below the global threshold are hidden. The overview plots support pan & zoom for the vertical axis, and the default zoom level is set to fit the complete document text. The area currently visible in the main text view is represented by a viewport rectangle in each plot (see Figure 7.3(c)). Each overview supports details on demand and navigation over the text by hovering and clicking, respectively. The users can also hide the overview

[4] A demo video for DoSVis is available at `https://vimeo.com/240178420` (last accessed in February 2019).

plots and navigate to the previous/next occurrence of the corresponding stance category by using the buttons located under each plot (see Figure 7.3(e)).

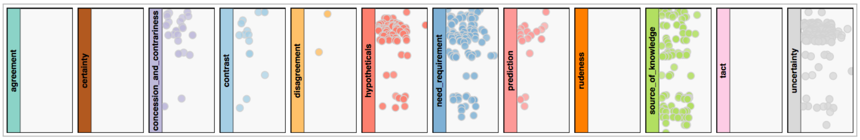
Besides the interactions with a single overview plot, the users can drag-and-drop the plots onto each other. This results in a new plot providing the overview of utterances which are labeled with the corresponding combination of categories. Such plots for the combinations of two and three categories, respectively, are displayed in Figure 7.3(d). In order to distinguish such combination plots from regular category overview plots, we have used rectangular markers with a dark gray color. The opacity mapping and global filtering behavior for the visual items are based on the lowest confidence value with regard to the category combination. Such combination overview plots support the same interactions as regular category overview plots, except for the "hide" button being replaced by the "remove" button (cf. Figure 7.3(e+f)).

DoSVis also provides a detailed text view (displayed in Figure 7.3(g)) with stance category labels and details on demand, thus supporting both distant and close reading approaches [197]. We use sets of non-intrusive rectangular glyphs located above utterances to represent the categories detected by the classifier (see Figure 7.3(h)). These glyphs share the color coding, opacity mapping, and filtering behavior with the overview plots. They are also connected with linking and brushing—see the elements highlighted in yellow in Figure 7.3(b+d+g). One additional design element used for the glyphs in the main text view is a low confidence warning represented by an exclamation mark, as depicted in Figure 7.3(i). Such marks are displayed for the classification results with confidence values lower than the global threshold controlled by the corresponding slider.

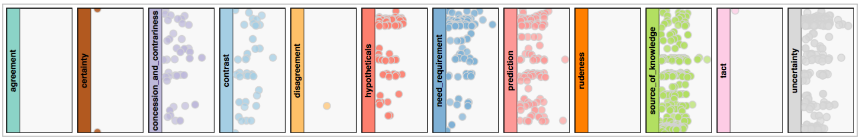
7.2.3 Use Cases

With the current application of stance visualization, we focus on use cases beyond social media monitoring. One of them is the exploration of business reports: an analyst or an investor may be interested not only in the reported financial results, but also in the language used throughout the report. Our tool DoSVis could be used in this case to explore the results of automatic stance analysis similar to the existing application of sentiment analysis [211, 310]. The users would benefit from the opportunity to get an overview for the complete text and to navigate between stance occurrences to explore such longer texts in detail and verify the classification results.

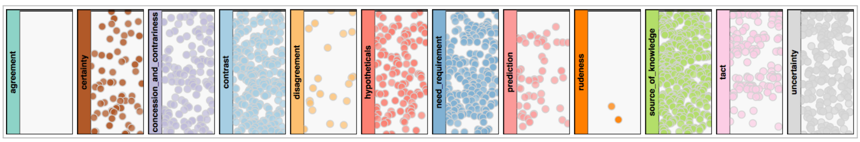
For example, the PDF versions of the 2015 annual reports from Tableau Software and Yahoo Inc. contain 98 and 180 pages, respectively. Their overviews in DoSVis are displayed in Figure 7.4(a+b) at the selected classification confidence level of 66%. It is interesting to note that both reports contain a rather large number of expressions of UNCERTAINTY which is detected in approximately 8% of utterances in both cases. The density of such expressions is particularly high in



(a) Tableau Software 2015 annual report



(b) Yahoo Inc. 2015 annual report



(c) "The Hound of the Baskervilles" by Arthur Conan Doyle

Figure 7.4: Overviews of stance categories detected in several documents with the LR classifier at 66% classification confidence. *Reprinted from [239] © 2018 SciTePress.*

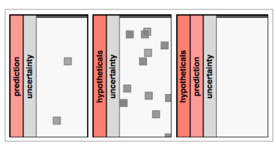


Figure 7.5: Overviews of several stance category combinations detected for the data in Figure 7.4(c). *Reprinted from [239] © 2018 SciTePress.*

the early sections of the reports where forward-looking statements are located. The occurrences of UNCERTAINTY combined with HYPOTHETICALS or PREDICTION are mainly found in the same regions of the text. The comparison between the two documents with regard to specific categories reveals that the Tableau Software report has a larger proportion of detected HYPOTHETICALS (3.79% vs 2.67% of utterances) and NEED AND REQUIREMENT (5.01% vs 3.08%) than the Yahoo Inc. report, and a lower proportion of PREDICTION (1.00% vs 3.91%). It is also interesting to note that categories such as AGREEMENT, DISAGREEMENT, TACT, and RUDENESS are almost absent in the results, which can be explained by the genre of these documents.

Another application of our approach is related to the exploration of works of literature. Scholars in digital humanities [362] could make use of the support for distant and close reading provided by DoSVis. Figure 7.4(c) displays an overview of Arthur Conan Doyle's "The Hound of the Baskervilles" and provides the user with a general impression of the stance category occurrences in the text. In contrast to the financial reports described above, it is easy to notice that the novel contains much more occurrences of categories such as CERTAINTY, DISAGREEMENT, and TACT. Our approach could, therefore, be interesting to the scholars in digital humanities and linguistics with regard to the analysis of differences between genres of text by using category overviews as sort of a fingerprint [214]. Furthermore, the scholars could make use of the opportunity to analyze occurrences of stance category combinations by drag-and-dropping the overview plots. Several recent papers on stance analysis [381, 393] discuss co-occurrences of such stance categories as PREDICTION with UNCERTAINTY and HYPOTHETICALS with UNCERTAINTY, respectively, in political blog data. Figure 7.5 provides an overview of corresponding category combinations in "The Hound of the Baskervilles", which can be interesting to the researchers in digital humanities. The user can immediately get insights about the distribution of these stance category combinations, e.g., there are just two instances of PREDICTION with UNCERTAINTY, and no occurrences of combinations of all three categories are detected at the current classification confidence level. By clicking visual items or using the navigation buttons, the user can then navigate to the corresponding utterances for close reading. In this case, exploratory analysis with DoSVis would allow the user to identify concrete interesting cases as opposed to interpreting overall category statistics computed with non-interactive analyses.

7.2.4 Discussion

In this subsection, we discuss several aspects of our work related to underlying data processing methods as well as scalability concerns.

7.2.4.1 Stance Classification

The existing methods of automatic stance classification do not reach the same levels of precision/accuracy [296] as, for instance, sentiment classification methods, especially for topic-independent tasks [390]. This raises concerns related to the users' trust in classification results and the corresponding visualization, especially when low confidence values are reported by the classifier. Nevertheless, our proposed visualization approach allows the users to explore the classification results in detail and make the final judgment themselves. DoSVis can also easily make use of improved classifiers available in the future.

7.2.4.2 Preprocessing

In order to apply our approach to the analysis of various reports and books available as PDF documents, text data must be extracted and classified utterance after utterance. For longer documents, manual preprocessing is not feasible, and automatic conversion of PDF to plain text often results in noisy or almost unusable data [87]. It would also be desirable to preserve the original layout of document pages in many cases. We consider this as part of the future work which could be based on the previously described approaches [280,406].

7.2.4.3 Scalability

We have tested DoSVis with documents of several sizes/lengths, the longest being the 2017 Economic Report of the President of the US (599 pages). Our tool is able to display the corresponding classification results, albeit the performance of some interactions is rather low. The largest delays are caused by the web browser's layout events for the main text view. The potential solution is to avoid displaying the complete document text in such cases and use some form of sectioning instead—for instance, Asokarajan et al. propose a visualization strategy relying on multiple text scales [19,20]. As for the other scalability concerns, the overviews for such large documents are affected by overplotting. Our current implementation relies on pan & zoom to allow the users focus on shorter text segments and avoid this effect. Alternative solutions could involve some forms of semantic zooming, although it could potentially affect other interactions.

7.2.5 Summary

In this section, we have demonstrated how stance classification results can be used for visual exploration of a text document such as a business report or a novel. We have described our tool DoSVis which provides an interactive visualization of multiple stance categories detected in the text. DoSVis can be used to estimate the number of utterances with detected stance in a given text, compare the results for several stance categories, and explore the text in detail. With the stance classification accuracy improving over time, we believe that such an approach will be useful for scholars and practitioners, as illustrated by our potential use cases. We plan to provide our prototype to the expert users in order to get their feedback and refine our implementation. Our plans for further development of DoSVis also include a user study in order to evaluate some of our design decisions.

While DoSVis focuses on individual text documents, our future work includes the development of novel visual representations for stance detected in text corpora, temporal and streaming text data, and text data associated with geospatial and relational attributes.

7.3 Visualization of Sentiment and Stance for Supporting Argument Extraction with Topics2Themes

The task of qualitative text analysis, in particular, the identification of arguments and themes, requires a lot of effort from the analyst (see Figure 7.6). Computational extraction of main topics in a document or a corpus has been shown to be an effective first step for such analyses [28, 395]. However, the typical output of topic modeling algorithms at the detailed level is also overwhelming. The fields of information visualization and visual analytics provide approaches for representing and interacting with textual data and results of various text analyses (including topic modeling [105] and sentiment & stance analysis [240]) to solve this problem.

In our previous work, we have introduced an interactive visualization tool, called Topics2Themes [387], that is used to assist the task of extraction and annotation of arguments in texts (as demonstrated in Figure 7.7) by providing a Jigsaw-like list interface [156, 400]. Topics2Themes was primarily designed to support analyses of vaccination-related texts with a limited number of opinions or stances towards this issue, such as FOR, UNDECIDED, or AGAINST[5]. In this section [244][6], we describe an application of a customized version of Topics2Themes to a different genre of data (political comments from Reddit) and a different set of supported sentiment and stance analyses (with multiple categories).

The data processing pipeline of Topics2Themes [385] includes the following steps: 1) optional classification or manual tagging of stances associated with text documents; 2) preprocessing including stop word removal, collocation detection, and clustering of semantically similar words; and 3) topic modeling with either the LDA or NMF algorithm. For the present work, we have customized Topics2Themes to use the classifiers developed as part of the StaViCTA project [393] for the first step. Then, we applied the tool to a data set of about 200 political comments from Reddit created during spring 2018. Each document was automatically labeled with its dominant sentiment category (POSITIVE, NEUTRAL, or NEGATIVE) by the VADER sentiment classifier [191] and a set of detected stance categories, such as CERTAINTY or CONTRAST, by our custom stance classifier [393] (the complete list is provided in Table 7.1).

[5] A demo video for the original version of Topics2Themes is available at `https://vimeo.com/257474950` (last accessed in February 2019).

[6] This section is based on the following publication: Kostiantyn Kucher, Maria Skeppstedt, and Andreas Kerren. Application of interactive computer-assisted argument extraction to opinionated social media texts. In Poster Abstracts of the 11th International Symposium on Visual Information Communication and Interaction, VINCI '18, pages 102–103. ACM, 2018. doi:10.1145/3231622.3232505 © 2018 The Authors.

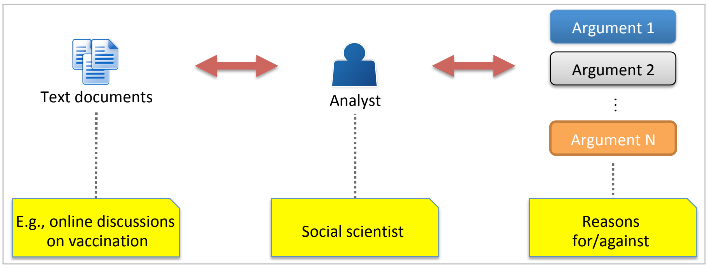


Figure 7.6: Motivation for the Topics2Themes approach: manual analysis of multiple arguments, opinions, or themes in multiple texts may be cumbersome or even infeasible for the analysts.

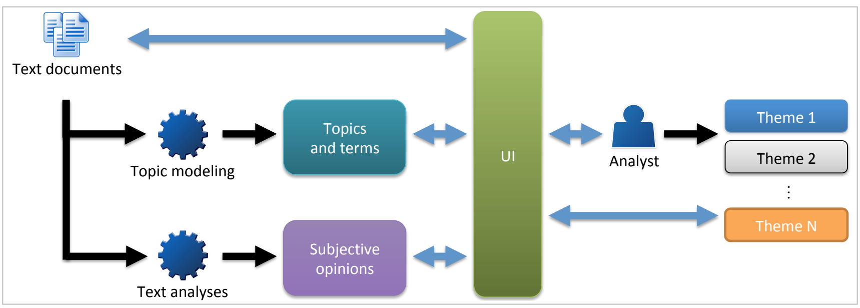


Figure 7.7: Overview of the Topics2Themes approach: a corpus of text documents is processed automatically with both topic modeling and text classification methods, and the results are presented to the user with an interactive UI. The user can define various *themes* related to multiple documents in order to formulate main arguments and viewpoints recurring in the data.

Table 7.1: Classification categories supported by the custom version of Topics2Themes

Sentiment:	POSITIVE, NEUTRAL, NEGATIVE
Stance:	AGREEMENT, CERTAINTY, CONCESSION AND CONTRARINESS, CONTRAST, DISAGREEMENT, HYPOTHETICALS, NEED AND REQUIREMENT, PREDICTION, RUDENESS, SOURCE OF KNOWLEDGE, TACT, UNCERTAINTY

Reprinted from [244] © 2018 The Authors.

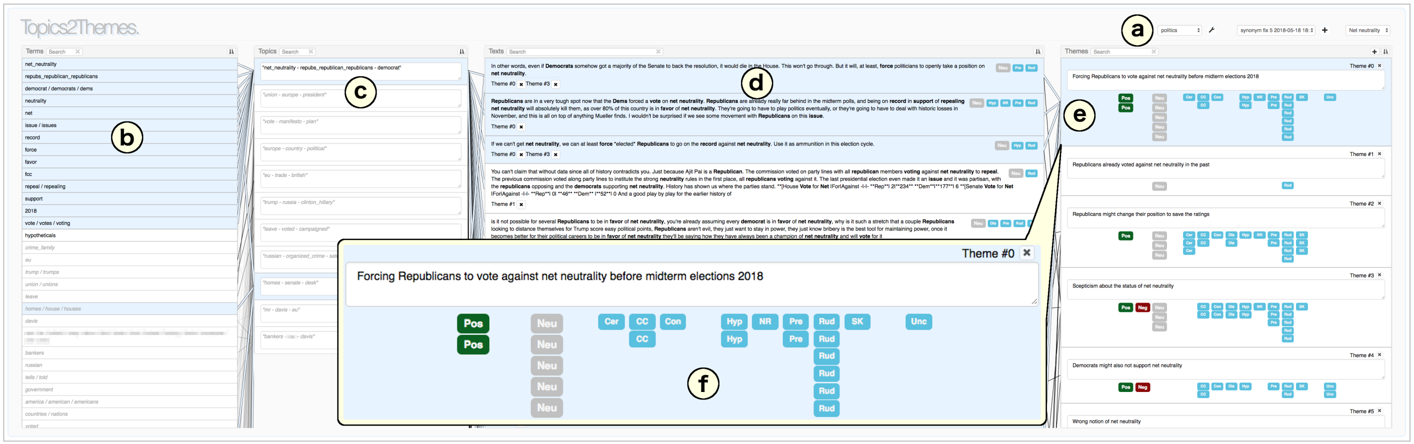


Figure 7.8: The screenshot of Topics2Themes customized for multiple sentiment & stance categories discovered in political comments from Reddit: (a) data loading controls; (b) list of topic terms; (c) list of topics (the first one is selected); (d) list of documents (displaying detected sentiments & stances and associated themes); (e) list of user-defined themes (displaying sentiments & stances for all associated documents); and (f) zoomed-in cutout of the first theme in (e). Here, the user hovered over the first theme displayed in (e), which affected highlighting of elements and links in other lists (in blue color). *Reprinted from [244] © 2018 The Authors.*

Using the frontend of Topics2Themes displayed in Figure 7.8, we were able to select an interesting topic on Internet neutrality among the output of the NMF algorithm (see Figure 7.8(c)). By reading the related documents and identifying the recurring themes (see Figure 7.8(d–e)), we established the main arguments in the ongoing discussion about the upcoming U.S. Senate vote on Internet neutrality[7]. The resulting user-labeled themes displayed in Figure 7.8(e–f) also provide a barchart-like overview of various sentiment and stance categories discovered in the associated documents (see Table 7.1), thus providing us with an opportunity to compare the opinions related to the themes.

Here, we briefly demonstrated the potential applications of our interactive tool Topics2Themes to political texts from social media. Topics2Themes allows the users to visually explore the output of topic modeling and stance classification algorithms, conduct close reading of the original texts, and annotate arguments for various viewpoints by defining recurring themes. Our future work includes collaboration with domain experts, evaluation of our proposed approach, and integration of Topics2Themes into larger visual stance analysis workflows.

7.4 Summary

In this chapter, we have discussed three stance visualization approaches that complement the work introduced in Chapters 4–6 by supporting further parts of the design space described in Chapter 3. StanceXplore uses the information about individual stance categories dominating in tweets and allows the users to investigate thematic and geographic data aspects in addition to temporal and text data. DoSVis focuses on longer individual documents and data domains such as literary fiction and business reports in contrast to collections of short documents from social media. Topics2Themes is designed to facilitate qualitative text analysis and argument extraction tasks, and its customized version is able to support multiple sentiment and stance categories reported by our classifiers. These three approaches demonstrate the potential of further applications of stance visualization and visual stance analysis to various data types, domains, and user tasks. Additional considerations for such future work are discussed in the next and final chapter of this dissertation.

[7]The vote took place on May 16, 2018: `https://edition.cnn.com/2018/05/16/politics/net-neutrality-vote-senate-democrats/` (last accessed in February 2019).

Chapter 8

Conclusions and Future Work

Contents

8.1 Research Findings 193
8.1.1 Analysis of Research Objectives 193
8.1.2 Summary of Scientific Contributions 198
8.2 Discussion 199
8.2.1 Design Decisions and Lessons Learned 199
8.2.2 Research Limitations and Validation Concerns 201
8.2.3 Ethical and Societal Aspects 203
8.3 Future Work 204

This dissertation started with the introduction of the research problem of sentiment and stance visualization, the goal of our research work, and the concrete objectives to be reached. In the final chapter of the dissertation, we discuss these objectives again in relation to the findings described in the previous chapters and then summarize the overall contributions of this work. Afterwards, we discuss various aspects of our research presented in this dissertation, including our design decisions, research limitations, and validation concerns. Additionally, we briefly address ethical concerns and societal implications relevant to the research practices and technologies discussed in this dissertation. Finally, we outline the directions for future work involving sentiment and stance visualization.

8.1 Research Findings

First of all, we are going to compare our findings to the objectives formulated in Section 1.2 and then focus on the main contributions (cf. Section 1.4).

8.1.1 Analysis of Research Objectives

O1 *Position the existing sentiment and stance visualization techniques in the wider context of text visualization*

Table 8.1: Our proposed approaches in the design space of text visualization techniques

Technique	Online Social Media	Communication	Patents	Reviews / (Medical) Reports	Literature/Poems	Scientific Articles / Papers	Editorial Media	Document	Corpora	Streams	Geospatial	Time series	Networks	Text Summarization / Topic Analysis / Entity Extraction	Discourse Analysis	Stance Analysis	Sentiment Analysis	Event Analysis	Trend Analysis / Pattern Analysis	Lexical/Syntactical Analysis	Relation/Connection Analysis	Translation / Text Alignment Analysis	Region of Interest	Clustering/Classification/ Categorization	Comparison	Overview	Monitoring	Navigation/Exploration	Uncertainty Tackling	2D	3D	Line Plot / River	Pixel/Area/Matrix	Node-Link	Clouds/Galaxies	Maps	Text	Glyph/Icon	Radial	Linear	Metric
uVSAT	●							●	●	○		●		●		○	●		●				●	●	●	●		●		●		●		●	●		●	●		●	●
ALVA	●								●			●	●			●			●	●	●			●	●	●		●		●		●	●	●	●		●	●		●	●
StanceVis Prime	●								●	○		●		●		●	●		●				●	●	●	●		●	●	●		●	●		●		●	●		●	●
StanceXplore	●								●		●	●		●		●			●					●	●	●		●	●	●		●	●			●	●	●		●	●
DoSVis				●	●		●	●						●		●								●	●	●		●	●	●					●		●	●		●	
Topics2Themes	●								●				●	●		●	○				●			●	●	●		●		●				●			●	●		●	

Note: Supported categories are marked by ●, partial support denoted by ○.

In Chapter 2, we briefly introduced the wide spectrum of problems studied within information visualization and visual analytics. It provides the motivation for the breadth of categorizations required to describe design spaces for particular fields within these disciplines, for instance, tree visualization [364] or temporal data visualization [4, 421]. Therefore, we designed a similarly broad and detailed categorization for the field of text visualization presented in Chapter 3 (more specifically, Section 3.1) and positioned more than 400 text visualization techniques in this design space after careful manual analysis of the corresponding publications. The results of this stage are available in the online survey browser that has been actively used by the research community during the past few years. Various analyses of the categorization results as well as the additional metadata, such as co-authorship of techniques/publications, provide us with a "big picture" of the state of the art in text visualization. Sentiment and stance visualization techniques are included in this design space, thus achieving research objective **O1**. The summary of our own visualization approaches within the text visualization categorization is provided in Table 8.1.

O2 *Define a design space for sentiment and stance visualization techniques*

As described in Chapter 2, the models of sentiment and stance can vary from a dichotomy of POSITIVITY/NEGATIVITY to multidimensional models of emotion and stance. The designers of the corresponding visualizations have to take this into account in addition to various possible data types, representation and interaction options, etc. To address this problem, we designed a categorization for sentiment visualization techniques (including stance visualization) based on the more general categorization described above and a manual analysis of more than 150 visualization techniques from peer-reviewed publications. The design space defined by this categorization includes the aspects related to (1) data domain (such as social media or literature), (2) data source type (e.g., individual documents or corpora/collections), (3) special data properties (for instance, temporal data), (4) analytic tasks (e.g., polarity or stance analysis), (5) visualization tasks (overview, comparison, etc.), (6) visual variable/channel for encoding sentiment (e.g., color or size), and (7) visual representation/metaphor for sentiment (a line plot, a glyph, etc.). We discuss the existing sentiment visualization techniques in the context of this categorization in Chapter 3 (more specifically, Section 3.2) and introduce a dedicated online survey browser. These results are complemented with additional analyses that provide an overview of the state of the art in sentiment visualization and indicate the interest for this research problem within InfoVis/VA and other disciplines. They also provide evidence of low support currently existing for the task of stance visualization, as discussed in Section 3.3, which motivates our efforts in order to address this research gap. Overall, these contributions have allowed us to achieve research objective **O2** and to position our own approaches in the same design space (see Table 8.2).

Table 8.2: Our proposed approaches in the design space of sentiment visualization techniques

Technique	Online Social Media	Communication	Reviews / (Medical) Reports	Literature/Poems	Scientific Articles / Papers	Editorial Media	Document	Corpora	Streams	Geospatial	Time Series	Networks	Polarity Analysis / Subjectivity Detection	Opinion Mining / Aspect-based Sentiment Analysis	Emotion/Affect Analysis	Stance Analysis	Region of Interest	Clustering/Classification/ Categorization	Comparison	Overview	Monitoring	Navigation/Exploration	Uncertainty Tackling	Color	Position/Orientation	Size	Shape	Texture/Pattern	Line Plot / River	Pixel/Area/Matrix	Node-Link	Clouds/Galaxies	Maps	Text	Glyph/Icon
uVSAT	●						●	●	○		●		●	●	●	○	●	●	●	●		●		●	●	●			●			●		●	●
ALVA	●							●			●	●				●		●	●	●		●		●	●			●	●	●	●	●		●	●
StanceVis Prime	●							●	○		●		●	●		●	●	●	●	●		●	●	●	●	●	●		●	●		●		●	●
StanceXplore	●							●		●	●			●		●		●	●	●		●	●	●		●			●	●			●	●	●
DoSVis			●	●		●	●									●			●	●		●	●	●	●		●					●		●	●
Topics2Themes	●							●				●	○	●		●		●	●	●		●		●		●					●			●	●

Note: Supported categories are marked by ●, partial support denoted by ○.

O3 *Enable the design and development of stance visualization techniques by facilitating the underlying research on stance analysis*

As established in Chapter 1, the research work described in this dissertation was undertaken as part of an interdisciplinary project on stance analysis that included domain experts in linguistics and computational linguistics as our collaborators. Therefore, some of our activities and contributions were directed at supporting the tasks related to (1) collection of textual data that was potentially interesting and useful for stance analysis, (2) annotation of stance categories in such data, and (3) implementation of an automatic stance classifier. In order to facilitate the early project efforts, we designed and implemented a visual analytics system, called uVSAT, which is described in Chapter 4. uVSAT supports analyses based on detection of markers (key terms) of multiple emotion categories as well as several stance categories (CERTAINTY and UNCERTAINTY). uVSAT was used by our collaborators to explore temporal text data collected from blogs and forums over time and to compile a data set of text documents, mainly from political blogs, for the subsequent stage of the project. To facilitate the tasks related to data annotation and training of a stance classifier with an active learning approach at the next stage, we designed and implemented a visual analytics solution, called ALVA, which is discussed in Chapter 5. Besides helping the users to manage the aforementioned processes via a graphical user interface, ALVA supports visual analyses related to (1) the annotated data, (2) the annotation process itself, and (3) the status of active learning. To address one of the core issues related to visual encoding of our main data type, annotations with multiple non-exclusive stance categories, we designed a novel visual representation, called CatCombos, that uses the spatial positions of annotation item groups to indicate the corresponding sets of categories. ALVA was used by our collaborators until the end of the active learning process, and then the resulting stance classifier could be used for application purposes. Thus, with the contributions of Chapters 4 and 5, we have attained research objective **O3**.

O4 *Instantiate the sentiment and stance visualization design space by implementing techniques for textual data from social media as well as other data domains and application scenarios*

The final objective set for this work is related to design and implementation of sentiment and stance visualization techniques as part of our research project. Besides supporting the tasks and the data provided by our collaborators, we used the design spaces and the insights about the previous approaches discussed in Chapter 3 to motivate our design choices and to address some of the research gaps existing within sentiment and stance visualization. We contributed multiple visualization approaches with the main focus on stance visualization, which are summarized within the contexts of our design spaces in Tables 8.1 and 8.2.

The first two of these approaches, uVSAT and ALVA (see Chapters 4 and 5, respectively), are discussed above. With regard to the categorization, uVSAT has only partial support for stance analysis as it is based mostly upon emotion categories. Its backend collects the data in an ongoing fashion, however, the visualization itself does not directly support streaming/dynamic data representation. The tasks of ongoing collection of temporal text data from social media and its visual analysis are also supported by our recent approach, called StanceVis Prime (see Chapter 6), however, in contrast to uVSAT, it makes use of multiple categories of stance. StanceVis Prime allows the users to (1) explore and compare multiple data series from several data sources, (2) identify regions of interest with large numbers of sentiment and stance occurrences, and (3) investigate the underlying sets of text posts with both close and distant reading supported. Our additional contributions demonstrate further applications of stance analysis and visualization. In Section 7.1, we discuss a visualization tool, called StanceXplore, which provides support for exploratory visual analysis of a data set of Twitter posts processed with one of our stance classifiers. Compared to other stance visualization approaches, the unique contribution of StanceXplore is related to its support of geospatial information present in the tweets. With DoSVis (see Section 7.2), we address other parts of our design space: a longer individual document as the data source with literature and reports as the data domains. Finally, in Section 7.3, we demonstrate how our visualization approach, called Topics2Themes (not explicitly discussed in this work in its original form [387]), can be extended and customized to support multiple sentiment and stance categories in order to facilitate interactive computer-assisted argument extraction from social media texts. To summarize, the contributions of Chapters 4–7 complement the existing techniques discussed in Section 3.3 with regard to the design space regions not addressed by the previous stance visualization techniques, hence allowing us to attain research objective **O4**.

Based on the accomplishment of all the objectives, we can state that the **goal** of this dissertation set in Section 1.2, *to define, categorize, and implement means for visual representation and visual analysis of sentiment and stance in textual data, in particular, for the data originating in social media*, has been achieved.

8.1.2 Summary of Scientific Contributions

We can now summarize the main scientific contributions of the work presented in this dissertation (cf. Section 1.4):

1. Formulation of the main research challenges for sentiment and stance visualization based on the analysis of the corresponding theoretical and computational models.

2. Analysis and categorization of the existing text visualization techniques, accompanied by a publicly available survey browser.

3. Analysis and categorization of the existing sentiment and stance visualization techniques in the dedicated design space, accompanied by a publicly available survey browser.

4. Analysis of user tasks related to visual analysis of stance phenomena based on the existing sentiment analysis approach, accompanied by design and implementation of the corresponding visual analytics approach, called uVSAT, which supports exploration of temporal text data and identification of candidate documents in order to collect a training data set for a stance classifier.

5. Analysis of user tasks related to visual analysis of annotated data for stance classification, accompanied by design and implementation of the corresponding visual analytics approach, called ALVA, which supports data annotation, visual analysis, and classifier training using active learning for multi-label stance classification.

6. Analysis of user tasks related to visual analysis of sentiment and stance in social media texts, accompanied by design and implementation of the corresponding visual analytics approach, called StanceVis Prime, which supports exploratory data analysis of sentiment and stance classification results for temporal text data from several social media sources.

7. Demonstration of further applications of sentiment and stance visualization to various tasks, data types, and data domains defined in our design space with several visualization tools (StanceXplore, DoSVis, and Topics2Themes).

8.2 Discussion

In this section, we focus on the discussion of various decisions made in this research work, its limitations, and overall ethical considerations related to applications of automatic analysis and visualization of sentiment and stance in social media data.

8.2.1 Design Decisions and Lessons Learned

The research tasks reported in this dissertation were carried out either directly as part of the interdisciplinary StaViCTA project or in close relation to it, and there are several aspects of the process and the outcomes to reflect upon here.

Collaboration with Domain Experts Establishing a common ground between the researchers from three disciplines (linguistics, computational linguistics, and visualization) early on and maintaining it throughout the course of collaboration was crucial for our interdisciplinary project. The importance of communication for successful joint work is discussed in the paper of Kirby and Meyer on multidisciplinary and interdisciplinary visualization collaborations [223] as well as in several recent reports about collaborations between experts in visualization and digital humanities [43, 176]. As an example of a background knowledge gap, Riehmann et al. [339] mention how their collaborator used to insist that information visualization methods *"obviously"* refer just to computer-assisted drawing, and the notion of visualization actually refers to something else entirely. In the case of our StaViCTA project, the activities started with an initial kickoff meeting including introductory lectures on the respective disciplines, which were also followed by mandatory assignments for PhD students to ensure common ground. Planning and discussion of (1) project tasks (including the tasks and designs for visualization approaches), (2) results, and (3) publication strategies took place during regular online (monthly or bimonthly) and onsite (two–three times per year) project meetings (in addition to regular digital communication). The collaboration was most educational with regard to experiencing the peculiarities of academic cultures existing in our disciplines. For instance, the perceived value and the publication process of journal articles vs conference papers are very different in linguistics and computer science. As an anecdote, our collaborators in linguistics once received a decision about a major revision required for their journal article submission, and the deadline was stated to be in *one year* after that date, which would be very unusual for visualization venues nowadays.

User- and Data-Driven Design Process As mentioned above, our design process for visualization and visual analytics approaches was driven by the collaboration with domain experts: we (1) elicited their requirements during the project meetings, (2) used the data and computational methods provided by them, and (3) incorporated their feedback in our designs and implementations. Van Wijk [440] describes potential collaboration issues that are related to the gaps between the visualization researchers and domain experts in the background knowledge and the interests/intents with regard to the resulting visualization approaches. The possible scenarios or models of collaboration aimed to resolve such issues include, for instance, (1) following a user-centered design approach, (2) focusing simply on the programming/engineering tasks while abandoning any research ambitions in visualization, or (3) choosing and addressing application-related problems based on the *curiosity* of the visualization researchers. Our work in StaViCTA has arguably followed a mixture of these approaches: (1) we used the requirements and the feedback of our collaborators to develop the visualization tools (subject to data and computational model availability), for instance, uVSAT (Chapter 4) and StanceVis Prime (Chapter 6), while (2) some of our efforts had to be directed towards data management and infrastructure, where visual methods were not even required and expected initially, e.g., the annotation interface and user management in ALVA (Chapter 5). With regard to (3) curiosity, we used our design spaces and survey results (Chapter 3), assessed the available computational methods and data, and asked ourselves the *"What if?.."* questions: for instance, the idea for DoSVis (Section 7.2) was conceived this way. As the implication of the design process, most of our visualization contributions and publications in StaViCTA fall into the category of application/design studies [366, 466]. Our designs mainly aimed to be useful for the domain experts rather than necessarily provide novel visual representation or interaction techniques (cf. the *"simple is good"* discussion by Russell [348]). For example, one of the initial intentions for the visualization team in the project was to address streaming data visualization [89, 97] task for stance analysis. However, the discussions with our collaborators in linguistics and the analysis of their requirements showed that the main interest of our users was to get an overview and explore the collected temporal text data in a consistent, reproducible way. The scenarios where real-time monitoring of streaming social media data is crucial, such as emergency situations [41, 55, 270], were not in the focus of our users' interests. Nevertheless, supporting stance visualization for such purposes is still an interesting and important future work prospect.

Online Survey Browsers While visualization and visual analysis of stance remains a highly specialized problem in the visualization community, the biggest impact of the work described in this dissertation was arguably made by the categorizations and surveys described in Chapter 3, available via the online survey browsers to the researchers and the general public alike. In particular, we have received a lot of positive feedback for TextVis Browser from the research community over the past years. By mid-February 2019, TextVis Browser has been visited by approximately 46,500 users from 160 countries since its release in summer 2014 (according to Google Analytics). We see such online browsers as an important tool for most research areas in general and intend to continue maintaining and updating our own browsers in the foreseeable future.

Teaching Applications Also relevant to the previous point, TextVis Browser has proven to be a useful tool for teaching InfoVis courses. We have used it ourselves at LNU, and we have also received notes about such a usage from several other universities. Finally, we have also used some of the visualization approaches discussed in this dissertation for our InfoVis courses at LNU. For example, several students with non-computer science background were once provided with the article on uVSAT [243] and the access to the actual tool with a task of critical review. According to their feedback, it inspired them with ideas for applications of temporal and text data visualization for their own disciplines.

8.2.2 Research Limitations and Validation Concerns

To open the discussion of concerns and limitations related to the methodology and the results described in this dissertation, a brief note about the role and position of information visualization and visual analytics as scientific disciplines [314] should be made. According to van Wijk [441], InfoVis can be viewed as science, technology, and art—there are arguments and evidence supporting all these aspects. Considering typical research work on individual representation or interaction techniques, for instance, the expected elements of the scientific method framework such as (1) observed phenomena, (2) hypotheses, and (3) experiments and collected data can be mapped to (1) the users' perception of data represented or modified visually, (2) suppositions of usability of such techniques, and (3) user studies involving quantitative and qualitative evaluation of effectiveness, efficiency, and/or other metrics. Fekete et al. [129] provide further arguments about compatibility of InfoVis with the scientific method and Karl Popper's epistemology system, while stating that InfoVis as a scientific discipline is unique: it focuses on general development of insights from the data rather than understanding a specific knowledge domain. In contrast to the experiment-driven workflow of natural sciences, multiple activities in InfoVis and VA are related to design and development of software systems, which is close to the design science process model [322] with steps such (1) as problem identification, (2) design and implementation of a solution, and (3) its validation.

Validation/evaluation of InfoVis and VA approaches still remains an open challenge in our community with its own dedicated events such as the BELIV workshop and the VAST Challenge. Purchase [331] describes guidelines for conducting controlled experiments and iterative system evaluations for the related discipline of human-computer interaction. Task-based user studies following similar guidelines are also usually desirable, and often required during the peer review process, for InfoVis and VA research publications. However, they are not always feasible due to time and resource constraints, task formulation issues, task interdependencies, etc. Detailed surveys, challenges, and calls for action with regard to evaluation in InfoVis and VA are provided, for instance, in the works by Lam et al. [248], Isenberg et al. [193], and Kosara [226]. Additionally, Munzner [306] describes a nested model for designing and validating visualization-related approaches ranging from specific algorithms and experiments to complete systems targeting the end users. The works by Stasko [399] and Wall et al. [446] provide additional pointers with regard to evaluation of the overall value added by visualization approaches for the end users.

This leads us to the discussion of limitations of our own work, including the validation aspects. As mentioned in Section 8.2, our primary research efforts related to specific sentiment and stance visualization approaches (rather than categorization models and meta-analyses) were driven by collaboration with domain experts and reported as application/design study publications. Thus, one limitation of our reported research is its dependency and potential bias with regard to the chosen domain-specific theoretical framework, computational models, and data. To mitigate this limitation, we typically employed criticism [227] internally during the design and development stages. We also discussed the problem-specific and generalizable aspects of our approaches in the corresponding publications, thus employing reflection [287]. The other important concern is related to the sufficiency of validation/evaluation activities carried out as part of our work, which did not include any formal user studies with quantitative measurements. Most of our contributions claimed in Section 8.1.2 fit the higher levels of Munzner's nested model [306], more specifically, *domain problem characterization* and *data/operation abstraction design*; evaluation methods suggested for these levels include interviews, observation of the target users, and collection of anecdotal evidence of utility. Weber et al. [466] mention participatory design practices, observational studies, and expert reviews among the possible evaluation approaches for application-driven research in visualization. In our project work, we employed the methods such as (1) interviews and participatory design with the domain experts, (2) use cases and case studies [366] involving real-world text data, and (3) critical discussion and expert reviews [426]. Most of our publications were also preceded by peer-reviewed poster papers, which provided us with numerous opportunities to receive constructive feedback [227] from the visualization experts in our community. Since the case studies and expert reviews involved ourselves (the visualization designers) and our collaborators, we acknowledge the existing threat to validity, as discussed for such design studies by Sedlmair et al. [366]. Further evaluations with independent experts or other end users should be carried out to provide stronger evidence; however, it is subject to availability of domain experts or users knowledgeable in sentiment and, especially, stance analysis. Some of our contributions also include novel representations and techniques, which correspond to the third level in Munzner's model [306]. For such specialized visual representations (e.g., CatCombos in ALVA introduced in Section 5.4.1) and semi-isolated groups of interactive representations (for instance, DTW views in StanceVis Prime discussed in Section 6.4.2), further task-based evaluation and comparison with baseline approaches should be carried out in the future to warrant their application for more general purposes rather than visualization of stance analysis results.

8.2.3 Ethical and Societal Aspects

As mentioned in the previous section, our research discussed in this dissertation did not involve any laboratory studies with human subjects. With the exception of our publicly available survey browsers, the interactions with the users of our tools mainly involved our project collaborators and colleagues. With regard to the tools and the data used in this research, for design and development tasks we used (1) publicly available software libraries, (2) research publications, (3) lists of marker words for sentiment and stance (see Chapter 4), and (4) stance annotations created by our collaborators (see Chapter 5). The data sources for our visualization tools included (1) aggregated data series and URLs to public blog and forum documents provided by our project partners from Gavagai AB, accompanied by the actual texts of these publicly available documents (see Chapter 4), some of which were later used for the stance annotation task (see Chapter 5); (2) publicly available business reports and works of fiction (see Section 7.2); and (3) social media texts and corresponding aggregated data series based on the publicly available data provided by Twitter and Reddit. The textual data was preprocessed, in most cases classified with regard to sentiment and stance, and visualized using our tools described in this dissertation.

Besides this discussion of how *our own* work collected and used the data, a discussion of the *general* concerns related to social media and data analytics should take place, motivated by such examples as multiple Facebook privacy scandals[1] and introduction of GDPR regulations in the European Union in 2018[2]. O'Leary [315] even argues for establishment of a code of conduct for "Big Data" analytics, including the imperatives not to injure other people and to try to improve their lives through such analytic activities. The issues related to the data misuse already include, for instance, automated discrimination of individuals or groups by data mining algorithms, as discussed by Carmichael et al. [65].

On the other hand, computational and visual data analytics methods, including sentiment and stance analysis, can also be leveraged for the challenges significant for the modern society, as we continue to witness the explosion of the amounts of digital data and the increasing role of artificial intelligence technologies [363]. For instance, the texts of public tweets posted by a user on Twitter could be used to predict the user's personality in order to customize the interface and provide a better user experience, as suggested by Golbeck et al. [153]. Hate speech detection is another important task for computational methods [151], which could be facilitated by sentiment and stance classification. Finally, the detection of *"fake news"* is also in the focus of work of many researchers in computational linguistics [66, 152, 159, 301]—sentiment analysis and visualization are already being applied for this task, for instance, by Harris [165], and it would only be natural to apply stance analysis accordingly.

8.3 Future Work

The work described in this dissertation has contributed to the research fields of sentiment and stance visualization, however, multiple open challenges still exist within and beyond these fields. Some of the more specific points regarding our own work and concrete visualization approaches were given above in the summaries of Chapters 3–7; below, we discuss the more general challenges.

Evaluation of Sentiment and Stance Visualizations While we have acknowledged the need for much more thorough validation of sentiment and stance

[1] `https://www.bbc.com/news/technology-46618582` (last accessed in February 2019)
[2] `https://www.bbc.com/news/technology-44239126` (last accessed in February 2019)

visualization approaches in Section 8.2.2, this issue applies not only to our own work, but most of the existing techniques, too. The work by Shamim et al. [371] provides some initial contributions for this challenge, however, further efforts are required for evaluating and reporting best practices for sentiment and stance visualization approaches with (1) individual representations, (2) combinations of coordinated multiple views, and (3) visual analytics systems involving complex computational models.

Research Gaps Identified in the Design Spaces By comparing the design space and the data about the existing sentiment and, especially, stance visualization techniques (see Chapter 3), we can identify multiple research gaps and scenarios for future work to support various (1) data domains, (2) data types, (3) user tasks, and (4) visual representations. Even more options are available if we consider the analytic tasks for text visualization which could benefit from involvement of sentiment and/or stance information: for example, digital humanities experts [198] could make use of visual analysis of multiple stance categories as part of their toolbox when exploring narratives and events in fiction and historical documents. Maintaining the existing design spaces and surveys up to date with emerging approaches and topics is a challenge on its own for our research community, which leads us to the next point.

Tighter Integration with Computational Models During the past few years, the attention of the visualization community has been attracted to the research topics related to machine learning models (especially, deep learning), explainable artificial intelligence, and so on. Several recent surveys [112, 120, 171, 264, 269] mentioned in Section 3.4 describe the state of the art in InfoVis and VA with regard to integration, visual analysis/interpretation, and adjustment of such models. The challenge of tighter integration of computational and visual models is also valid for sentiment and stance analysis techniques. Our own work on ALVA described in Chapter 5 is relevant to this problem, but further efforts are required to support (1) data management and annotation, (2) training computational models supporting sentiment and stance, and (3) visual analysis and interaction with such data and models. Some of the recent contributions related to these tasks include the approaches for facilitating the active learning process by Huang et al. [184] and interactive construction of lexicon-based concepts by Park et al. [320]; the VIAL approach for interactive classifier training by Bernard et al. [32, 342]; the approach for interactive analysis of text annotations by Baumann et al. [27]; and NLIZE by Liu et al. [266], a recent visual analytics approach for analysis and interpretation of natural language inference models.

Better Support for Analytical Workflows The previous challenge is undoubtedly related to visual analytics; at the same time, the level of direct support for certain analytical tasks in visual sentiment and stance analysis systems should be increased. For instance, in Chapter 2 we mentioned the sensemaking process model for intelligence analysis by Pirolli and Card [325] which includes steps such as (1) formulating hypotheses, (2) linking them to the evidence in the data, and (3) presenting/disseminating the results of the case. Andrienko et al. [15] describe the process of visual analytics as (1) constructing a behavioral model for the given problem and data, (2) using the model to answer questions and gain knowledge, and, finally, (3) externalizing the model as well as the model provenance [219]. From the point of view of user interaction and visualization, support for such workflows should include explicit graphical means for (1) formulating and testing hypotheses, (2) compiling and exporting the results of analyses, and (3) working in a reproducible manner with regard to preserving sessions, intermediate results, and navigation/interaction history (e.g., see Chapter 4 for the description of history diagrams and query links in uVSAT, which address this provenance task). Support for (4) collaborative work [192] should also be improved. While the existing general-purpose analytical tools targeted at professional data analysts [474, 475] support some of such tasks, the approaches for visualization and visual analysis of sentiment and stance should aim to address them. This can help to increase the value and usability of such approaches for the target users (for instance, researchers in the humanities) and to generally improve the visibility of this field for the users from academia, industry, and other areas.

Integration with/into Other Approaches and Tools Finally, we envision further applications of sentiment and stance visualization that could benefit the existing and future tools, both within and beyond the approaches developed by the visualization community, for instance, the tools for close and distant reading used within the digital humanities community [197, 198]. One could imagine a PDF document viewer with sparklines [30, 432], glyphs [40], or scented widgets [471] that encode the detected sentiment and stance occurrences, similar to our standalone DoSVis tool described in Section 7.2.4. Another possibility is to support such analyses with multimedia analytics [79] approaches involving texts accompanying video or audio data: the existing works of Diakopoulos et al. [103] and Hohman et al. [179], for instance, could provide inspiration for future efforts in this direction.

As a summary, given the importance of social media as a public communication channel and the crucial role of language and text for our society in general, we can predict that computational and visual analyses of various aspects of subjectivity in textual data will remain an important topic in research and practical applications in the foreseeable future.

Bibliography

[1] Ahmed Abbasi and Hsinchun Chen. Categorization and analysis of text in computer mediated communication archives using visualization. In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '07, pages 11–18. ACM, 2007. doi:10.1145/1255175.1255178.

[2] Brett Adams, Dinh Phung, and Svetha Venkatesh. Eventscapes: Visualizing events over time with emotive facets. In *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, pages 1477–1480. ACM, 2011. doi:10.1145/2072298.2072044.

[3] Rebecca Agheyisi and Joshua A. Fishman. Language attitude studies: A brief survey of methodological approaches. *Anthropological Linguistics*, 12(5):137–157, 1970.

[4] Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. *Visualization of Time-Oriented Data*. Springer, 2011. doi:10.1007/978-0-85729-079-3.

[5] Conrad Albrecht-Buehler, Benjamin Watson, and David Ayman Shamma. Visualizing live text streams using motion and temporal pooling. *IEEE Computer Graphics and Applications*, 25(3):52–59, May 2005. doi:10.1109/MCG.2005.70.

[6] Aretha B. Alencar, Katy Börner, Fernando V. Paulovich, and Maria Cristina F. de Oliveira. Time-aware visualization of document collections. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, SAC '12, pages 997–1004. ACM, 2012. doi:10.1145/2245276.2245469.

[7] Aretha B. Alencar, Maria Cristina F. de Oliveira, and Fernando V. Paulovich. Seeing beyond reading: A survey on visual text analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6):476–492, 2012. doi:10.1002/widm.1071.

[8] Eric Alexander, Joe Kohlmann, Robin Valenza, Michael Witmore, and Michael Gleicher. Serendip: Topic model-driven visual exploration of text corpora. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '14, pages 173–182. IEEE, 2014. doi:10.1109/VAST.2014.7042493.

[9] Mohammad Alharbi and Robert S. Laramee. SoS TextVis: An extended survey of surveys on text visualization. *Computers*, 8(1), 2019. doi:10.3390/computers8010017.

[10] Bandar Alhumaidi A. Almutairi. Visualizing patterns of appraisal in texts and corpora. *Text & Talk*, 33(4–5):691–723, 2013. doi:10.1515/text-2013-0031.

[11] Basak Alper, Huahai Yang, Eben Haber, and Eser Kandogan. OpinionBlocks: Visualizing consumer reviews. In *Proceedings of the IEEE Workshop on Interactive Visual Text Analytics for Decision Making*, TextVis '11, 2011. URL: http://vialab.science.uoit.ca/textvis2011/.

[12] Sara Alspaugh, Nava Zokaei, Andrea Liu, Cindy Jin, and Marti A. Hearst. Futzing and moseying: Interviews with professional data analysts on exploration practices. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):22–31, January 2019. doi:10.1109/TVCG.2018.2865040.

[13] Enrique Amigó, Jorge Carrillo De Albornoz, Irina Chugur, Adolfo Corujo, Julio Gonzalo, Tamara Martín, Edgar Meij, Maarten De Rijke, and Damiano Spina. Overview of RepLab 2013: Evaluating online reputation monitoring systems. In *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, volume 8138 of *Lecture Notes in Computer Science*, pages 333–352. Springer Berlin Heidelberg, 2013. doi:10.1007/978-3-642-40802-1_31.

[14] Keith Andrews, Wolfgang Kienreich, Vedran Sabol, Jutta Becker, Georg Droschl, Frank Kappe, Michael Granitzer, Peter Auer, and Klaus Tochtermann. The InfoSky visual explorer: Exploiting hierarchical structure and document similarities. *Information Visualization*, 1(3-4):166–181, 2002. doi:10.1057/palgrave.ivs.9500023.

[15] Natalia Andrienko, Tim Lammarsch, Gennady Andrienko, Georg Fuchs, Daniel A. Keim, Silvia Miksch, and Alexander Rind. Viewing visual analytics as model building. *Computer Graphics Forum*, 37(6):275–299, 2018. doi:10.1111/cgf.13324.

[16] Daniel Angus, Andrew Smith, and Janet Wiles. Conceptual recurrence plots: Revealing patterns in human discourse. *IEEE Transactions on Visualization and Computer Graphics*, 18(6):988–997, June 2012. doi:10.1109/TVCG.2011.100.

[17] Michelle Annett and Grzegorz Kondrak. A comparison of sentiment analysis techniques: Polarizing movie blogs. In Sabine Bergler, editor, *Advances in Artificial Intelligence*, volume 5032 of *Lecture Notes in Computer Science*, pages 25–35. Springer Berlin Heidelberg, 2008. doi:10.1007/978-3-540-68825-9_3.

[18] Ron Artstein and Massimo Poesio. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596, December 2008. doi:10.1162/coli.07-034-R2.

[19] Bharathi Asokarajan, Ronak Etemadpour, June Abbas, Sam Huskey, and Chris Weaver. Visualization of Latin textual variants using a pixel-based text analysis tool. In *Proceedings of the EuroVis Workshop on Visual Analytics*, EuroVA '16. The Eurographics Association, 2016. doi:10.2312/eurova.20161119.

[20] Bharathi Asokarajan, Ronak Etemadpour, June Abbas, Sam Huskey, and Chris Weaver. TexTile: A pixel-based focus+context tool for analyzing variants across multiple text scales. In *Short Papers of the EG/VGTC Conference on Visualization*, EuroVis '17. The Eurographics Association, 2017. doi:10.2312/eurovisshort.20171132.

[21] Angeliki Athanasiadou and Elżbieta Tabakowska, editors. *Speaking of Emotions: Conceptualisation and Expression*. Cognitive Linguistics Research. Mouton de Gruyter, 1998.

[22] Moshe Azar. Argumentative text as rhetorical structure: An application of rhetorical structure theory. *Argumentation*, 13(1):97–114, February 1999. doi:10.1023/A:1007794409860.

[23] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation*, volume 10 of *LREC '10*, pages 2200–2204. ELRA, 2010. URL: http://lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf.

[24] Nadeem Bader, Osnat Mokryn, and Joel Lanir. Exploring emotions in online movie reviews for online browsing. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, IUI '17 Companion, pages 35–38. ACM, 2017. doi:10.1145/3030024.3040982.

[25] R. C. Balabantaray, Mudasir Mohammad, and Nibha Sharma. Multi-class Twitter emotion classification: A new approach. *International Journal of Applied Information Systems*, 4(1):48–53, September 2012. doi:10.5120/ijais12-450651.

[26] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media*, ICWSM '09, pages 361–362. AAAI Press, 2009. URL: https://aaai.org/ocs/index.php/ICWSM/09/paper/view/154.

[27] Martin Baumann, Steffen Koch, Harutyun Minasyan, and Thomas Ertl. Zooming on tokens: Seamless display modes for annotation analysis. In *Poster Abstracts of IEEE VIS*, 2018.

[28] Eric P. S. Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K. Gay. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*, 68(6):1397–1410, June 2017. doi:10.1002/asi.23786.

[29] Fabian Beck, Sebastian Koch, and Daniel Weiskopf. Visual analysis and dissemination of scientific literature collections with SurVis. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):180–189, January 2016. doi:10.1109/TVCG.2015.2467757.

[30] Fabian Beck and Daniel Weiskopf. Word-sized graphics for scientific texts. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1576–1587, June 2017. doi:10.1109/TVCG.2017.2674958.

[31] Robert Bembenik and Piotr Andruszkiewicz. Towards automatic argument extraction and visualization in a deliberative model of online consultations for local governments. In Jaroslav Pokorný, Mirjana Ivanović, Bernhard Thalheim, and Petr Šaloun, editors, *Proceedings of the 20th East European Conference on Advances in Databases and Information Systems (ADBIS 2016)*, volume 9809 of *Lecture Notes in Computer Science*, pages 74–86. Springer International Publishing, 2016. doi:10.1007/978-3-319-44039-2_6.

[32] Jürgen Bernard, Matthias Zeppelzauer, Michael Sedlmair, and Wolfgang Aigner. VIAL: A unified process for visual interactive labeling. *The Visual Computer*, 34(9):1189–1207, September 2018. doi:10.1007/s00371-018-1500-3.

[33] Jürgen Bernard, Nils Wilhelm, Maximilian Scherer, Thorsten May, and Tobias Schreck. TimeSeriesPaths: Projection-based explorative analysis of multivariate time series data. *Journal of WSCG*, 20(2):97–106, 2012. URL: http://wscg.zcu.cz/WSCG2012/!_2012-Journal-Full-2.pdf.

[34] Donald J. Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *Proceedings of the AAAI Workshop on Knowledge Discovery in Databases*, KDD '94, pages 359–370. AAAI Press, 1994. URL: http://aaai.org/Library/Workshops/1994/ws94-03-031.php.

[35] Jacques Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, 1983. Translated by W. J. Berg.

[36] Douglas Biber. Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2):97–116, 2006. doi:10.1016/j.jeap.2006.05.001.

[37] Douglas Biber and Edward Finegan. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text — Interdisciplinary Journal for the Study of Discourse*, 9(1):93–124, 1989. doi:10.1515/text.1.1989.9.1.93.

[38] Bootstrap, 2011. Accessed February 4, 2019. URL: https://getbootstrap.com/.

[39] Ingwer Borg and Patrick J. F. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005. doi:10.1007/0-387-28981-X.

[40] Rita Borgo, Johannes Kehrer, David H. S. Chung, Eamonn Maguire, Robert S. Laramee, Helwig Hauser, Matthew Ward, and Min Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics 2013 — State of the Art Reports*. The Eurographics Association, 2013. doi:10.2312/conf/EG2013/stars/039-063.

[41] Harald Bosch, Dennis Thom, Florian Heimerl, Edward Püttmann, Steffen Koch, Robert Krüger, Michael Wörner, and Thomas Ertl. ScatterBlogs2: Real-time monitoring of microblog messages through user-guided filtering. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2022–2031, December 2013. doi:10.1109/TVCG.2013.186.

[42] Ameni Boumaiza. A survey on sentiment analysis and visualization. *Journal of Emerging Technologies in Web Intelligence*, 7(1):35–43, November 2015. doi:10.12720/jetwi.7.1.35-43.

[43] Adam James Bradley, Mennatallah El-Assady, Katharine Coles, Eric Alexander, Min Chen, Christopher Collins, Stefan Jänicke, and David Joseph Wrisley. Visualization and the digital humanities: Moving toward stronger collaborations. *IEEE Computer Graphics and Applications*, 38(6):26–38, November 2018. doi:10.1109/MCG.2018.2878900.

[44] Ulrik Brandes and Steven R. Corman. Visual unrolling of network evolution and the analysis of dynamic discourse. *Information Visualization*, 2(1):40–50, 2003. doi:10.1057/palgrave.ivs.9500037.

[45] Adrian M.P. Braşoveanu, Alexander Hubmann-Haidvogel, and Arno Scharl. Interactive visualization of emerging topics in multiple social media streams. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 530–533. ACM, 2012. doi:10.1145/2254556.2254655.

[46] Richard Brath and Ebad Banissi. Using text in visualizations for micro/macro readings. In *Proceedings of the IUI Workshop on Visual Text Analytics*, TextVis '15, 2015. URL: http://vialab.science.uoit.ca/textvis2015/.

[47] Richard Brath and Ebad Banissi. Typographic sets: Labeled set elements with font attributes. In *Proceedings of the International Workshop on Set Visualization and Reasoning*, SetVR 2016, pages 29–43. CEUR-WS.org, 2016. URL: http://ceur-ws.org/Vol-1655.

[48] Matthew Brehmer and Tamara Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, December 2013. doi:10.1109/TVCG.2013.124.

[49] Anthony Brew, Derek Greene, Daniel Archambault, and Pádraig Cunningham. Deriving insights from national happiness indices. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops*, ICDMW '11, pages 53–60. IEEE, 2011. doi:10.1109/ICDMW.2011.61.

[50] Ken W. Brodlie, Lesley Ann Carpenter, Rae A. Earnshaw, Julian R. Gallop, Roger J. Hubbold, Anne M. Mumford, Chris D. Osland, and Peter Quarendon, editors. *Scientific Visualization: Techniques and Applications*. Springer-Verlag, 1992.

[51] Samuel Brody and Noemie Elhadad. An unsupervised aspect-sentiment model for online reviews. In *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 804–812. ACL, 2010. URL: http://aclweb.org/anthology/N10-1122.

[52] Michael Brooks, Katie Kuksenok, Megan K. Torkildson, Daniel Perry, John J. Robinson, Taylor J. Scott, Ona Anicello, Ariana Zukowski, Paul Harris, and Cecilia R. Aragon. Statistical affect detection in collaborative chat. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 317–328. ACM, 2013. doi:10.1145/2441776.2441813.

[53] Michael Brooks, John J. Robinson, Megan K. Torkildson, Sungsoo (Ray) Hong, and Cecilia R. Aragon. Collaborative visual analysis of sentiment in Twitter events. In Yuhua Luo, editor, *Proceedings of the 11th International Conference on Cooperative Design, Visualization, and Engineering (CDVE '14)*, volume 8683 of *Lecture Notes in Computer Science*, pages 1–8. Springer International Publishing, 2014. doi:10.1007/978-3-319-10831-5_1.

[54] Lee Byron and Martin Wattenberg. Stacked graphs — Geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1245–1252, November 2008. doi:10.1109/TVCG.2008.166.

[55] Nadya A. Calderon, Richard Arias-Hernandez, and Brian Fisher. Studying animation for real-time visual analytics: A design study of social media analytics in emergency management. In *Proceedings of the 47th Hawaii International Conference on System Sciences*, HICSS '14, pages 1364–1373. IEEE, 2014. doi:10.1109/HICSS.2014.176.

[56] Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI '14, pages 1515–1521. AAAI Press, 2014. URL: http://dl.acm.org/citation.cfm?id=2892753.2892763.

[57] Erik Cambria, Björn Schuller, Yunqing Xia, and Catherine Havasi. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2):15–21, March 2013. doi:10.1109/MIS.2013.30.

[58] Nan Cao and Weiwei Cui. *Introduction to Text Visualization*, volume 1 of *Atlantis Briefs in Artificial Intelligence*. Atlantis Press, 2016. doi:10.2991/978-94-6239-186-4.

[59] Nan Cao, Yu-Ru Lin, Xiaohua Sun, D. Lazer, Shixia Liu, and Huamin Qu. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2649–2658, December 2012. `doi:10.1109/TVCG.2012.291`.

[60] Nan Cao, Lu Lu, Yu-Ru Lin, Fei Wang, and Zhen Wen. SocialHelix: Visual analysis of sentiment divergence in social media. *Journal of Visualization*, 18(2):221–235, 2015. `doi:10.1007/s12650-014-0246-x`.

[61] Nan Cao, Conglei Shi, Sabrina Lin, Jie Lu, Yu-Ru Lin, and Ching-Yung Lin. TargetVue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):280–289, January 2016. `doi:10.1109/TVCG.2015.2467196`.

[62] Nan Cao, Jimeng Sun, Yu-Ru Lin, David Gotz, Shixia Liu, and Huamin Qu. FacetAtlas: Multifaceted visualization for rich text corpora. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1172–1181, November 2010. `doi:10.1109/TVCG.2010.154`.

[63] Cornelia Caragea, Anna Cinzia Squicciarini, Sam Stehle, Kishore Neppalli, and Andrea H. Tapia. Mapping moods: Geo-mapped sentiment analysis during Hurricane Sandy. In *Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management*, ISCRAM '14, pages 642–651. The Pennsylvania State University, 2014. URL: `http://iscram.org/legacy/ISCRAM2014/papers/p29.pdf`.

[64] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, 1999. URL: `https://dl.acm.org/citation.cfm?id=300679`.

[65] Laura Carmichael, Sophie Stalla-Bourdillon, and Steffen Staab. Data mining and automated discrimination: A mixed legal/technical perspective. *IEEE Intelligent Systems*, 31(6):51–55, November 2016. `doi:10.1109/MIS.2016.96`.

[66] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on Twitter. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 675–684. ACM, 2011. `doi:10.1145/1963405.1963500`.

[67] Daniel Cernea and Andreas Kerren. A survey of technologies on the rise for emotion-enhanced interaction. *Journal of Visual Languages and Computing*, 31, Part A:70–86, 2015. `doi:10.1016/j.jvlc.2015.10.001`.

[68] Daniel Cernea, Igor Truderung, Andreas Kerren, and Achim Ebert. Web-Comets: A tab-oriented approach for browser history visualization. In *Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications — Volume 1: IVAPP*, VISIGRAPP '13, pages 439–450. SciTePress, 2013. `doi:10.5220/0004226004390450`.

[69] Ofelia Cervantes, Francisco Gutiérrez, Ernesto Gutiérrez, Esteban Castillo, J. Alfredo Sánchez, and Wanggen Wan. Expression: Visualizing affective content from social streams. In *Proceedings of the Latin American Conference on Human Computer Interaction*, CLIHC '15, pages 10:1–10:8. ACM, 2015. `doi:10.1145/2824893.2824903`.

[70] Jon Chamberlain, Udo Kruschwitz, and Orland Hoeber. Scalable visualisation of sentiment and stance. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, LREC 2018, pages 4181–4182. ELRA, 2018. URL: `http://lrec-conf.org/proceedings/lrec2018/summaries/760.html`.

[71] Senthil Chandrasegaran, Sriram Karthik Badam, Lorraine Kisselburgh, Karthik Ramani, and Niklas Elmqvist. Integrating visual analytics support for grounded theory practice in qualitative text analysis. *Computer Graphics Forum*, 36(3):201–212, 2017. URL: `http://dx.doi.org/10.1111/cgf.13180`, `doi:10.1111/cgf.13180`.

[72] Annie T. Chen, Jessica Martell, and Pamella Lach. Supporting discovery through contextual representation: Developing a tool for visually exploring slave narratives. In *Proceedings of the Workshop on Visualization for the Digital Humanities*, VIS4DH '16, 2016. URL: `http://vis4dh.dbvis.de/papers/2016/`.

[73] Chaomei Chen, Fidelia Ibekwe-Sanjuan, Eric SanJuan, and Chris Weaver. Visual analysis of conflicting opinions. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, VAST '06, pages 59–66. IEEE, 2006. `doi:10.1109/VAST.2006.261431`.

[74] Nan-Chen Chen, Michael Brooks, Rafal Kocielnik, Sungsoo Ray Hong, Jeff Smith, Sanny Lin, Zening Qu, and Cecilia Aragon. Lariat: A visual analytics tool for social media researchers to explore Twitter datasets. In *Proceedings of the 50th Hawaii International Conference on System Sciences*, HICSS '17, pages 1881–1890, 2017. `doi:10.24251/HICSS.2017.228`.

[75] Nan-Chen Chen, Laurie Beth Feldman, Judith F. Kroll, and Cecilia R. Aragon. Emoticons and linguistic alignment: How visual analytics can elicit storytelling. In *Poster Abstracts of the IEEE Conference on Visual Analytics Science and Technology*, VAST '14, pages 237–238. IEEE, 2014. `doi:10.1109/VAST.2014.7042508`.

[76] Siming Chen, Lijing Lin, and Xiaoru Yuan. Social media visual analytics. *Computer Graphics Forum*, 36(3):563–587, June 2017. `doi:10.1111/cgf.13211`.

[77] Yu-Sheng Chen, Lieu-Hen Chen, and Yasufumi Takama. Proposal of LDA-based sentiment visualization of hotel reviews. In *Proceedings of the 2015 IEEE International Conference on Data Mining Workshops*, ICDMW 2015, pages 687–693. IEEE, 2015. `doi:10.1109/ICDMW.2015.72`.

[78] Yu-Sheng Chen, Lieu-Hen Chen, Toru Yamaguchi, and Yasufumi Takama. Visualization system for analyzing user opinion. In *Proceedings of the IEEE/SICE International Symposium on System Integration*, SII 2015, pages 646–649. IEEE, 2015. `doi:10.1109/SII.2015.7405055`.

[79] Nancy A. Chinchor, James J. Thomas, Pak Chung Wong, Michael G. Christel, and William Ribarsky. Multimedia analysis + visual analytics = multimedia analytics. *IEEE Computer Graphics and Applications*, 30(5):52–60, September 2010. `doi:10.1109/MCG.2010.92`.

[80] Isaac Cho, Ryan Wesslen, Svitlana Volkova, Bill Ribarsky, and Wenwen Dou. CrystalBall: A visual analytic system for future event discovery and analysis from social media data. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '17, pages 25–35, 2017. `doi:10.1109/VAST.2017.8585658`.

[81] William B. Claster, Malcolm Cooper, and Philip Sallis. Thailand — Tourism and conflict: Modeling sentiment from Twitter tweets using Naïve Bayes and unsupervised artificial neural nets. In *Proceedings of the Second International Conference on Computational Intelligence, Modelling and Simulation*, CIMSim '10, pages 89–94. IEEE, 2010. `doi:10.1109/CIMSiM.2010.98`.

[82] Lawrence H. Cohen. Measurement of life events. In *Life Events and Psychological Functioning: Theoretical and Methodological Issues*, pages 11–30. SAGE Publications, 1988.

[83] Christopher Collins, Sheelagh Carpendale, and Gerald Penn. DocuBurst: Visualizing document content using language structure. *Computer Graphics Forum*, 28(3):1039–1046, 2009. `doi:10.1111/j.1467-8659.2009.01439.x`.

[84] Christopher Collins, Fernanda B. Viégas, and Martin Wattenberg. Parallel tag clouds to explore and analyze facted text corpora. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, VAST '09, pages 91–98. IEEE, 2009. `doi:10.1109/VAST.2009.5333443`.

[85] ColorBrewer 2.0 — Color advice for cartography, 2009. Accessed February 4, 2019. URL: `http://colorbrewer2.org/`.

[86] The Compass DeRose guide to emotion words, 2005. Accessed February 4, 2019. URL: `http://derose.net/steve/resources/emotionwords/ewords.html`.

[87] Alexandru Constantin, Steve Pettifer, and Andrei Voronkov. PDFX: Fully-automated PDF-to-XML conversion of scientific literature. In *Proceedings of the ACM Symposium on Document Engineering*, DocEng '13, pages 177–180. ACM, 2013. `doi:10.1145/2494266.2494271`.

[88] Michael Correll, Michael Witmore, and Michael Gleicher. Exploring collections of tagged text for literary scholarship. *Computer Graphics Forum*, 30(3):731–740, 2011. `doi:10.1111/j.1467-8659.2011.01922.x`.

[89] Joseph A. Cottam, Andrew Lumsdaine, and Chris Weaver. Watch this: A taxonomy for dynamic data visualization. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '12, pages 193–202. IEEE, 2012. `doi:10.1109/VAST.2012.6400552`.

[90] Stefano Cresci, Andrea D'Errico, Davide Gazzé, Angelica Lo Duca, Andrea Marchetti, and Maurizio Tesconi. Tour-pedia: A web application for sentiment visualization in tourism domain. In *Proceedings of the LREC 2014 OpeNER Workshop*, pages 18–21, 2014.

[91] Tarik Crnovrsanin, Chris Muelder, Carlos Correa, and Kwan-Liu Ma. Proximity-based visualization of movement trace data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, VAST '09, pages 11–18, 2009. `doi:10.1109/VAST.2009.5332593`.

[92] Eric Cuenca, Arnaud Sallaberry, Florence Ying Wang, and Pascal Poncelet. MultiStream: A multiresolution streamgraph approach to explore hierarchical time series. *IEEE Transactions on Visualization and Computer Graphics*, 24(12):3160–3173, December 2018. `doi:10.1109/TVCG.2018.2796591`.

[93] Weiwei Cui, Huamin Qu, Hong Zhou, Wenbin Zhang, and Steve Skiena. Watch the story unfold with TextWheel: Visualization of large-scale news streams. *ACM Transactions on Intelligent Systems and Technology*, 3(2):20:1–20:17, February 2012. `doi:10.1145/2089094.2089096`.

[94] D3 — Data-driven documents, 2011. Accessed February 4, 2019. URL: `http://d3js.org/`.

[95] Xiangfeng Dai and Robert Prout. Unlocking Super Bowl insights: Weighted word embeddings for Twitter sentiment classification. In *Proceedings of the 3rd Multidisciplinary International Social Networks Conference, SocialInformatics, Data Science*, MISNC, SI, DS 2016, pages 12:1–12:6. ACM, 2016. `doi:10.1145/2955129.2955148`.

[96] Amitava Das, Sivaji Bandyopadhyay, and Björn Gambäck. Sentiment analysis: What is the end user's requirement? In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, WIMS '12, pages 35:1–35:10. ACM, 2012. `doi:10.1145/2254129.2254173`.

[97] Aritra Dasgupta, Dustin L. Arendt, Lyndsey R. Franklin, Pak Chung Wong, and Kristin A. Cook. Human factors in streaming data analysis: Challenges and opportunities for information visualization. *Computer Graphics Forum*, 37(1):254–272, 2018. `doi:10.1111/cgf.13264`.

[98] Maria Cristina Ferreira de Oliveira and Haim Levkowitz. From visual data exploration to visual data mining: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 9(3):378–394, July 2003. `doi:10.1109/TVCG.2003.1207445`.

[99] Philip DeCamp, Amber Frid-Jimenez, Jethran Guiness, and Deb Roy. Gist icons: Seeing meaning in large bodies of literature. In *Poster Abstracts of the IEEE Symposium on Information Visualization*, InfoVis '05, 2005.

[100] Thomas A. DeFanti, Maxine D. Brown, and Bruce H. McCormick. Visualization: Expanding scientific and engineering research opportunities. *Computer*, 22(8):12–25, August 1989. `doi:10.1109/2.35195`.

[101] Vasundhara Dehiya and Klaus Mueller. Analyzing Hillary Clinton's emails. In *Poster Abstracts of IEEE VIS*, 2016.

[102] Luigi Di Caro and Matteo Grella. Sentiment analysis via dependency parsing. *Computer Standards & Interfaces*, 35(5):442–453, 2013. `doi:10.1016/j.csi.2012.10.005`.

[103] Nicholas Diakopoulos, Mor Naaman, and Funda Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, VAST '10, pages 115–122. IEEE, 2010. `doi:10.1109/VAST.2010.5652922`.

[104] Nicholas Diakopoulos, Amy X Zhang, Dag Elgesem, and Andrew Salway. Identifying and analyzing moral evaluation frames in climate change blog discourse. In *Proceedings of the International Conference on Weblogs and Social Media*, ICWSM '14, pages 583–586. AAAI Press, 2014. URL: `http://aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8064`.

[105] Wenwen Dou and Shixia Liu. Topic- and time-oriented visual text analysis. *IEEE Computer Graphics and Applications*, 36(4):8–13, July 2016. `doi:10.1109/MCG.2016.73`.

[106] Wenwen Dou, Xiaoyu Wang, Drew Skau, William Ribarsky, and Michelle X. Zhou. LeadLine: Interactive visual analysis of text data through event identification and exploration. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '12, pages 93–102. IEEE, 2012. `doi:10.1109/VAST.2012.6400485`.

[107] Wenwen Dou, Li Yu, Xiaoyu Wang, Zhiqiang Ma, and W. Ribarsky. HierarchicalTopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2002–2011, December 2013. `doi:10.1109/TVCG.2013.162`.

[108] Johanna Drucker. Graphical approaches to the digital humanities. In Susan Schreibman, Ray Siemens, and John Unsworth, editors, *A New Companion to Digital Humanities*, pages 238–250. John Wiley & Sons, 2016.

[109] Xiaolin Du, Yunming Ye, Raymond Y.K. Lau, and Yueping Li. OpinionRings: Inferring and visualizing the opinion tendency of socially connected users. *Decision Support Systems*, 75:11–24, 2015. `doi:10.1016/j.dss.2015.04.007`.

[110] John W. Du Bois. The stance triangle. In Robert Englebretson, editor, *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, volume 164 of *Pragmatics & Beyond New Series*, pages 139–182. John Benjamins, 2007. `doi:10.1075/pbns.164`.

[111] Dongxu Duan, Weihong Qian, Shimei Pan, Lei Shi, and Chuang Lin. VISA: A visual sentiment analysis system. In *Proceedings of the 5th International Symposium on Visual Information Communication and Interaction*, VINCI '12, pages 22–28. ACM, 2012. `doi:10.1145/2397696.2397700`.

[112] John J. Dudley and Per Ola Kristensson. A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems*, 8(2):8:1–8:37, June 2018. `doi:10.1145/3185517`.

[113] Stephen G. Eick, Joseph L. Steffen, and Eric E. Sumner. Seesoft — A tool for visualizing line oriented software statistics. *IEEE Transactions on Software Engineering*, 18(11):957–968, November 1992. `doi:10.1109/32.177365`.

[114] Paul Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3–4):169–200, May 1992.

[115] Mennatallah El-Assady, Valentin Gold, Carmela Acevedo, Christopher Collins, and Daniel A. Keim. ConToVi: Multi-party conversation exploration using topic-space views. *Computer Graphics Forum*, 35(3):431–440, 2016. `doi:10.1111/cgf.12919`.

[116] Mennatallah El-Assady, Rita Sevastjanova, Bela Gipp, Daniel A. Keim, and Christopher Collins. NEREx: Named-entity relationship exploration in multi-party conversations. *Computer Graphics Forum*, 36(3):213–225, 2017. `doi:10.1111/cgf.13181`.

[117] Mennatallah El-Assady, Rita Sevastjanova, Daniel Keim, and Christopher Collins. Thread-Reconstructor: Modeling reply-chains to untangle conversational text through visual analytics. *Computer Graphics Forum*, 37(3):351–365, 2018. `doi:10.1111/cgf.13425`.

[118] Geoffrey Ellis and Alan Dix. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1216–1223, November 2007. `doi:10.1109/TVCG.2007.70535`.

[119] Emotion detection and recognition market — Global forecast to 2021, December 2016. Accessed February 4, 2019. URL: `http://www.researchandmarkets.com/research/kkrrgz/emotion_detection`.

[120] Alex Endert, William Ribarsky, Cagatay Turkay, B.L. William Wong, Ian Nabney, Ignacio Díaz Blanco, and Fabrice Rossi. The state of the art in integrating machine learning into visual analytics. *Computer Graphics Forum*, 36(8):458–486, 2017. doi:10.1111/cgf.13092.

[121] Robert Englebretson, editor. *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, volume 164 of *Pragmatics & Beyond New Series*. John Benjamins, 2007. doi:10.1075/pbns.164.

[122] Philippe Esling and Carlos Agon. Time-series data mining. *ACM Computing Surveys*, 45(1):12:1–12:34, December 2012. doi:10.1145/2379776.2379788.

[123] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD '96, pages 226–231. AAAI Press, 1996. URL: http://aaai.org/Papers/KDD/1996/KDD96-037.pdf.

[124] Andrea Esuli and Fabrizio Sebastiani. SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation*, LREC '06, pages 417–422. ELRA, 2006. URL: http://lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.

[125] James D. Evans. *Straightforward Statistics for the Behavioral Sciences*. Brooks/Cole Publishing, 1996.

[126] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. Opinion Space: A scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1175–1184. ACM, 2010. doi:10.1145/1753326.1753502.

[127] Paolo Federico, Florian Heimerl, Steffen Koch, and Silvia Miksch. A survey on visual approaches for analyzing scientific literature and patents. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2179–2198, September 2017. doi:10.1109/TVCG.2016.2610422.

[128] Jean-Daniel Fekete and Nicole Dufournaud. Compus: Visualization and analysis of structured documents for understanding social life in the 16th century. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, DL '00, pages 47–55. ACM, 2000. doi:10.1145/336597.336632.

[129] Jean-Daniel Fekete, Jarke J. van Wijk, John T. Stasko, and Chris North. The value of information visualization. In Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, volume 4950 of *Lecture Notes in Computer Science*, pages 1–18. Springer Berlin Heidelberg, 2008. doi:10.1007/978-3-540-70956-5_1.

[130] Cristian Felix, Steven Franconeri, and Enrico Bertini. Taking word clouds apart: An empirical investigation of the design space for keyword summaries. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):657–666, January 2018. doi:10.1109/TVCG.2017.2746018.

[131] Roy T. Fielding and Richard N. Taylor. Principled design of the modern web architecture. *ACM Transactions on Internet Technology*, 2(2):115–150, May 2002. doi:10.1145/514183.514185.

[132] Flask, 2010. Accessed February 4, 2019. URL: http://flask.pocoo.org/.

[133] Johnny R. J. Fontaine, Klaus R. Scherer, and Cristina Soriano. *Components of Emotional Meaning: A Sourcebook*. Oxford University Press, 2013. doi:10.1093/acprof:oso/9780199592746.001.0001.

[134] Eric Freeman and David Gelernter. Lifestreams: A storage model for personal data. *ACM SIGMOD Record*, 25(1):80–86, March 1996. doi:10.1145/381854.381893.

[135] Siwei Fu, Yong Wang, Yi Yang, Qingqing Bi, Fangzhou Guo, and Huamin Qu. VisForum: A visual analysis system for exploring user groups in online forums. *ACM Transactions on Interactive Intelligent Systems*, 8(1):3:1–3:21, February 2018. doi:10.1145/3162075.

[136] Siwei Fu, Jian Zhao, Weiwei Cui, and Huamin Qu. Visual analysis of MOOC forums with iForum. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):201–210, January 2017. doi:10.1109/TVCG.2016.2598444.

[137] Tomohiro Fukuhara, Hiroshi Nakagawa, and Toyoaki Nishida. Understanding sentiment of people from news articles: Temporal sentiment analysis of social events. In *Proceedings of The 2nd International Conference on Weblogs and Social Media*, ICWSM '07, 2007.

[138] Susan R. Fussell, editor. *The Verbal Communication of Emotions: Interdisciplinary Perspectives*. Taylor & Francis, 2002.

[139] Guia Gali, Symon Oliver, Fanny Chevalier, and Sara Diamond. Visualizing sentiments in business-customer relations with metaphors. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 1493–1498. ACM, 2012. doi:10.1145/2212776.2223661.

[140] Michael Gamon, Anthony Aue, Simon Corston-Oliver, and Eric Ringger. Pulse: Mining customer opinions from free text. In A. Fazel Famili, Joost N. Kok, José M. Peña, Arno Siebes, and Ad Feelders, editors, *Proceedings of 6th International Symposium on Intelligent Data Analysis (IDA 2005)*, volume 3646 of *Lecture Notes in Computer Science*, pages 121–132. Springer Berlin Heidelberg, 2005. doi:10.1007/11552253_12.

[141] Michael Gamon, Sumit Basu, Dmitriy Belenko, Danyel Fisher, Matthew Hurst, and Arnd Christian König. BLEWS: Using blogs to provide context for news articles. In *Proceedings of the International Conference on Weblogs and Social Media*, ICWSM '08, pages 60–67. AAAI Press, 2008. URL: http://aaai.org/Library/ICWSM/2008/icwsm08-015.php.

[142] Qihong Gan, Min Zhu, Mingzhao Li, Ting Liang, Yu Cao, and Baoyao Zhou. Document visualization: An overview of current research. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(1):19–36, 2014. doi:10.1002/wics.1285.

[143] Emden R. Gansner, Yifan Hu, and Stephen Kobourov. GMap: Visualizing graphs and clusters as maps. In *Proceedings of the IEEE Pacific Visualization Symposium*, PacificVis '10, pages 201–208. IEEE, 2010. doi:10.1109/PACIFICVIS.2010.5429590.

[144] Viola Ganter and Michael Strube. Finding hedges by chasing weasels: Hedge detection using Wikipedia tags and shallow linguistic features. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 173–176. ACL, 2009. URL: http://aclweb.org/anthology/P09-2044.

[145] Mingkun Gao, Hyo Jin Do, and Wai-Tat Fu. An intelligent interface for organizing online opinions on controversial topics. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, IUI '17, pages 119–123. ACM, 2017. doi:10.1145/3025171.3025230.

[146] GeneralInquirer, 2002. Accessed February 4, 2019. URL: http://wjh.harvard.edu/~inquirer/.

[147] Zhao Geng, Tom Cheesman, Robert S. Laramee, Kevin Flanagan, and Stephan Thiel. ShakerVis: Visual analysis of segment variation of German translations of Shakespeare's Othello. *Information Visualization*, 14(4):273–288, 2015. doi:10.1177/1473871613495845.

[148] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011. doi:10.1177/1473871611416549.

[149] Dylan Glynn and Mette Sjölin, editors. *Subjectivity and Epistemicity: Corpus, Discourse, and Literary Approaches to Stance*. Lund Studies in English. Lund University Press, 2014.

[150] Stephane Gobron, Junghyun Ahn, Georgios Paltoglou, Michael Thelwall, and Daniel Thalmann. From sentence to emotion: A real-time three-dimensional graphics metaphor of emotions extracted from text. *The Visual Computer*, 26(6):505–519, 2010. doi:10.1007/s00371-010-0446-x.

[151] Jennifer Golbeck, Zahra Ashktorab, Rashad O. Banjo, Alexandra Berlinger, Siddharth Bhagwan, Cody Buntain, Paul Cheakalos, Alicia A. Geller, Quint Gergory, Rajesh Kumar Gnanasekaran, Raja Rajan Gunasekaran, Kelly M. Hoffman, Jenny Hottle, Vichita Jienjitlert, Shivika Khare, Ryan Lau, Marianna J. Martindale, Shalmali Naik, Heather L. Nixon, Piyush Ramachandran,

Kristine M. Rogers, Lisa Rogers, Meghna Sardana Sarin, Gaurav Shahane, Jayanee Thanki, Priyanka Vengataraman, Zijian Wan, and Derek Michael Wu. A large labeled corpus for online harassment research. In *Proceedings of the ACM Conference on Web Science*, WebSci '17, pages 229–233. ACM, 2017. `doi:10.1145/3091478.3091509`.

[152] Jennifer Golbeck, Matthew Mauriello, Brooke Auxier, Keval H. Bhanushali, Christopher Bonk, Mohamed Amine Bouzaghrane, Cody Buntain, Riya Chanduka, Paul Cheakalos, Jennine B. Everett, Waleed Falak, Carl Gieringer, Jack Graney, Kelly M. Hoffman, Lindsay Huth, Zhenya Ma, Mayanka Jha, Misbah Khan, Varsha Kori, Elo Lewis, George Mirano, William T. Mohn IV, Sean Mussenden, Tammie M. Nelson, Sean Mcwillie, Akshat Pant, Priya Shetye, Rusha Shrestha, Alexandra Steinheimer, Aditya Subramanian, and Gina Visnansky. Fake news vs satire: A dataset and analysis. In *Proceedings of the ACM Conference on Web Science*, WebSci '18, pages 17–21. ACM, 2018. `doi:10.1145/3201064.3201100`.

[153] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting personality from Twitter. In *Proceedings of the IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing*, PASSAT/SocialCom '11, pages 149–156. IEEE, 2011. `doi:10.1109/PASSAT/SocialCom.2011.33`.

[154] Valentin Gold, Christian Rohrdantz, and Mennatallah El-Assady. Exploratory text analysis using lexical episode plots. In *Short Papers of the EG/VGTC Conference on Visualization*, EuroVis '15. The Eurographics Association, 2015. `doi:10.2312/eurovisshort.20151130`.

[155] Diego Gomez-Zara, Miriam Boon, and Larry Birnbaum. Who is the hero, the villain, and the victim?: Detection of roles in news articles using natural language techniques. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces*, IUI '18, pages 311–315. ACM, 2018. `doi:10.1145/3172944.3172993`.

[156] Carsten Görg, Zhicheng Liu, Jaeyeon Kihm, Jaegul Choo, Haesun Park, and John Stasko. Combining computational analyses and interactive visualization for document exploration and sensemaking in Jigsaw. *IEEE Transactions on Visualization and Computer Graphics*, 19(10):1646–1663, 2013. `doi:10.1109/TVCG.2012.324`.

[157] Michelle L. Gregory, Nancy Chinchor, Paul Whitney, Richard Carter, Elizabeth Hetzler, and Alan Turner. User-directed sentiment analysis: Visualizing the affective content of documents. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, SST '06, pages 23–30. ACL, 2006. URL: `http://anthology.aclweb.org/W/W06/W06-03.pdf#page=33`.

[158] Herbert Paul Grice. Meaning. *The Philosophical Review*, 66(3):377–388, 1957.

[159] Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. TweetCred: Real-time credibility assessment of content on Twitter. In Luca Maria Aiello and Daniel McFarland, editors, *Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11–13, 2014, Proceedings*, volume 8851 of *Lecture Notes in Computer Science*, pages 228–243. Springer International Publishing, 2014. `doi:10.1007/978-3-319-13734-6_16`.

[160] Emitza Guzman. Visualizing emotions in software development projects. In *Proceedings of the 2013 First IEEE Working Conference on Software Visualization*, VISSOFT '13, pages 1–4. IEEE, 2013. `doi:10.1109/VISSOFT.2013.6650529`.

[161] Emitza Guzman, Padma Bhuvanagiri, and Bernd Bruegge. FAVe: Visualizing user feedback for software evolution. In *Proceedings of the 2014 Second IEEE Working Conference on Software Visualization*, VISSOFT '14, pages 167–171. IEEE, 2014. `doi:10.1109/VISSOFT.2014.33`.

[162] Hyoji Ha, Gi-nam Kim, Wonjoo Hwang, Hanmin Choi, and Kyungwon Lee. CosMovis: Analyzing semantic network of sentiment words in movie reviews. In *Poster Abstracts of the IEEE Symposium on Large Data Analysis and Visualization*, LDAV '14, pages 113–114. IEEE, 2014. `doi:10.1109/LDAV.2014.7013215`.

[163] Lieve Hamers, Yves Hemeryck, Guido Herweyers, Marc Janssen, Hans Keters, Ronald Rousseau, and André Vanhoutte. Similarity measures in scientometric research: The Jaccard index

versus Salton's cosine formula. *Information Processing & Management*, 25(3):315–318, 1989. doi:10.1016/0306-4573(89)90048-4.

[164] Ming C. Hao, Christian Rohrdantz, Halldór Janetzko, Daniel A. Keim, Umeshwar Dayal, Lars-Erik Haug, Meichun Hsu, and Florian Stoffel. Visual sentiment analysis of customer feedback streams using geo-temporal term associations. *Information Visualization*, 12(3–4):273–290, 2013. doi:10.1177/1473871613481691.

[165] Christopher Harris. Searching for diverse perspectives in news articles: Using an LSTM network to classify sentiment. In *Proceedings of the ACM IUI 2018 Workshop on Exploratory Search and Interactive Data Analytics*, ESIDA '18. CEUR-WS.org, 2018. URL: http://ceur-ws.org/Vol-2068.

[166] Kazi Saidul Hasan and Vincent Ng. Why are you taking this stance? Identifying and classifying reasons in ideological debates. In *Proceedings of the 2012 Conference on Empirical Methods on Natural Language Processing*, volume 14 of *EMNLP 2014*, pages 751–762. ACL, 2014. URL: http://aclweb.org/anthology/D14-1083.

[167] Susan Havre, Beth Hetzler, and Lucy Nowell. ThemeRiver: Visualizing theme changes over time. In *Proceedings of the IEEE Symposium on Information Visualization*, InfoVis '00, pages 115–123. IEEE, 2000. doi:10.1109/INFVIS.2000.885098.

[168] Hongmei He, Ondrej Sýkora, and Imrich Vrt'o. Crossing minimisation heuristics for 2-page drawings. *Electronic Notes in Discrete Mathematics*, 22:527–534, 2005. doi:10.1016/j.endm.2005.06.088.

[169] Yulan He. A Bayesian modeling approach to multi-dimensional sentiment distributions prediction. In *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, WISDOM '12, pages 1:1–1:8. ACM, 2012. doi:10.1145/2346676.2346677.

[170] Marti A. Hearst. TileBars: Visualization of term distribution information in full text information access. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 59–66. ACM Press/Addison-Wesley Publishing Co., 1995. doi:10.1145/223904.223912.

[171] Florian Heimerl and Michael Gleicher. Interactive analysis of word vector embeddings. *Computer Graphics Forum*, 37(3):253–265, 2018. doi:10.1111/cgf.13417.

[172] Florian Heimerl, Steffen Koch, Harald Bosch, and Thomas Ertl. Visual classifier training for text document retrieval. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2839–2848, December 2012. doi:10.1109/TVCG.2012.277.

[173] Julian Heinrich and Daniel Weiskopf. State of the art of parallel coordinates. In *Eurographics 2013 — State of the Art Reports*. The Eurographics Association, 2013. doi:10.2312/conf/EG2013/stars/095-116.

[174] Patrick Hennig, Philipp Berger, Maximilian Brehm, Bastien Grasnick, Jonathan Herdt, and Christoph Meinel. Hot spot detection — An interactive cluster heat map for sentiment analysis. In *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics*, DSAA 2015. IEEE, 2015. doi:10.1109/DSAA.2015.7344885.

[175] Patrick Hennig, Philipp Berger, Christoph Meinel, Lukas Pirl, and Lukas Schulze. Exploring emotions over time within the blogosphere. In *Proceedings of the 2014 International Conference on Data Science and Advanced Analytics*, DSAA 2014, pages 587–592. IEEE, 2014. doi:10.1109/DSAA.2014.7058131.

[176] Uta Hinrichs, Mennatallah El-Assady, Adam James Bradley, Stefania Forlini, and Christopher Collins. Risk the drift! Stretching disciplinary boundaries through critical collaborations between the humanities and visualization. In *Proceedings of the 2nd Workshop on Visualization for the Digital Humanities*, VIS4DH '17, 2017. URL: http://vis4dh.dbvis.de/papers/2017/.

[177] Orland Hoeber, Larena Hoeber, Maha El Meseery, Kenneth Odoh, and Radhika Gopi. Visual Twitter analytics (Vista): Temporally changing sentiment and the discovery of emergent themes within sport event tweets. *Online Information Review*, 40(1):25–41, 2016. doi:10.1108/OIR-02-2015-0067.

[178] Benjamin Höferlin, Rudolf Netzel, Markus Höferlin, Daniel Weiskopf, and Gunther Heidemann. Inter-active learning of ad-hoc classifiers for video visual analytics. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '12, pages 23–32. IEEE, 2012. doi:10.1109/VAST.2012.6400492.

[179] Fred Hohman, Sandeep Soni, Ian Stewart, , and John Stasko. A Viz of Ice and Fire: Exploring entertainment video using color and dialogue. In *Proceedings of the 2nd Workshop on Visualization for the Digital Humanities*, VIS4DH '17, 2017. URL: http://vis4dh.dbvis.de/papers/2017/.

[180] Enamul Hoque and Giuseppe Carenini. ConVis: A visual text analytic system for exploring blog conversations. *Computer Graphics Forum*, 33(3):221–230, 2014. doi:10.1111/cgf.12378.

[181] Enamul Hoque and Giuseppe Carenini. ConVisIT: Interactive topic modeling for exploring asynchronous online conversations. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, IUI '15, pages 169–180. ACM, 2015. doi:10.1145/2678025.2701370.

[182] Enamul Hoque and Giuseppe Carenini. MultiConVis: A visual text analytics system for exploring a collection of online conversations. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, IUI '16, pages 96–107. ACM, 2016. doi:10.1145/2856767.2856782.

[183] David W. Hosmer, Jr., Stanley Lemeshow, and Rodney X. Sturdivant. *Applied Logistic Regression*. John Wiley & Sons, Inc., 2013. doi:10.1002/9781118548387.

[184] Lulu Huang, Stan Matwin, Eder J. de Carvalho, and Rosane Minghim. Active learning with visualization for text data. In *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*, ESIDA '17, pages 69–74. ACM, 2017. doi:10.1145/3038462.3038469.

[185] Shih-Wen Huang, Pei-Fen Tu, Wai-Tat Fu, and Mohammad Amanzadeh. Leveraging the crowd to improve feature-sentiment analysis of user reviews. In *Proceedings of the International Conference on Intelligent User Interfaces*, IUI '13, pages 3–14. ACM, 2013. doi:10.1145/2449396.2449400.

[186] Shah Rukh Humayoun, Saman Ardalan, Ragaad AlTarawneh, and Achim Ebert. TExVis: An interactive visual tool to explore Twitter data. In *Proceedings of the EG/VGTC Conference on Visualization — Short Papers*, EuroVis '17. The Eurographics Association, 2017. doi:10.2312/eurovisshort.20171149.

[187] Michael Hundt, Bruno Schneider, Mennatallah El-Assady, Daniel A. Keim, and Alexandra Diehl. Visual analysis of geolocated echo chambers in social media. In *Poster Abstracts of the EG/VGTC Conference on Visualization*, EuroVis '17. The Eurographics Association, 2017. doi:10.2312/eurp.20171185.

[188] Susan Hunston. *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. Routledge, 2011.

[189] Susan Hunston. Flavours of corpus linguistics. Paper given at Charles University, Prague 2012 and at Corpus Linguistics 2011, Birmingham, 2014.

[190] Samuel Huron, Romain Vuillemot, and Jean-Daniel Fekete. Visual sedimentation. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2446–2455, December 2013. doi:10.1109/TVCG.2013.227.

[191] C.J. Hutto and Eric Gilbert. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, ICWSM '14. AAAI Press, 2014. URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109.

[192] Petra Isenberg, Niklas Elmqvist, Jean Scholtz, Daniel Cernea, Kwan-Liu Ma, and Hans Hagen. Collaborative visualization: Definition, challenges, and research agenda. *Information Visualization*, 10(4):310–326, 2011. doi:10.1177/1473871611412817.

[193] Tobias Isenberg, Petra Isenberg, Jian Chen, Michael Sedlmair, and Torsten Möller. A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2818–2827, December 2013. doi:10.1109/TVCG.2013.126.

[194] Dominik Jäckle, Fabian Fischer, Tobias Schreck, and Daniel A. Keim. Temporal MDS plots for analysis of multivariate data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):141–150, January 2016. doi:10.1109/TVCG.2015.2467553.

[195] Donald A. Jackson, Keith M. Somers, and Harold H. Harvey. Similarity coefficients: Measures of co-occurrence and association or simply measures of occurrence? *The American Naturalist*, 133(3):436–453, 1989. doi:10.1086/284927.

[196] Alexandra Jaffe. *Stance: Sociolinguistic Perspectives*. Oxford University Press, 2009.

[197] Stefan Jänicke, Greta Franzini, Muhammad Faisal Cheema, and Gerik Scheuermann. On close and distant reading in digital humanities: A survey and future challenges. In *Proceedings of the EG/VGTC Conference on Visualization — STARs*, EuroVis '15. The Eurographics Association, 2015. doi:10.2312/eurovisstar.20151113.

[198] Stefan Jänicke, Greta Franzini, Muhammad Faisal Cheema, and Gerik Scheuermann. Visual text analysis in digital humanities. *Computer Graphics Forum*, 36(6):226–250, 2017. doi:10.1111/cgf.12873.

[199] Stefan Jänicke, Annette Geßner, Marco Büchler, and Gerik Scheuermann. Visualizations for text re-use. In *Proceedings of the 5th International Conference on Information Visualization Theory and Applications — Volume 1: IVAPP*, VISIGRAPP '14, pages 59–70, Lisbon, Portugal, 2014. SciTePress. doi:10.5220/0004692500590070.

[200] Wolfgang Jentner, Mennatallah El-Assady, Bela Gipp, and Daniel A. Keim. Feature alignment for the analysis of verbatim text transcripts. In *Proceedings of the EuroVis Workshop on Visual Analytics*, EuroVA '17. The Eurographics Association, 2017. doi:10.2312/eurova.20171113.

[201] Jericho HTML parser, 2012. Accessed February 4, 2019. URL: http://jericho.htmlparser.net/.

[202] Peerumporn Jiranantanagorn and Haifeng Shen. Sentiment analysis and visualisation in a backchannel system. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, OzCHI '16, pages 353–357. ACM, 2016. doi:10.1145/3010915.3010992.

[203] Yohan Jo and Alice H. Oh. Aspect and sentiment unification model for online review analysis. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 815–824. ACM, 2011. doi:10.1145/1935826.1935932.

[204] Ian T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag New York, 1986. doi:10.1007/978-1-4757-1904-8.

[205] jQuery, 2011. Accessed February 4, 2019. URL: https://jquery.com/.

[206] JSON (JavaScript Object Notation), 2002. Accessed February 4, 2019. URL: http://json.org.

[207] Ata Kabán and Mark A. Girolami. A dynamic probabilistic model to visualise topic evolution in text streams. *Journal of Intelligent Information Systems*, 18(2–3):107–125, 2002. doi:10.1023/A:1013673310093.

[208] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ACL '14, pages 655–665. ACL, 2014. URL: http://www.anthology.aclweb.org/P/P14/P14-1062.pdf.

[209] Xin Kang and Fuji Ren. Sampling latent emotions and topics in a hierarchical Bayesian network. In *Proceedings of the 2011 7th International Conference on Natural Language Processing and Knowledge Engineering*, NLP-KE '11, pages 37–42. IEEE, 2011. doi:10.1109/NLPKE.2011.6138166.

[210] Joseph 'Jofish' Kaye, Anita Lillie, Deepak Jagdish, James Walkup, Rita Parada, and Koichi Mori. Nokia Internet Pulse: A long term deployment and iteration of a Twitter visualization. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 829–844. ACM, 2012. doi:10.1145/2212776.2212856.

[211] Colm Kearney and Sha Liu. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33:171–185, 2014. doi:10.1016/j.irfa.2014.02.006.

[212] Daniel A. Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Visual analytics: Definition, process, and challenges. In Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, volume 4950 of *Lecture Notes in Computer Science*, pages 154–175. Springer Berlin Heidelberg, 2008. doi:10.1007/978-3-540-70956-5_7.

[213] Daniel A. Keim, Miloš Krstajić, Christian Rohrdantz, and Tobias Schreck. Real-time visual analytics for text streams. *Computer*, 46(7):47–55, July 2013. doi:10.1109/MC.2013.152.

[214] Daniel A. Keim and Daniela Oelke. Literature fingerprinting: A new method for visual literary analysis. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, VAST '07, pages 115–122. IEEE, 2007. doi:10.1109/VAST.2007.4389004.

[215] Renato Kempter, Valentina Sintsova, Claudiu Musat, and Pearl Pu. EmotionWatch: Visualizing fine-grained emotions in event-related tweets. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, ICWSM '14, pages 236–245. AAAI Press, 2014. URL: http://aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8117.

[216] Andreas Kerren, Kostiantyn Kucher, Yuan-Fang Li, and Falk Schreiber. BioVis Explorer: A visual guide for biological data visualization techniques. *PLOS ONE*, 12(11), November 2017. doi:10.1371/journal.pone.0187341.

[217] Andreas Kerren, Kostiantyn Kucher, Yuan-Fang Li, and Falk Schreiber. MDS-based visual survey of biological data visualization techniques. In *Poster Abstracts of the EG/VGTC Conference on Visualization*, EuroVis '17, pages 85–87. The Eurographics Association, 2017. doi:10.2312/eurp.20171175.

[218] Andreas Kerren, Mimi Kyusakova, and Carita Paradis. From culture to text to interactive visualization of wine reviews. In Francis T. Marchese and Ebad Banissi, editors, *Knowledge Visualization Currents: From Text to Art to Culture*, chapter 5, pages 85–110. Springer, 2012. doi:10.1007/978-1-4471-4303-1_5.

[219] Andreas Kerren and Falk Schreiber. Toward the role of interaction in visual analytics. In *Proceedings of the Winter Simulation Conference*, WSC '12, pages 420:1–420:13. Winter Simulation Conference, 2012. URL: http://dl.acm.org/citation.cfm?id=2429759.2430303.

[220] Pooja Kherwa, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, and Prashast Kumar Singh. An approach towards comprehensive sentimental data analysis and opinion mining. In *Proceedings of the 2014 IEEE International Advance Computing Conference*, IACC '14, pages 606–612. IEEE, 2014. doi:10.1109/IAdCC.2014.6779394.

[221] Kyoungok Kim and Jaewook Lee. Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recognition*, 47(2):758–768, 2014. doi:10.1016/j.patcog.2013.07.022.

[222] Nam Wook Kim, Benjamin Bach, Hyejin Im, Sasha Schriber, Markus Gross, and Hanspeter Pfister. Visualizing nonlinear narratives with story curves. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):595–604, January 2018. doi:10.1109/TVCG.2017.2744118.

[223] Robert M. Kirby and Miriah Meyer. Visualization collaborations: What works and why. *IEEE Computer Graphics and Applications*, 33(6):82–88, November 2013. doi:10.1109/MCG.2013.101.

[224] Steffen Koch, Markus John, Michael Wörner, Andreas Müller, and Thomas Ertl. VarifocalReader — In-depth visual analysis of large text documents. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1723–1732, December 2014. doi:10.1109/TVCG.2014.2346677.

[225] Teuvo Kohonen. The self-organizing map. *Neurocomputing*, 21(1):1–6, 1998. doi:10.1016/S0925-2312(98)00030-7.

[226] Robert Kosara. An empire built on sand: Reexamining what we think we know about visualization. In *Proceedings of the Sixth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '16, pages 162–168. ACM, 2016. doi:10.1145/2993901.2993909.

[227] Robert Kosara, Fritz Drury, Lars Erik Holmquist, and David H. Laidlaw. Visualization criticism. *IEEE Computer Graphics and Applications*, 28(3):13–15, May 2008. doi:10.1109/MCG.2008.63.

[228] Janez Kranjc, Jasmina Smailović, Vid Podpečan, Miha Grčar, Martin Žnidaršič, and Nada Lavrač. Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the ClowdFlows platform. *Information Processing & Management*, 51(2):187–203, 2015. doi:10.1016/j.ipm.2014.04.001.

[229] Uros Krcadinac, Jelena Jovanovic, Vladan Devedzic, and Philippe Pasquier. Textual affect communication and evocation using abstract generative visuals. *IEEE Transactions on Human-Machine Systems*, 46(3):370–379, June 2016. doi:10.1109/THMS.2015.2504081.

[230] Miloš Krstajić, Christian Rohrdantz, Michael Hund, and Andreas Weiler. Getting there first: Real-time detection of real-world incidents on Twitter. In *Proceedings of the 2nd IEEE Workshop on Interactive Visual Text Analytics "Task-Driven Analysis of Social Media"*, TextVis '12, 2012. URL: http://vialab.science.uoit.ca/textvis2012/.

[231] Miloš Krstajić, Florian Mansmann, Andreas Stoffel, Martin Atkinson, and Daniel A. Keim. Processing online news streams for large-scale semantic analysis. In *Proceedings of the 26th International IEEE Conference on Data Engineering Workshops*, ICDEW '10, pages 215–220. IEEE, 2010. doi:10.1109/ICDEW.2010.5452710.

[232] Wojtek J. Krzanowski. *Principles of Multivariate Analysis*. Oxford Statistical Science Series. OUP Oxford, 2000.

[233] Chong Kuang, Jiayu Tang, Zhiyuan Liu, and Maosong Sun. ImgWordle: Image and text visualization for events in microblogging services. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, AVI '14, pages 371–372. ACM, 2014. doi:10.1145/2598153.2600041.

[234] Kostiantyn Kucher, Daniel Cernea, and Andreas Kerren. Visualizing excitement of individuals and groups. In *Proceedings of the ACM IUI 2016 Workshop on Emotion and Visualization*, EmoVis 2016, pages 15–22. Linköping University Electronic Press, 2016. doi:10.3384/ecp10303.

[235] Kostiantyn Kucher and Andreas Kerren. Text visualization browser: A visual survey of text visualization techniques. In *Poster Abstracts of the IEEE Conference on Information Visualization*, InfoVis '14, 2014.

[236] Kostiantyn Kucher and Andreas Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *Proceedings of the IEEE Pacific Visualization Symposium*, PacificVis '15, pages 117–121. IEEE, 2015. doi:10.1109/PACIFICVIS.2015.7156366.

[237] Kostiantyn Kucher, Andreas Kerren, Carita Paradis, and Magnus Sahlgren. Visual analysis of stance markers in online social media. In *Poster Abstracts of the IEEE Conference on Visual Analytics Science and Technology*, VAST '14, pages 259–260. IEEE, 2014. doi:10.1109/VAST.2014.7042519.

[238] Kostiantyn Kucher, Andreas Kerren, Carita Paradis, and Magnus Sahlgren. Visual analysis of text annotations for stance classification with ALVA. In *Poster Abstracts of the EG/VGTC Conference on Visualization*, EuroVis '16, pages 49–51. The Eurographics Association, 2016. doi:10.2312/eurp.20161139.

[239] Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. DoSVis: Document stance visualization. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP '18) — Volume 3: IVAPP*, IVAPP '18, pages 168–175. SciTePress, 2018. doi:10.5220/0006539101680175.

[240] Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. The state of the art in sentiment visualization. *Computer Graphics Forum*, 37(1):71–96, February 2018. doi:10.1111/cgf.13217.

[241] Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. Visual analysis of sentiment and stance in social media texts. In *Poster Abstracts of the EG/VGTC Conference on Visualization*, EuroVis '18, pages 49–51. The Eurographics Association, 2018. doi:10.2312/eurp.20181127.

[242] Kostiantyn Kucher, Carita Paradis, Magnus Sahlgren, and Andreas Kerren. Active learning and visual analytics for stance classification with ALVA. *ACM Transactions on Interactive Intelligent Systems*, 7(3):14:1–14:31, October 2017. doi:10.1145/3132169.

[243] Kostiantyn Kucher, Teri Schamp-Bjerede, Andreas Kerren, Carita Paradis, and Magnus Sahlgren. Visual analysis of online social media to open up the investigation of stance phenomena. *Information Visualization*, 15(2):93–116, April 2016. doi:10.1177/1473871615575079.

[244] Kostiantyn Kucher, Maria Skeppstedt, and Andreas Kerren. Application of interactive computer-assisted argument extraction to opinionated social media texts. In *Poster Abstracts of the 11th International Symposium on Visual Information Communication and Interaction*, VINCI '18, pages 102–103. ACM, 2018. doi:10.1145/3231622.3232505.

[245] Katie Kuksenok, Michael Brooks, John J. Robinson, Daniel Perry, Megan K. Torkildson, and Cecilia Aragon. Automating large-scale annotation for analysis of social media content. In *Proceedings of the 2nd IEEE Workshop on Interactive Visual Text Analytics "Task-Driven Analysis of Social Media"*, TextVis '12, 2012. URL: http://vialab.science.uoit.ca/textvis2012/.

[246] Tadahiko Kumamoto, Hitomi Wada, and Tomoya Suzuki. Visualizing temporal changes in impressions from tweets. In *Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services*, iiWAS '14, pages 116–125. ACM, 2014. doi:10.1145/2684200.2684279.

[247] Bum Chul Kwon, Sung-Hee Kim, Sukwon Lee, Jaegul Choo, Jina Huh, and Ji Soo Yi. VisOHC: Designing visual analytics for online health communities. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):71–80, January 2016. doi:10.1109/TVCG.2015.2467555.

[248] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, September 2012. doi:10.1109/TVCG.2011.279.

[249] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning*, ICML '14, pages 1188–1196. PMLR, 2014. URL: http://proceedings.mlr.press/v32/le14.pdf.

[250] Hyowon Lee, Paul Ferguson, Neil O'Hare, Cathal Gurrin, and Alan F. Smeaton. Integrating interactivity into visualising sentiment analysis of blogs. In *Proceedings of the First International Workshop on Intelligent Visual Interfaces for Text Analysis*, IVITA '10, pages 17–20. ACM, 2010. doi:10.1145/2002353.2002360.

[251] Frank Lentricchia and Andrew DuBois, editors. *Close Reading: The Reader*. Duke University Press, 2002.

[252] Eldon Y. Li, Chien Hsiang Liao, and Hsiuju Rebecca Yen. Co-authorship networks and research impact: A social capital perspective. *Research Policy*, 42(9):1515–1530, 2013. doi:10.1016/j.respol.2013.06.012.

[253] Hanchuan Li, Haichen Shen, Shengliang Xu, and Congle Zhang. Visualizing NLP annotations for crowdsourcing. *CoRR*, abs/1508.06044, 2015. URL: http://arxiv.org/abs/1508.06044.

[254] Huana Li and Fuji Ren. The study on text emotional orientation based on a three-dimensional emotion space model. In *Proceedings of the International Conference on Natural Language Processing and Knowledge Engineering*, NLP-KE '09. IEEE, 2009. doi:10.1109/NLPKE.2009.5313815.

[255] Qiusheng Li, Yadong Wu, Song Wang, Maosong Lin, Xinmiao Feng, and Haiyang Wang. VisTravel: Visualizing tourism network opinion from the user generated content. *Journal of Visualization*, 19(3):489–502, 2016. doi:10.1007/s12650-015-0330-x.

[256] Xiaojun Li, Lin Dai, and Hanxiao Shi. Opinion mining of camera reviews based on semantic role labeling. In *Proceedings of the Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, volume 5 of *FSKD '10*, pages 2372–2375. IEEE, 2010. doi:10.1109/FSKD.2010.5569525.

[257] Chenghua Lin and Yulan He. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CKIM '09, pages 375–384. ACM, 2009. doi:10.1145/1645953.1646003.

[258] Chenghua Lin, Yulan He, and Richard Everson. Sentence subjectivity detection with weakly-supervised learning. In *Proceedings of Fifth International Joint Conference on Natural Language Processing*, IJCNLP '11, pages 1153–1161. ACL, 2011.

[259] Xia Lin. Visualization for the document space. In *Proceedings of the IEEE Conference on Visualization*, Visualization '92, pages 274–281. IEEE, 1992. doi:10.1109/VISUAL.1992.235198.

[260] Bing Liu. Sentiment analysis and subjectivity. In Nitin Indurkhya and Fred J. Damerau, editors, *Handbook of Natural Language Processing*, pages 627–666. Chapman & Hall, 2010.

[261] Bing Liu, Minqing Hu, and Junsheng Cheng. Opinion Observer: Analyzing and comparing opinions on the Web. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, pages 342–351. ACM, 2005. doi:10.1145/1060745.1060797.

[262] Hugo Liu, Ted Selker, and Henry Lieberman. Visualizing the affective structure of a text document. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, pages 740–741. ACM, 2003. doi:10.1145/765891.765961.

[263] Shenghua Liu, Wenjun Zhu, Ning Xu, Fangtao Li, Xue-qi Cheng, Yue Liu, and Yuanzhuo Wang. Co-training and visualizing sentiment evolvement for tweet events. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion, pages 105–106. ACM, 2013. doi:10.1145/2487788.2487836.

[264] Shixia Liu, Xiting Wang, Christopher Collins, Wenwen Dou, Fangxin Ouyang, Mennatallah El-Assady, Liu Jiang, and Daniel A. Keim. Bridging text visualization and mining: A task-driven survey. *IEEE Transactions on Visualization and Computer Graphics*, 2018. doi:10.1109/TVCG.2018.2834341.

[265] Shixia Liu, Yingcai Wu, Enxun Wei, Mengchen Liu, and Yang Liu. StoryFlow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2436–2445, December 2013. doi:10.1109/TVCG.2013.196.

[266] Shusen Liu, Zhimin Li, Tao Li, Vivek Srikumar, Valerio Pascucci, and Peer-Timo Bremer. NLIZE: A perturbation-driven visual interrogation tool for analyzing and interpreting natural language inference models. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):651–660, January 2019. doi:10.1109/TVCG.2018.2865230.

[267] Shusen Liu, Dan Maljovec, Bei Wang, Peer-Timo Bremer, and Valerio Pascucci. Visualizing high-dimensional data: Advances in the past decade. In *Eurographics Conference on Visualization (EuroVis) — STARs*, EuroVis '15. The Eurographics Association, 2015. doi:10.2312/eurovisstar.20151115.

[268] Xiaotong Liu, Anbang Xu, Liang Gou, Haibin Liu, Rama Akkiraju, and Han-Wei Shen. SocialBrands: Visual analysis of public perceptions of brands on social media. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '16, pages 71–80. IEEE, 2016. doi:10.1109/VAST.2016.7883513.

[269] Yafeng Lu, Rolando Garcia, Brett Hansen, Michael Gleicher, and Ross Maciejewski. The state-of-the-art in predictive visual analytics. *Computer Graphics Forum*, 36(3):539–562, 2017. doi:10.1111/cgf.13210.

[270] Yafeng Lu, Xia Hu, Feng Wang, Shamanth Kumar, Huan Liu, and Ross Maciejewski. Visualizing social media sentiment in disaster scenarios. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 1211–1215. ACM, 2015. doi:10.1145/2740908.2741720.

[271] Yafeng Lu, Robert Krüger, Dennis Thom, Feng Wang, Steffen Koch, Thomas Ertl, and Ross Maciejewski. Integrating predictive analytics and social media. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '14, pages 193–202. IEEE, 2014. doi:10.1109/VAST.2014.7042495.

[272] Yafeng Lu, Michael Steptoe, Sarah Burke, Hong Wang, Jiun-Yi Tsai, Hasan Davulcu, Douglas Montgomery, Steven R. Corman, and Ross Maciejewski. Exploring evolving media discourse through event cueing. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):220–229, January 2016. doi:10.1109/TVCG.2015.2467991.

[273] Yafeng Lu, Hong Wang, Steven Landis, and Ross Maciejewski. A visual analytics framework for identifying topic drivers in media events. *IEEE Transactions on Visualization and Computer Graphics*, 24(9):2501–2515, September 2018. doi:10.1109/TVCG.2017.2752166.

[274] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies — Volume 1*, HLT '11, pages 142–150. ACL, 2011. URL: http://anthology.aclweb.org/P/P11/P11-1015.pdf.

[275] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(November):2579–2605, 2008. URL: http://jmlr.org/papers/v9/vandermaaten08a.html.

[276] Bernardo Magnini, Anne-Lyse Minard, Mohammed R.H. Qwaider, and Manuela Speranza. TextPro-AL: An active learning platform for flexible and efficient production of training data for NLP tasks. In *Proceedings of the 26th International Conference on Computational Linguistics — System Demonstrations*, COLING '16, pages 131–135. ACL, 2016. URL: https://aclweb.org/anthology/C/C16/C16-2028.pdf.

[277] Jalal Mahmud, Geli Fei, Anbang Xu, Aditya Pal, and Michelle Zhou. Predicting attitude and actions of Twitter users. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, IUI '16, pages 2–6. ACM, 2016. doi:10.1145/2856767.2856800.

[278] Raheleh Makki, Stephen Brooks, and Evangelos E. Milios. Context-specific sentiment lexicon expansion via minimal user interaction. In *Proceedings of the 5th International Conference on Information Visualization Theory and Applications — Volume 1: IVAPP*, VISIGRAPP '14, pages 178–186. SciTePress, 2014. doi:10.5220/0004745301780186.

[279] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[280] Song Mao, Azriel Rosenfeld, and Tapas Kanungo. Document structure analysis algorithms: A literature survey. *SPIE Proceedings*, 5010, 2003. doi:10.1117/12.476326.

[281] Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden, and Robert C. Miller. TwitInfo: Aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 227–236. ACM, 2011. doi:10.1145/1978942.1978975.

[282] Juana I. Marín-Arrese, Marta Carretero, Jorge Arús Hita, and Johan van der Auwera, editors. *English Modality: Core, Periphery and Evidentiality*, volume 81 of *Topics in English Linguistics*. De Gruyter Mouton, 2013.

[283] Edison Marrese-Taylor, Juan D. Velásquez, and Felipe Bravo-Marquez. Opinion Zoom: A modular tool to explore tourism opinions on the Web. In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies*, volume 3 of *WI-IAT '13*, pages 261–264. IEEE, 2013. doi:10.1109/WI-IAT.2013.193.

[284] James R. Martin and Peter R. White. *The Language of Evaluation: Appraisal in English*. Palgrave Macmillan, 2005. doi:10.1057/9780230511910.

[285] Rafael M. Martins and Andreas Kerren. Efficient dynamic time warping for big data streams. In *Proceedings of the 3rd Workshop on Real-time & Stream Analytics in Big Data & Stream Data Management at IEEE Big Data '18*, pages 2924–2929. IEEE, 2018. doi:10.1109/BigData.2018.8621878.

[286] Rafael M. Martins, Vasiliki Simaki, Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. StanceXplore: Visualization for the interactive exploration of stance in social media. In *Proceedings of the 2nd Workshop on Visualization for the Digital Humanities*, VIS4DH '17, 2017. URL: http://vis4dh.dbvis.de/papers/2017/.

[287] Miriah Meyer and Jason Dykes. Reflection on reflection in applied visualization research. *IEEE Computer Graphics and Applications*, 38(6):9–16, November 2018. doi:10.1109/MCG.2018.2874523.

[288] Qingliang Miao, Qiudan Li, and Ruwei Dai. AMAZING: A sentiment mining and retrieval system. *Expert Systems with Applications*, 36(3, Part 2):7192–7198, 2009. doi:10.1016/j.eswa.2008.09.035.

[289] Stanley Milgram and Denise Jodelet. Psychological maps of Paris. In Harold M. Proshansky, William H. Ittelson, and Leanne G. Rivlin, editors, *Environmental Psychology: People and Their Physical Settings*, pages 104–124. Holt, Rinehart, and Winston, New York, 2nd edition, 1976.

[290] George A. Miller. WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39–41, November 1995. doi:10.1145/219717.219748.

[291] Gilad Mishne and Maarten De Rijke. MoodViews: Tools for blog mood analysis. In *Proceedings of the AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pages 153–154. AAAI Press, 2006. URL: https://aaai.org/Library/Symposia/Spring/2006/ss06-03-029.php.

[292] Lewis Mitchell, Morgan R. Frank, Kameron Decker Harris, Peter Sheridan Dodds, and Christopher M. Danforth. The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLOS ONE*, 8(5), May 2013. doi:10.1371/journal.pone.0064417.

[293] Saif Mohammad. Portable features for classifying emotional text. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '12, pages 587–591. ACL, 2012. URL: http://aclweb.org/anthology/N12-1071.

[294] Saif M. Mohammad. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4):730–741, 2012. doi:10.1016/j.dss.2012.05.030.

[295] Saif M. Mohammad. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In Herbert L. Meiselman, editor, *Emotion Measurement*, pages 201–237. Woodhead Publishing, 2016. doi:10.1016/B978-0-08-100508-8.00009-6.

[296] Saif M. Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the International Workshop on Semantic Evaluation*, SemEval '16. ACL, 2016. URL: http://aclweb.org/anthology/S16-1003.

[297] Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. *ACM Transactions on Internet Technology*, 17(3):26:1–26:23, June 2017. doi:10.1145/3003433.

[298] MongoDB, 2009. Accessed February 4, 2019. URL: https://mongodb.com/.

[299] Franco Moretti. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, 2005.

[300] Franco Moretti. *Distant Reading*. Verso Books, 2013.

[301] Meredith Ringel Morris, Scott Counts, Asta Roseway, Aaron Hoff, and Julia Schwarz. Tweeting is believing?: Understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 441–450. ACM, 2012. doi:10.1145/2145204.2145274.

[302] Thomas Morton and Jeremy LaCivita. WordFreak: An open tool for linguistic annotation. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Demonstrations — Volume 4*, NAACL-Demonstrations '03, pages 17–18. ACL, 2003. `doi:10.3115/1073427.1073436`.

[303] Christoph Müller and Michael Strube. Multi-level annotation in MMAX. In *Proceedings of the Fourth SIGdial Workshop on Discourse and Dialogue*, SIGdial '03, pages 198–207. ACL, 2003. URL: `http://aclweb.org/anthology/W03-2117`.

[304] Myriam Munezero, Calkin Suero Montero, Maxim Mozgovoy, and Erkki Sutinen. EmoTwitter — A fine-grained visualization system for identifying enduring sentiments in tweets. In Alexander Gelbukh, editor, *Proceedings of the 16th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing '15)*, volume 9042 of *Lecture Notes in Computer Science*, pages 78–91. Springer International Publishing, 2015. `doi:10.1007/978-3-319-18117-2_6`.

[305] Myriam Munezero, Calkin Suero Montero, Erkki Sutinen, and John Pajunen. Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE Transactions on Affective Computing*, 5(2):101–111, April 2014. `doi:10.1109/TAFFC.2014.2317187`.

[306] Tamara Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, November 2009. `doi:10.1109/TVCG.2009.111`.

[307] Alena Neviarouskaya, Masaki Aono, Helmut Prendinger, and Mitsuru Ishizuka. Intelligent interface for textual attitude analysis. *ACM Transactions on Intelligent Systems and Technology*, 5(3):48:1–48:20, September 2014. `doi:10.1145/2535912`.

[308] Mark E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[309] Vu Dung Nguyen, Blesson Varghese, and Adam Barker. The royal birth of 2013: Analysing and visualising public sentiment in the UK using Twitter. In *Proceedings of the IEEE International Conference on Big Data*, BigData '13, pages 46–54. IEEE, 2013. `doi:10.1109/BigData.2013.6691669`.

[310] Clemens Nopp and Allan Hanbury. Detecting risks in the banking system by sentiment analysis. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, EMNLP '15, pages 591–600. ACL, 2015. URL: `http://aclweb.org/anthology/D15-1071`.

[311] Jaume Nualart-Vilaplana, Mario Pérez-Montoro, and Mitchell Whitelaw. How we draw texts: A review of approaches to text visualization and exploration. *El profesional de la información*, 23(3):221–235, 2014. `doi:10.3145/epi.2014.may.02`.

[312] Daniela Oelke, Ming C. Hao, Christian Rohrdantz, Daniel A. Keim, Umeshwar Dayal, Lars-Erik Haug, and Halldór Janetzko. Visual opinion analysis of customer feedback data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, VAST '09, pages 187–194. IEEE, 2009. `doi:10.1109/VAST.2009.5333919`.

[313] Philip V. Ogren. Knowtator: A protégé plug-in for annotated corpus construction. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology — Companion Volume: Demonstrations*, NAACL-Demonstrations '06, pages 273–275. ACL, 2006. `doi:10.3115/1225785.1225791`.

[314] Samir Okasha. *Philosophy of Science: Very Short Introduction*. Oxford University Press, 2nd edition, 2016.

[315] Daniel E. O'Leary. Ethics for big data and analytics. *IEEE Intelligent Systems*, 31(4):81–84, July 2016. `doi:10.1109/MIS.2016.70`.

[316] João Oliveira, Mike Pinto, Pedro Saleiro, and Jorge Teixeira. SentiBubbles: Topic modeling and sentiment visualization of entity-centric tweets. In *Proceedings of the Ninth International C* Conference on Computer Science & Software Engineering*, C3S2E '16, pages 123–124. ACM, 2016. `doi:10.1145/2948992.2949023`.

[317] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135, 2008. doi:10.1561/1500000011.

[318] Carita Paradis. Between epistemic modality and degree: The case of "really". In Roberta Facchinetti, Frank Palmer, and Manfred Krug, editors, *Modality in Contemporary English*, volume 44 of *Topics in English Linguistics*, pages 191–222. De Gruyter Mouton, 2003.

[319] Carita Paradis. Meanings of words: Theory and application. In Ulrike Haß and Petra Storjohann, editors, *Handbuch Wort und Wortschatz*, volume 3 of *Handbücher Sprachwissen*, pages 274–294. De Gruyter Mouton, 2015.

[320] Deokgun Park, Seungyeon Kim, Jurim Lee, Jaegul Choo, Nicholas Diakopoulos, and Niklas Elmqvist. ConceptVector: Text visual analytics via interactive lexicon building using word embedding. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):361–370, January 2018. doi:10.1109/TVCG.2017.2744478.

[321] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Ëdouard Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, October 2011. URL: http://jmlr.org/papers/v12/pedregosa11a.html.

[322] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee. A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3):45–77, 2007. doi:10.2753/MIS0742-1222240302.

[323] Martín Pérez-Pérez, Daniel Glez-Peña, Florentino Fdez-Riverola, and Anália Lourenço. Marky: A tool supporting annotation consistency in multi-user and iterative document annotation projects. *Computer Methods and Programs in Biomedicine*, 118(2):242–251, 2015. doi:10.1016/j.cmpb.2014.11.005.

[324] Carmelo Pino, Isaak Kavasidis, and Concetto Spampinato. GeoSentiment: A tool for analyzing geographically distributed event-related sentiments. In *Proceedings of the 13th IEEE Annual Consumer Communications Networking Conference*, CCNC 2016, pages 270–271. IEEE, 2016. doi:10.1109/CCNC.2016.7444775.

[325] Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the International Conference on Intelligence Analysis*, volume 5, pages 2–4, 2005.

[326] John C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In Alexander J. Smola, Peter L. Bartlett, Bernhard Schölkopf, and Dale Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, 1999.

[327] Robert Plutchik. A general psychoevolutionary theory of emotion. In Robert Plutchik and Henry Kellerman, editors, *Theories of Emotion*, volume 1 of *Emotion: Theory, Research, and Experience*, pages 3–33. Academic Press, 1980.

[328] Forough Poursabzi-Sangdeh, Jordan Boyd-Graber, Leah Findlater, and Kevin Seppi. ALTO: Active learning with topic overviews for speeding label induction and document labeling. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, ACL '16, pages 1158–1169. ACL, 2016. URL: http://aclweb.org/anthology/P/P16/P16-1110.pdf.

[329] Radityo Eko Prasojo, Fariz Darari, and Mouna Kacimi. ORCAESTRA: Organizing news comments using aspect, entity and sentiment extraction. In *Poster Abstracts of IEEE VIS*, 2015.

[330] Sara Pupi, Giulia Di Pietro, and Carlo Aliprandi. Ent-it-up. In Constantine Stephanidis, editor, *HCI International 2014 — Posters' Extended Abstracts*, volume 435, part 2 of *Communications in Computer and Information Science*, pages 3–8. Springer International Publishing, 2014. doi:10.1007/978-3-319-07854-0_1.

[331] Helen C. Purchase. *Experimental Human-Computer Interaction: A Practical Guide with Visual Examples*. Cambridge University Press, 2012.

[332] Robin G. Qiu, Ramya R. Ravi, and Lawrence L. Qiu. Aggregating and visualizing public opinions and sentiment trends on the US higher education. In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, iiWAS '15, pages 33:1–33:5. ACM, 2015. `doi:10.1145/2837185.2837261`.

[333] Afshin Rahimi, Magnus Sahlgren, Andreas Kerren, and Carita Paradis. The StaViCTA group report for RepLab 2014: Reputation dimensions task. In *Working Notes for CLEF 2014 Conference*, pages 1519–1527. CEUR-WS.org, 2014. URL: `http://ceur-ws.org/Vol-1180`.

[334] Paulo E. Rauber, Alexandre X. Falcão, and Alexandru C. Telea. Visualizing time-dependent data using dynamic t-SNE. In *Short Papers of the EG/VGTC Conference on Visualization*, EuroVis '16. The Eurographics Association, 2016. `doi:10.2312/eurovisshort.20161164`.

[335] Kumar Ravi and Vadlamani Ravi. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, 89:14–46, 2015. `doi:10.1016/j.knosys.2015.06.015`.

[336] Tim Regan and Linda Becker. Visualizing the text of Philip Pullman's trilogy "His Dark Materials". In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordCHI '10, pages 759–764. ACM, 2010. `doi:10.1145/1868914.1869023`.

[337] Earl Rennison. Galaxy of News: An approach to visualizing and understanding expansive news landscapes. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology*, UIST '94, pages 3–12. ACM, 1994. URL: `http://doi.acm.org/10.1145/192426.192429`, `doi:10.1145/192426.192429`.

[338] Rickshaw: A JavaScript toolkit for creating interactive time-series graphs, 2011. Accessed July 26, 2018. URL: `https://github.com/shutterstock/rickshaw`.

[339] Patrick Riehmann, Dora Kiesel, Martin Kohlhaas, and Bernd Froehlich. Visualizing a thinker's life. *IEEE Transactions on Visualization and Computer Graphics*, 25(4):1803–1816, April 2019. `doi:10.1109/TVCG.2018.2824822`.

[340] Ricardo A. Rios, Paulo A. Pagliosa, Renato P. Ishii, and Rodrigo F. de Mello. TSViz: A data stream architecture to online collect, analyze, and visualize tweets. In *Proceedings of the Symposium on Applied Computing*, SAC '17, pages 1031–1036. ACM, 2017. `doi:10.1145/3019612.3019811`.

[341] John Risch, Anne Kao, Stephen R. Poteet, and Yuan-Jye Jason Wu. Text visualization for visual text analytics. In Simeon J. Simoff, Michael H. Böhlen, and Arturas Mazeika, editors, *Visual Data Mining*, volume 4404 of *Lecture Notes in Computer Science*, pages 154–171. Springer Berlin Heidelberg, 2008. `doi:10.1007/978-3-540-71080-6_11`.

[342] Christian Ritter, Christian Altenhofen, Matthias Zeppelzauer, Arjan Kuijper, Tobias Schreck, and Jürgen Bernard. Personalized visual-interactive music classification. In *Proceedings of the EuroVis Workshop on Visual Analytics*, EuroVA '18. The Eurographics Association, 2018. `doi:10.2312/eurova.20181109`.

[343] Jonathan C. Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization*, CMV 2007, pages 61–71. IEEE, 2007. `doi:10.1109/CMV.2007.20`.

[344] George G. Robertson and Jock D. Mackinlay. The document lens. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*, UIST '93, pages 101–108. ACM, 1993. `doi:10.1145/168642.168652`.

[345] Christian Rohrdantz, Ming C. Hao, Umeshwar Dayal, Lars-Erik Haug, and Daniel A. Keim. Feature-based visual sentiment analysis of text document streams. *ACM Transactions on Intelligent Systems and Technology*, 3(2):26:1–26:25, February 2012. `doi:10.1145/2089094.2089102`.

[346] Christian Rohrdantz, Andreas Niekler, Annette Hautli, Miriam Butt, and Daniel A. Keim. Lexical semantics and distribution of suffixes: A visual analysis. In *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*, LINGVIS '12, pages 7–15. ACL, 2012. URL: `http://aclweb.org/anthology/W12-0202`.

[347] Tobias Ruppert, Andreas Bannach, Jürgen Bernard, Hendrik Lücke-Tieke, Alex Ulmer, and Jörn Kohlhammer. Supporting collaborative political decision making: An interactive policy process visualization system. In *Proceedings of the 9th International Symposium on Visual Information Communication and Interaction*, VINCI '16, pages 104–111. ACM, 2016. `doi:10.1145/2968220.2968223`.

[348] Daniel M. Russell. Simple is good: Observations of visualization use amongst the Big Data digerati. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '16, pages 7–12. ACM, 2016. `doi:10.1145/2909132.2933287`.

[349] James A. Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–294, 1977. `doi:10.1016/0092-6566(77)90037-X`.

[350] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey Ellis, and Daniel A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, December 2014. `doi:10.1109/TVCG.2014.2346481`.

[351] Dominik Sacha, Leishi Zhang, Michael Sedlmair, John A. Lee, Jaakko Peltonen, Daniel Weiskopf, Stephen C. North, and Daniel A. Keim. Visual interaction with dimensionality reduction: A structured literature analysis. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):241–250, January 2017. `doi:10.1109/TVCG.2016.2598495`.

[352] Ramik Sadana, Timothy Major, Alistair Dove, and John Stasko. OnSet: A visualization technique for large-scale binary set data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1993–2002, December 2014. `doi:10.1109/TVCG.2014.2346249`.

[353] Hassan Saif, Yulan He, Miriam Fernandez, and Harith Alani. Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management*, 52(1):5–19, 2016. `doi:10.1016/j.ipm.2015.01.005`.

[354] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523, 1988. `doi:10.1016/0306-4573(88)90021-0`.

[355] Gerard Salton, Anita Wong, and Chung-Shu Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, November 1975. `doi:10.1145/361219.361220`.

[356] Ulrich Schäfer and Christian Spurk. TAKE Scientist's Workbench: Semantic search and citation-based visual navigation in scholar papers. In *Proceedings of the IEEE Fourth International Conference on Semantic Computing*, ISCS '10, pages 317–324. IEEE, 2010. `doi:10.1109/ICSC.2010.40`.

[357] Teri Schamp-Bjerede, Carita Paradis, Andreas Kerren, Magnus Sahlgren, Kostiantyn Kucher, and Afshin Rahimi. Hedges and tweets: Certainty and uncertainty in epistemic markers in microblog feeds. In *Abstracts of the 47th Annual Meeting of the Societas Linguistica Europaea*, SLE '14, page 199, 2014.

[358] Teri Schamp-Bjerede, Carita Paradis, Kostiantyn Kucher, Andreas Kerren, and Magnus Sahlgren. The signifier, signified & stance: A study on the happy/sad emoticon as emotionizers. In *Poster Abstracts of The First Conference of the International Association for Cognitive Semiotics*, IACS 2014, page 219. Lund University, 2014.

[359] Teri Schamp-Bjerede, Carita Paradis, Kostiantyn Kucher, Andreas Kerren, and Magnus Sahlgren. Turning face: Emoticons as reinforcers/attenuators. In *Poster Abstracts of the International Conference on Conceptual Structure, Discourse, and Language*, CSDL '14, 2014.

[360] Arno Scharl, Alexander Hubmann-Haidvogel, Alistair Jones, Daniel Fischl, Ruslan Kamolov, Albert Weichselbraun, and Walter Rafelsberger. Analyzing the public discourse on works of fiction — Detection and visualization of emotion in online coverage about HBO's Game of Thrones. *Information Processing & Management*, 52(1):129–138, 2016. `doi:10.1016/j.ipm.2015.02.003`.

[361] Arno Scharl, Alexander Hubmann-Haidvogel, Albert Weichselbraun, Gerhard Wohlgenannt, Heinz-Peter Lang, and Marta Sabou. Extraction and interactive exploration of knowledge from aggregated news and social media content. In *Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '12, pages 163–168. ACM, 2012. doi:10.1145/2305484.2305511.

[362] Susan Schreibman, Ray Siemens, and John Unsworth. *A New Companion to Digital Humanities*. John Wiley & Sons, 2016.

[363] Dagmar Schuller and Björn W. Schuller. The age of artificial emotional intelligence. *Computer*, 51(9):38–46, September 2018. doi:10.1109/MC.2018.3620963.

[364] Hans-Jörg Schulz. TreeVis.net: A tree visualization reference. *IEEE Computer Graphics and Applications*, 31(6):11–15, November 2011. Accessed February 4, 2019. URL: http://treevis.net, doi:10.1109/MCG.2011.103.

[365] George A.F. Seber and Alan J. Lee. *Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, second edition, 2012.

[366] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, December 2012. doi:10.1109/TVCG.2012.213.

[367] Michael Sedlmair, Tamara Munzner, and Melanie Tory. Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2634–2643, December 2013. doi:10.1109/TVCG.2013.153.

[368] Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, November 2010. doi:10.1109/TVCG.2010.179.

[369] Burr Settles. Closing the loop: Fast, interactive semi-supervised annotation with queries on features and instances. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1467–1478. ACL, 2011. URL: http://aclweb.org/anthology/D11-1136.

[370] Burr Settles. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–114, 2012. doi:10.2200/S00429ED1V01Y201207AIM018.

[371] Azra Shamim, Vimala Balakrishnan, and Muhammad Tahir. Evaluation of opinion visualization techniques. *Information Visualization*, 14(4):339–358, 2015. doi:10.1177/1473871614550537.

[372] Conglei Shi, Weiwei Cui, Shixia Liu, Panpan Xu, Wei Chen, and Huamin Qu. RankExplorer: Visualization of ranking changes in large time series data. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2669–2678, December 2012. doi:10.1109/TVCG.2012.253.

[373] Ben Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Longman Publishing Co., Inc., 1986.

[374] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, VL '96, pages 336–343. IEEE, 1996. doi:10.1109/VL.1996.545307.

[375] Ben Shneiderman and Aleks Aris. Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):733–740, September 2006. doi:10.1109/TVCG.2006.166.

[376] Ayush Shrestha, Ben Miller, Ying Zhu, and Yi Zhao. Storygraph: Extracting patterns from spatio-temporal data. In *Proceedings of the ACM SIGKDD Workshop on Interactive Data Exploration and Analytics*, IDEA '13, pages 95–103. ACM, 2013. doi:10.1145/2501511.2501525.

[377] Harri Siirtola, Tanja Säily, Terttu Nevalainen, and Kari-Jouko Räihä. Text Variation Explorer: Towards interactive visualization tools for corpus linguistics. *International Journal of Corpus Linguistics*, 19(3):417–429, 2014. doi:10.1075/ijcl.19.3.05sii.

[378] Bas Sijtsma, Pernilla Qvarfordt, and Francine Chen. Tweetviz: Visualizing tweets for business intelligence. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, pages 1153–1156. ACM, 2016. doi:10.1145/2911451.2911470.

[379] Artur Šilić and Bojana Dalbelo Bašić. Visualization of Text Streams: A Survey. In Rossitza Setchi, Ivan Jordanov, Robert J. Howlett, and Lakhmi C. Jain, editors, *Knowledge-Based and Intelligent Information and Engineering Systems*, volume 6277 of *Lecture Notes in Computer Science*, pages 31–43. Springer Berlin Heidelberg, 2010. doi:10.1007/978-3-642-15390-7_4.

[380] Shejuti Silvia, Ronak Etemadpour, June Abbas, Sam Huskey, and Chris Weaver. Visualizing variation in classical text with force directed storylines. In *Proceedings of the 1st Workshop on Visualization for the Digital Humanities*, VIS4DH '16, 2016. URL: http://vis4dh.dbvis.de/papers/2016/.

[381] Vasiliki Simaki, Carita Paradis, and Andreas Kerren. Stance classification in texts from blogs on the 2016 British referendum. In *Proceedings of the International Conference on Speech and Computer*, SPECOM '17, pages 700–709. Springer, 2017. doi:10.1007/978-3-319-66429-3_70.

[382] Vasiliki Simaki, Carita Paradis, and Andreas Kerren. Evaluating stance-annotated sentences from the Brexit Blog Corpus: A quantitative linguistic analysis. *ICAME Journal*, 42(1):133–166, 2018. doi:10.1515/icame-2018-0007.

[383] Vasiliki Simaki, Carita Paradis, Maria Skeppstedt, Magnus Sahlgren, Kostiantyn Kucher, and Andreas Kerren. Annotating speaker stance in discourse: The Brexit Blog Corpus. *Corpus Linguistics and Linguistic Theory*, 2017. doi:10.1515/cllt-2016-0060.

[384] Stéfan Sinclair and Geoffrey Rockwell. Text analysis and visualization. In Susan Schreibman, Ray Siemens, and John Unsworth, editors, *A New Companion to Digital Humanities*, pages 274–290. John Wiley & Sons, 2016.

[385] Maria Skeppstedt, Andreas Kerren, and Manfred Stede. Vaccine hesitancy in discussion forums: Computer-assisted argument mining with topic models. *Studies in Health Technology and Informatics*, 247:366–370, 2018. Proceedings of the 29th Medical Informatics Europe Conference (MIE '18) — Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth. doi:10.3233/978-1-61499-852-5-366.

[386] Maria Skeppstedt, Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. Language processing components of the StaViCTA project. In *Proceedings of the Workshop on Logic and Algorithms in Computational Linguistics*, LACompLing '17, pages 137–138, 2017.

[387] Maria Skeppstedt, Kostiantyn Kucher, Manfred Stede, and Andreas Kerren. Topics2Themes: Computer–assisted argument extraction by visual analysis of important topics. In *Proceedings of the 3rd Workshop on Visualization as Added Value in the Development, Use and Evaluation of Language Resources at LREC '18*, VisLR III. ELRA, 2018. URL: http://lrec-conf.org/workshops/lrec2018/W16/summaries/2_W16.html.

[388] Maria Skeppstedt, Carita Paradis, and Andreas Kerren. PAL, a tool for pre-annotation and active learning. *Journal for Language Technology and Computational Linguistics*, 31(1):91–110, 2016. URL: http://jlcl.org/2016_Heft1/jlcl-2016-1-5SkeppstedtEtAl.pdf.

[389] Maria Skeppstedt, Carita Paradis, Andreas Kerren, and Magnus Sahlgren. Finding infrequent phenomena in large corpora using distributional semantics. In *Poster Abstracts of the Symposium on Methods and Linguistic Theories*, MaLT '15, 2015.

[390] Maria Skeppstedt, Magnus Sahlgren, Carita Paradis, and Andreas Kerren. Active learning for detection of stance components. In *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media at COLING '16*, PEOPLES '16, pages 50–59. ACL, 2016. URL: http://aclweb.org/anthology/W16-4306.

[391] Maria Skeppstedt, Magnus Sahlgren, Carita Paradis, and Andreas Kerren. Unshared task: (Dis)agreement in online debates. In *Proceedings of the 3rd Workshop on Argument Mining at ACL '16, short papers track*, ArgMining '16, pages 154–159. ACL, 2016. URL: `http://aclweb.org/anthology/W/W16/W16-2818.pdf`.

[392] Maria Skeppstedt, Teri Schamp-Bjerede, Magnus Sahlgren, Carita Paradis, and Andreas Kerren. Detecting speculations, contrasts and conditionals in consumer reviews. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, WASSA '15, pages 162–168. ACL, 2015. URL: `http://aclweb.org/anthology/W/W15/W15-29.pdf#page=174`.

[393] Maria Skeppstedt, Vasiliki Simaki, Carita Paradis, and Andreas Kerren. Detection of stance and sentiment modifiers in political blogs. In *Proceedings of the International Conference on Speech and Computer*, SPECOM '17, pages 302–311. Springer, 2017. `doi:10.1007/978-3-319-66429-3_29`.

[394] Henry Small. Interpreting maps of science using citation context sentiments: A preliminary investigation. *Scientometrics*, 87(2):373–388, 2011. `doi:10.1007/s11192-011-0349-2`.

[395] Parinaz Sobhani, Diana Inkpen, and Stan Matwin. From argumentation mining to stance classification. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 67–77. ACL, 2015. URL: `http://aclweb.org/anthology/W15-0509`.

[396] Richard Socher, Jeffrey Pennington, Eric Huang, Andrew Ng, and Christopher Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 151–161. ACL, 2011. URL: `http://aclweb.org/anthology/D11-1014`.

[397] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, EMNLP '13, pages 1631–1642. ACL, 2013. URL: `http://aclweb.org/anthology/D13-1170`.

[398] Spark web framework, 2011. Accessed February 4, 2019. URL: `http://sparkjava.com/`.

[399] John Stasko. Value-driven evaluation of visualizations. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, pages 46–53. ACM, 2014. `doi:10.1145/2669557.2669579`.

[400] John Stasko, Carsten Görg, and Zhicheng Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008. `doi:10.1057/palgrave.ivs.9500180`.

[401] Chad A. Steed, Justin Beaver, Paul L. Bogen II, Margaret Drouhard, and Joshua Pyle. Text stream trend analysis using multiscale visual analytics with applications to social media systems. In *Proceedings of the IUI Workshop on Visual Text Analytics*, TextVis '15, 2015. URL: `http://vialab.science.uoit.ca/textvis2015/`.

[402] Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. BRAT: A web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 102–107. ACL, 2012. URL: `http://dl.acm.org/citation.cfm?id=2380921.2380942`.

[403] Florian Stoffel, Wolfgang Jentner, Michael Behrisch, Johannes Fuchs, and Daniel A. Keim. Interactive ambiguity resolution of named entities in fictional literature. *Computer Graphics Forum*, 36(3):189–200, 2017. `doi:10.1111/cgf.13179`.

[404] Carlo Strapparava and Alessandro Valitutti. WordNet Affect: An affective extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, volume 4 of *LREC '04*, pages 1083–1086. ELRA, 2004. URL: `http://lrec-conf.org/proceedings/lrec2004/pdf/369.pdf`.

[405] Hendrik Strobelt, Daniela Oelke, Bum Chul Kwon, Tobias Schreck, and Hanspeter Pfister. Guidelines for effective usage of text highlighting techniques. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):489–498, January 2016. doi:10.1109/TVCG.2015.2467759.

[406] Hendrik Strobelt, Daniela Oelke, Christian Rohrdantz, Andreas Stoffel, Daniel A. Keim, and Oliver Deussen. Document Cards: A top trumps visualization for documents. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1145–1152, November 2009. doi:10.1109/TVCG.2009.139.

[407] Pero Subasic and Alison Huettner. Affect analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy Systems*, 9(4):483–496, August 2001. doi:10.1109/91.940962.

[408] Ching-Ying Sung, Xun-Yi Huang, Yicong Shen, Fu-Yin Cherng, Wen-Chieh Lin, and Hao-Chuan Wang. ToPIN: A visual analysis tool for time-anchored comments in online educational videos. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 2185–2191. ACM, 2016. doi:10.1145/2851581.2892327.

[409] Jan Svartvik. Corpus linguistics 25+ years on. In Roberta Facchinetti, editor, *Corpus Linguistics 25 Years On*, pages 11–25. Rodopi, 2007.

[410] Yuzuru Tanahashi and Kwan-Liu Ma. Design considerations for optimizing storyline visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2679–2688, December 2012. doi:10.1109/TVCG.2012.212.

[411] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. Learning sentiment-specific word embedding for Twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ACL '14, pages 1555–1565. ACL, 2014. URL: http://aclweb.org/anthology/P14-1146.

[412] Tugba Taskaya and Khurshid Ahmad. Bimodal visualisation: A financial trading case study. In *Proceedings of the Seventh International Conference on Information Visualization*, IV '03, pages 320–326. IEEE, 2003. doi:10.1109/IV.2003.1217997.

[413] Annie Tat and Sheelagh Carpendale. CrystalChat: Visualizing personal chat history. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, volume 3 of *HICSS '06*. IEEE, 2006. doi:10.1109/HICSS.2006.107.

[414] Yla R. Tausczik and James W. Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010. doi:10.1177/0261927X09351676.

[415] Richard Taylor. Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, 6(1):35–39, 1990. doi:10.1177/875647939000600106.

[416] Martijn Tennekes and Edwin de Jonge. Tree Colors: Color schemes for tree-structured data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2072–2081, December 2014. doi:10.1109/TVCG.2014.2346277.

[417] James J. Thomas and Kristin A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE, 2005.

[418] Geoff Thompson and Susan Hunston, editors. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford University Press, 2000.

[419] Milan Tofiloski, Lydia Odilinye, Jasneet Sabharwal, and Fred Popowich. A taxonomy of the target-problem space for visual text analytics. In *Proceedings of the IUI Workshop on Visual Text Analytics*, TextVis '15, 2015. URL: http://vialab.science.uoit.ca/textvis2015/.

[420] Katrin Tomanek, Joachim Wermter, and Udo Hahn. Efficient annotation with the Jena ANnotation Environment (JANE). In *Proceedings of the Linguistic Annotation Workshop*, LAW '07, pages 9–16. ACL, 2007. URL: http://dl.acm.org/citation.cfm?id=1642059.1642061.

[421] Christian Tominski and Wolfgang Aigner. The TimeVis Browser. Accessed February 4, 2019. URL: http://survey.timeviz.net.

[422] Simon Tong and Daphne Koller. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2(November):45–66, 2001. URL: `http://jmlr.org/papers/v2/tong01a.html`.

[423] Kamil Topal and Gultekin Ozsoyoglu. Emotional classification and visualization of movies based on their IMDb reviews. *Information Discovery and Delivery*, 45(3):149–158, 2017. `doi:10.1108/IDD-05-2017-0045`.

[424] Megan K. Torkildson. Visualizing the performance of classification algorithms with additional re-annotated data. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 2767–2772. ACM, 2013. `doi:10.1145/2468356.2479507`.

[425] Megan K. Torkildson, Kate Starbird, and Cecilia Aragon. Analysis and visualization of sentiment and emotion on crisis tweets. In Yuhua Luo, editor, *Proceedings of the 11th International Conference on Cooperative Design, Visualization, and Engineering (CDVE '14)*, volume 8683 of *Lecture Notes in Computer Science*, pages 64–67. Springer International Publishing, 2014. URL: `http://dx.doi.org/10.1007/978-3-319-10831-5_9`, `doi:10.1007/978-3-319-10831-5_9`.

[426] Melanie Tory and Torsten Möller. Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications*, 25(5):8–11, September 2005. `doi:10.1109/MCG.2005.102`.

[427] J. Gregory Trafton, Susan S. Kirschenbaum, Ted L. Tsui, Robert T. Miyamoto, James A. Ballas, and Paula D. Raymond. Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human-Computer Studies*, 53(5):827–850, 2000. `doi:10.1006/ijhc.2000.0419`.

[428] Trump2Cash: A stock trading bot powered by Trump tweets, 2017. Accessed February 4, 2019. URL: `https://github.com/maxbbraun/trump2cash`.

[429] Nikos Tsirakis, Vasilis Poulopoulos, Panagiotis Tsantilas, and Iraklis Varlamis. Large scale opinion mining for social, news and blog data. *Journal of Systems and Software*, 2016. `doi:10.1016/j.jss.2016.06.012`.

[430] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. Mining multi-label data. In Oded Maimon and Lior Rokach, editors, *Data Mining and Knowledge Discovery Handbook*, pages 667–685. Springer US, 2010. `doi:10.1007/978-0-387-09823-4_34`.

[431] Mikalai Tsytsarau and Themis Palpanas. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3):478–514, May 2012. `doi:10.1007/s10618-011-0238-6`.

[432] Edward R. Tufte. *Beautiful Evidence*. Graphics Press, 2006.

[433] John W. Tukey. *Exploratory Data Analysis*. Addison-Wesley Publishing Company, 1977.

[434] Peter D. Turney. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424. ACL, 2002. `doi:10.3115/1073083.1073153`.

[435] Peter D. Turney and Patrick Pantel. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188, 2010. `doi:10.1613/jair.2934`.

[436] Twitter's API, 2006. Accessed February 4, 2019. URL: `https://developer.twitter.com/en/docs`.

[437] Maria Ulan, Sebastian Hönel, Rafael M. Martins, Morgan Ericsson, Welf Löwe, Anna Wingkvist, and Andreas Kerren. Quality models inside out: Interactive visualisation of software metrics by means of joint probabilities. In *Proceedings of the 6th IEEE Working Conference on Software Visualization*, VISSOFT '18, pages 65–75. IEEE, 2018. `doi:10.1109/VISSOFT.2018.00015`.

[438] Omar Valdiviezo, J. Alfredo Sánchez, and Ofelia Cervantes. Visualizing sentiment change in social networks. In *Proceedings of the 8th Latin American Conference on Human-Computer Interaction*, CLIHC '17, pages 1:1–1:7. ACM, 2017. `doi:10.1145/3151470.3151475`.

[439] Maarten van Meersbergen, Piek Vossen, Janneke van der Zwaan, Antske Fokkens, Willem van Hage, Inger Leemans, and Isa Maks. Storyteller: Visual analytics of perspectives on rich text interpretations. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, pages 37–45. ACL, 2017. URL: `http://aclweb.org/anthology/W17-4207`.

[440] Jarke J. Van Wijk. Bridging the gaps. *IEEE Computer Graphics and Applications*, 26(6):6–9, November 2006. `doi:10.1109/MCG.2006.120`.

[441] Jarke J. Van Wijk. Views on visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):421–432, July 2006. `doi:10.1109/TVCG.2006.80`.

[442] Erik Velldal. Predicting speculation: A simple disambiguation approach to hedge detection in biomedical literature. *Journal of Biomedical Semantics*, 2(5):1–14, 2011. `doi:10.1186/2041-1480-2-S5-S7`.

[443] Arie Verhagen. *Constructions of Intersubjectivity: Discourse, Syntax, and Cognition*. Oxford University Press, 2005. `doi:10.1093/acprof:oso/9780199226702.001.0001`.

[444] Fernanda B. Viégas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. ManyEyes: A site for visualization at Internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, November 2007. `doi:10.1109/TVCG.2007.70577`.

[445] Marilyn A. Walker, Pranav Anand, Rob Abbott, Jean E. Fox Tree, Craig Martell, and Joseph King. That is your evidence?: Classifying stance in online political debate. *Decision Support Systems*, 53(4):719–729, 2012. `doi:10.1016/j.dss.2012.05.032`.

[446] Emily Wall, Meeshu Agnihotri, Laura Matzen, Kristin Divis, Michael Haass, Alex Endert, and John Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, January 2019. `doi:10.1109/TVCG.2018.2865146`.

[447] Changbo Wang, Zhao Xiao, Yuhua Liu, Yanru Xu, Aoying Zhou, and Kang Zhang. SentiView: Sentiment analysis and visualization for Internet popular topics. *IEEE Transactions on Human-Machine Systems*, 43(6):620–630, November 2013. `doi:10.1109/THMS.2013.2285047`.

[448] Florence Ying Wang, Arnaud Sallaberry, Karsten Klein, Masahiro Takatsuka, and Mathieu Roche. SentiCompass: Interactive visualization for exploring and comparing the sentiments of time-varying Twitter data. In *Proceedings of the 8th IEEE Pacific Visualization Symposium*, PacificVis '15, pages 129–133. IEEE, 2015. `doi:10.1109/PACIFICVIS.2015.7156368`.

[449] Ji Wang, Kyle D. Dent, and Chris L. North. Fisheye word cloud for temporal sentiment exploration. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 1767–1772. ACM, 2013. `doi:10.1145/2468356.2468673`.

[450] Lan Wang and Yuan Wan. Sentiment classification of documents based on latent semantic analysis. In Song Lin and Xiong Huang, editors, *Advanced Research on Computer Education, Simulation and Modeling*, volume 176 of *Communications in Computer and Information Science*, pages 356–361. Springer, 2011. `doi:10.1007/978-3-642-21802-6_57`.

[451] Sida Wang and Christopher D. Manning. Baselines and bigrams: Simple, good sentiment and topic classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers — Volume 2*, ACL '12, pages 90–94. ACL, 2012. URL: `http://anthology.aclweb.org/P/P12/P12-2018.pdf`.

[452] Xiting Wang, Shixia Liu, Yang Chen, Tai-Quan Peng, Jing Su, Jing Yang, and Baining Guo. How ideas flow across multiple social groups. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '16, pages 51–60. IEEE, 2016. `doi:10.1109/VAST.2016.7883511`.

[453] Yong Wang, Hammad Haleem, Conglei Shi, Yanhong Wu, Xun Zhao, Siwei Fu, and Huamin Qu. Towards easy comparison of local businesses using online reviews. *Computer Graphics Forum*, 37(3):63–74, June 2018. `doi:10.1111/cgf.13401`.

[454] Yun Wang, Dongyu Liu, Huamin Qu, Qiong Luo, and Xiaojuan Ma. A guided tour of literature review: Facilitating academic paper reading with narrative visualization. In *Proceedings of the 9th International Symposium on Visual Information Communication and Interaction*, VINCI '16, pages 17–24. ACM, 2016. doi:10.1145/2968220.2968242.

[455] Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '00, pages 110–119. ACM, 2000. doi:10.1145/345513.345271.

[456] Franz Wanner, Thomas Ramm, and Daniel A. Keim. ForAVis: Explorative user forum analysis. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 14:1–14:10. ACM, 2011. doi:10.1145/1988688.1988705.

[457] Franz Wanner, Christian Rohrdantz, Florian Mansmann, Daniela Oelke, and Daniel A. Keim. Visual sentiment analysis of RSS news feeds featuring the US presidential election in 2008. In *Proceedings of the IUI Workshop on Visual Interfaces to the Social and the Semantic Web*, VISSW '09. CEUR-WS.org, 2009. URL: http://ceur-ws.org/Vol-443.

[458] Franz Wanner, Andreas Stoffel, Dominik Jäckle, Bum Chul Kwon, Andreas Weiler, and Daniel A. Keim. State-of-the-art report of visual analysis for event detection in text data streams. In *Proceedings of the Eurographics Conference on Visualization — STARs*, EuroVis '14, pages 125–139. The Eurographics Association, 2014. doi:10.2312/eurovisstar.20141176.

[459] Franz Wanner, Andreas Weiler, and Tobias Schreck. Topic Tracker: Shape-based visualization for trend and sentiment tracking in Twitter. In *Proceedings of the 2nd IEEE Workshop on Interactive Visual Text Analytics "Task-Driven Analysis of Social Media"*, TextVis '12, 2012. URL: http://vialab.science.uoit.ca/textvis2012/.

[460] Matthew Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3–4):194–210, 2002. doi:10.1057/palgrave.ivs.9500025.

[461] Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., 2004.

[462] Malcolm Ware, Eibe Frank, Geoffrey Holmes, Mark Hall, and Ian H. Witten. Interactive machine learning: Letting users build classifiers. *International Journal of Human-Computer Studies*, 55(3):281–292, 2001. doi:10.1006/ijhc.2001.0499.

[463] Massimo Warglien and Peter Gärdenfors. Semantics, conceptual spaces, and the meeting of minds. *Synthese*, 190(12):2165–2193, 2013. doi:10.1007/s11229-011-9963-z.

[464] Martin Wattenberg. Arc diagrams: Visualizing structure in strings. In *Proceedings of the IEEE Symposium on Information Visualization*, INFOVIS '02, pages 110–116. IEEE, 2002. doi:10.1109/INFVIS.2002.1173155.

[465] Martin Wattenberg and Fernanda B. Viégas. The Word Tree, an interactive visual concordance. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1221–1228, November 2008. doi:10.1109/TVCG.2008.172.

[466] Gunther H. Weber, Sheelagh Carpendale, David Ebert, Brian Fisher, Hans Hagen, Ben Shneiderman, and Anders Ynnerman. Apply or die: On the role and assessment of application papers in visualization. *IEEE Computer Graphics and Applications*, 37(3):96–104, May 2017. doi:10.1109/MCG.2017.51.

[467] Alan J. Wecker, Joel Lanir, Osnat Mokryn, Einat Minkov, and Tsvi Kuflik. Semantize: Visualizing the sentiment of individual document. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, AVI '14, pages 385–386. ACM, 2014. doi:10.1145/2598153.2600056.

[468] Stephen Wehrend and Clayton Lewis. A problem-oriented classification of visualization techniques. In *Proceedings of the First IEEE Conference on Visualization*, Visualization '90, pages 139–143. IEEE, 1990. doi:10.1109/VISUAL.1990.146375.

[469] Andreas Weiler, Michael Grossniklaus, and Marc H. Scholl. The Stor-e-Motion visualization for topic evolution tracking in text data streams. In *Proceedings of the 6th International Conference on Information Visualization Theory and Applications — Volume 1: IVAPP*, VISIGRAPP '15, pages 29–39. SciTePress, 2015. doi:10.5220/0005292900290039.

[470] April M. Wensel and Sara O. Sood. VIBES: Visualizing changing emotional states in personal stories. In *Proceedings of the 2nd ACM International Workshop on Story Representation, Mechanism and Context*, SRMC '08, pages 49–56. ACM, 2008. doi:10.1145/1462014.1462025.

[471] Wesley Willett, Jeffrey Heer, and Maneesh Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, November 2007. doi:10.1109/TVCG.2007.70589.

[472] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 347–354. ACL, 2005. doi:10.3115/1220575.1220619.

[473] James A. Wise, James J. Thomas, Kelly Pennock, David Lantrip, Marc Pottier, Anne Schur, and Vern Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proceedings of the IEEE Symposium on Information Visualization*, InfoVis '95, pages 51–58. IEEE, 1995. doi:10.1109/INFVIS.1995.528686.

[474] William Wright, David Schroh, Pascale Proulx, Alex Skaburskis, and Brian Cort. The sandbox for analysis: Concepts and methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 801–810. ACM, 2006. doi:10.1145/1124772.1124890.

[475] William Wright, David Sheffield, and Stephanie Santosa. Argument mapper: Countering cognitive biases in analysis with critical (visual) thinking. In *Proceedings of the 21st International Conference on Information Visualisation*, IV '17, pages 250–255. IEEE, July 2017. doi:10.1109/iV.2017.69.

[476] Jiayu Wu, Zhiyong Fu, Zhiyuan Liu, Xu Lin, Jiayu Tang, Jiajia Pan, and Chen Zhao. Creating reflections in public emotion visualization: Prototype exploration on traffic theme. In *Proceedings of the 9th ACM Conference on Creativity & Cognition*, C&C '13, pages 357–361. ACM, 2013. doi:10.1145/2466627.2466671.

[477] Jiayu Wu, Zhiyong Fu, Zhiyuan Liu, Jiajia Pan, Huiling Long, Xu Lin, Haoqing He, Xinxiong Chen, and Jiayu Tang. City Flow: Prototype exploration for visualizing urban traffic conversations. In *Proceedings of the International Conference on Privacy, Security, Risk and Trust and the International Conference on Social Computing*, SocialCom/PASSAT 2012, pages 481–489. IEEE, 2012. doi:10.1109/SocialCom-PASSAT.2012.40.

[478] Yingcai Wu, Nan Cao, David Gotz, Yap-Peng Tan, and Daniel A. Keim. A survey on visual analytics of social media data. *IEEE Transactions on Multimedia*, 18(11):2135–2148, November 2016. doi:10.1109/TMM.2016.2614220.

[479] Yingcai Wu, Zhutian Chen, Guodao Sun, Xiao Xie, Nan Cao, Shixia Liu, and Weiwei Cui. StreamExplorer: A multi-stage system for visually exploring events in social streams. *IEEE Transactions on Visualization and Computer Graphics*, 24(10):2758–2772, October 2018. doi:10.1109/TVCG.2017.2764459.

[480] Yingcai Wu, Shixia Liu, Kai Yan, Mengchen Liu, and Fangzhao Wu. OpinionFlow: Visual analysis of opinion diffusion on social media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1763–1772, December 2014. doi:10.1109/TVCG.2014.2346920.

[481] Yingcai Wu, Furu Wei, Shixia Liu, Norman Au, Weiwei Cui, Hong Zhou, and Huamin Qu. OpinionSeer: Interactive visualization of hotel customer feedback. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1109–1118, November 2010. doi:10.1109/TVCG.2010.183.

[482] Jin Xu, Yubo Tao, Hai Lin, Rongjie Zhu, and Yuyu Yan. Exploring controversy via sentiment divergences of aspects in reviews. In *Proceedings of the IEEE Pacific Visualization Symposium*, PacificVis '17, pages 240–249, 2017. `doi:10.1109/PACIFICVIS.2017.8031600`.

[483] Kaiquan Xu, Stephen Shaoyi Liao, Jiexun Li, and Yuxia Song. Mining comparative opinions from customer reviews for competitive intelligence. *Decision Support Systems*, 50(4):743–754, 2011. `doi:10.1016/j.dss.2010.08.021`.

[484] Panpan Xu, Yingcai Wu, Enxun Wei, Tai-Quan Peng, Shixia Liu, Jonathan J.H. Zhu, and Huamin Qu. Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2012–2021, December 2013. `doi:10.1109/TVCG.2013.221`.

[485] Koji Yatani, Michael Novati, Andrew Trusty, and Khai N. Truong. Review Spotlight: A user interface for summarizing user-generated reviews using adjective-noun word pairs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1541–1550. ACM, 2011. `doi:10.1145/1978942.1979167`.

[486] Seid Muhie Yimam, Iryna Gurevych, Richard Eckart de Castilho, and Chris Biemann. WebAnno: A flexible, web-based and visually supported system for distributed annotations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics — Conference System Demonstrations*, ACL '13, pages 1–6. ACL, 2013. URL: `http://aclweb.org/anthology/P/P13/P13-4.pdf`.

[487] Beth Yost and Chris North. Single complex glyphs versus multiple simple glyphs. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1889–1892. ACM, 2005. `doi:10.1145/1056808.1057048`.

[488] Zhiwen Yu, Zhitao Wang, Liming Chen, Bin Guo, and Wenjie Li. Featuring, detecting, and visualizing human sentiment in Chinese micro-blog. *ACM Transactions on Knowledge Discovery from Data*, 10(4):48:1–48:23, May 2016. `doi:10.1145/2821513`.

[489] Chenghai Zhang, Yuhua Liu, and Changbo Wang. Time-space varying visual analysis of micro-blog sentiment. In *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction*, VINCI '13, pages 64–71. ACM, 2013. `doi:10.1145/2493102.2493110`.

[490] Jianwei Zhang, Yukiko Kawai, Tadahiko Kumamoto, and Katsumi Tanaka. A novel visualization method for distinction of web news sentiment. In Gottfried Vossen, Darrell D. E. Long, and Jeffrey Xu Yu, editors, *Proceedings of the 10th International Conference on Web Information Systems Engineering (WISE '09)*, volume 5802 of *Lecture Notes in Computer Science*, pages 181–194. Springer Berlin Heidelberg, 2009. `doi:10.1007/978-3-642-04409-0_22`.

[491] Jian Zhao, Nan Cao, Zhen Wen, Yale Song, Yu-Ru Lin, and Christopher Collins. #FluxFlow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1773–1782, December 2014. `doi:10.1109/TVCG.2014.2346922`.

[492] Jian Zhao, Liang Gou, Fei Wang, and Michelle Zhou. PEARL: An interactive visual analytic tool for understanding personal emotion style derived from social media. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '14, pages 203–212. IEEE, 2014. `doi:10.1109/VAST.2014.7042496`.

[493] Jichang Zhao, Li Dong, Junjie Wu, and Ke Xu. MoodLens: An emoticon-based sentiment analysis system for Chinese tweets. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 1528–1531. ACM, 2012. `doi:10.1145/2339530.2339772`.

[494] Shuai Zhao, Yunhai Tong, Xinhai Liu, and Shaohua Tan. Correlating Twitter with the stock market through non-Gaussian SVAR. In *Proceedings of the Eighth International Conference on Advanced Computational Intelligence*, ICACI 2016, pages 257–264. IEEE, 2016. `doi:10.1109/ICACI.2016.7449835`.

[495] Yanyan Zhao, Bing Qin, Ting Liu, and Duyu Tang. Social sentiment sensor: A visualization system for topic detection and topic sentiment analysis on microblog. *Multimedia Tools and Applications*, 75(15):8843–8860, 2016. doi:10.1007/s11042-014-2184-y.