DoSVis: Document Stance Visualization

Kostiantyn Kucher[1], Carita Paradis[1] and Andreas Kerren[1]
[1]*Department of Computer Science, Linnaeus University, Växjö, Sweden*
[2]*Centre for Languages and Literature, Lund University, Lund, Sweden*
{*kostiantyn.kucher, andreas.kerren*}*@lnu.se, carita.paradis@englund.lu.se*

Keywords: Stance Visualization, Sentiment Visualization, Text Visualization, Stance Analysis, Sentiment Analysis, Text Analytics, Information Visualization, Interaction.

Abstract: Text visualization techniques often make use of automatic text classification methods. One of such methods is stance analysis, which is concerned with detecting various aspects of the writer's attitude towards utterances expressed in the text. Existing text visualization approaches for stance classification results are usually adapted to textual data consisting of individual utterances or short messages, and they are often designed for social media or debate monitoring tasks. In this paper, we propose a visualization approach called DoSVis (Document Stance Visualization) that focuses instead on individual text documents of a larger length. DoSVis provides an overview of multiple stance categories detected by our classifier at the utterance level as well as a detailed text view annotated with classification results, thus supporting both distant and close reading tasks. We describe our approach by discussing several application scenarios involving business reports and works of literature.

1 INTRODUCTION

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 (Jänicke et al., 2015; Kucher and Kerren, 2015). 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 (Pang and Lee, 2008; Mohammad, 2016). There is also research on sentiment analysis of business reports and CEO letters which studies the relation between the language and financial indicators (Kearney and Liu, 2014; Nopp and Hanbury, 2015). The existing sentiment visualization techniques for textual data support a variety of data domains, data source types, and user tasks (Kucher et al., 2017a). At the same time, few existing visualization techniques make use of another method related to sentiment analysis—stance analysis (Mohammad et al., 2016; Skeppstedt et al., 2016b; Simaki et al., 2017b). 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[1] 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 focused on political text data such as transcripts of debates (El-Assady et al., 2016), blog posts and comments (Kucher et al., 2016a; Kucher et al., 2016b), and tweets (Mohammad et al., 2016; Martins et al., 2017).

In this paper, 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

[1]Advances in the description and explanation of Stance in discourse using Visual and Computational Text Analytics (`http://cs.lnu.se/stavicta/`).

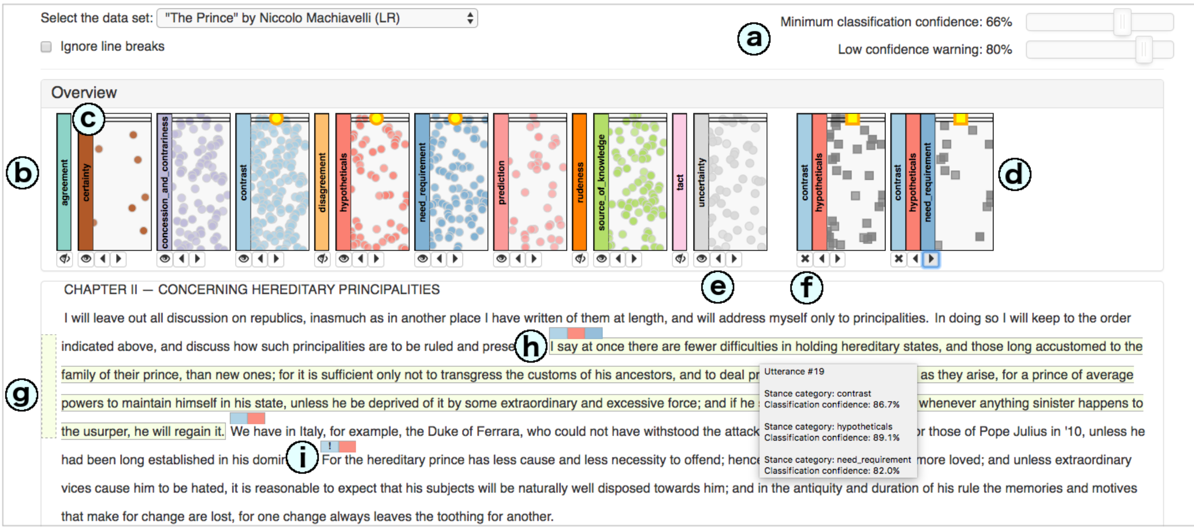


Figure 1: Visualization of a 16th century political treatise "The Prince" by Niccolò Machiavelli in our tool DoSVis: (a) sliders used for global filtering and toggling of warning symbols; (b) scatterplot-like overviews based on the detected occurrences of stance categories in the text; (c) a viewport 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.

such as social media posts, we look into scenarios involving exploration of longer documents such as business reports (Kearney and Liu, 2014) and works of literature (Sinclair and Rockwell, 2016). 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 Figure 1). The main contributions of this paper 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 article is organized as follows. In the next section, 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 3. We illustrate the applicability of our approach with several use cases in Section 4 and discuss some aspects of our findings in Section 5. Finally, we conclude this article in Section 6.

2 RELATED WORK

2.1 Stance Analysis and Visualization

A more conservative 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 (Skeppstedt et al., 2016b; Mohammad et al., 2016). The latter work describes the results of a stance analysis contest 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 (El-Assady et al., 2016). 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* (Kucher et al., 2016b) or *speculation* and *condition* (Skeppstedt et al., 2016a). Kucher et al. describe a visualization of their stance data set with a tool called ALVA (Kucher et al., 2016a; Kucher et al., 2017b). Similar to the other stance visualizations discussed above, ALVA focuses on the overview of a data set or corpus consisting of multiple utterances or sentences from blog posts and comments. Finally, StanceXplore (Martins et al., 2017) provides multiple coordinated views for exploratory visual analysis of a corpus of tweets labelled with multiple stance categories by a stance classifier. In contrast to all these works, our contribution proposed in this paper is designed for a detailed exploration of individual documents which are much larger/longer than social media posts.

2.2 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 (Jänicke et al., 2015) or text streams (Kucher and Kerren, 2015; Kucher et al., 2017a), 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 (Drucker, 2016). Providing an overview of the content of individual documents dates back to early techniques, such as SeeSoft (Eick et al., 1992) and TileBars (Hearst, 1995). Both provide pixel-based summaries for text segments constituting the documents. Affect Color Bar (Liu et al., 2003) 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 (Abbasi and Chen, 2007) 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 (Keim and Oelke, 2007). VarifocalReader (Koch et al., 2014) supports both distant and close reading (see (Jänicke et al., 2015), for example) by using topic segmentation, overview of text structure, and highlighting of automatically annotated words or chunks. Lexical Episode Plots (Gold et al., 2015) provide an overview of topics recurring throughout a text (more specifically, a transcript of political debates). uVSAT (Kucher et al., 2016b) uses scatterplot-like representations for overviews of stance markers detected in a text document. 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 (Chandrasegaran et al., 2017). 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 (Strobelt et al., 2016) to represent custom annotations or labels. For example, Ink Blots (Abbasi and Chen, 2007) highlight an approximate region based on the position of certain marker words or features. Serendip (Alexander et al., 2014) highlights words relevant to specific topics. uVSAT (Kucher et al., 2016b) 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 (Chandrasegaran et al., 2017). 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.

3 VISUALIZATION METHODOLOGY

The input data for our tool DoSVis is generated by a stance classifier pipeline currently developed by our project members (Kucher et al., 2016a; Kucher et al., 2017b; Simaki et al., 2017b; Skeppstedt et al., 2017a). The pipeline (see an illustration in Figure 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

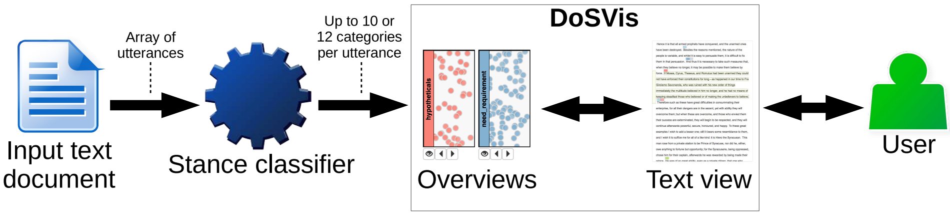


Figure 2: The architecture of our approach. DoSVis uses the output of the stance classifier for a text document divided into utterances. Each utterance may be simultaneously labelled with multiple stance categories.

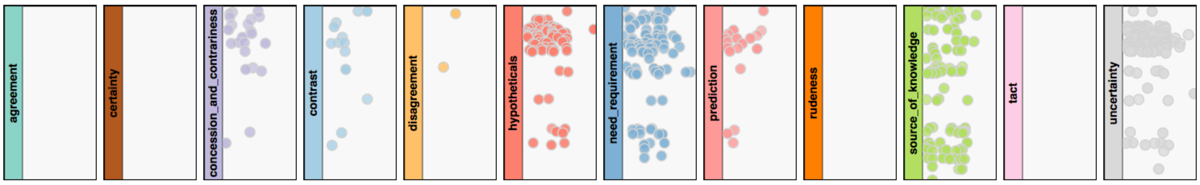
approach can actually be generalized to any set of categories or labels associated with utterances. We have tested this by using two versions of the stance classifier: (1) an SVM-based classifier with 10 stance categories (Kucher et al., 2017b), and (2) a logistic regression (LR)-based classifier with 12 stance categories (Skeppstedt et al., 2017a). Both of these classifiers also provide a form of confidence estimates for the classification decisions based on (1) Platt scaling (Platt, 1999) and (2) probability estimates (Hosmer et al., 2013), 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 labelled 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 using JavaScript and D3 (D3, 2011). Its user interface depicted in Figure 1 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 can be preferable in case of some documents converted from the PDF format (see Section 5). The sliders located at the top right (see Figure 1(a)) specify the classification confidence thresholds for displaying the classification results at all and displaying warning symbols (exclamation marks within the glyphs, see Figure 1(i)), respectively, in order to help the users focus on more reliable results.

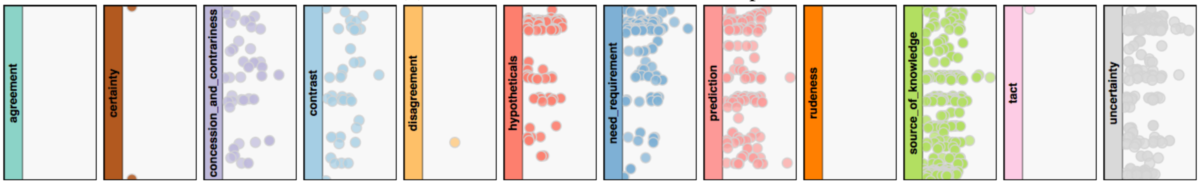
The overview of stance classification results consists of scatterplot-like representations for individual stance categories displayed in Figure 1(b). We have decided to follow this design with separate representations for categories due to the data considerations described above. Any utterance in our data can potentially be labelled 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 (Kucher et al., 2017b; Martins et al., 2017). 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 (ColorBrewer, 2009). 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 1(c)). Each overview supports details on demand and navigation over the text by hovering and clicking, respectively. The users can also hide the overview plots and navigate to the previous/next occurrence of the corresponding stance category by using the buttons located under each plot (see Figure 1(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 labelled with the corresponding combination of categories. Such plots for the combinations of two and three categories, respectively, are displayed in Figure 1(d). In order to distinguish such combination plots from regular category overview plots, we have used rectangular markers with a dark grey color. The opacity mapping and global filtering behaviour 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 1(e+f)).

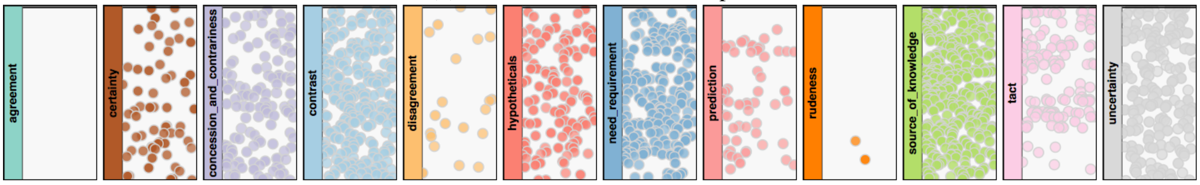
DoSVis also provides a detailed text view (displayed in Figure 1(g)) with stance category labels and details on demand, thus supporting both distant and close reading approaches (Jänicke et al., 2015). We use sets of non-intrusive rectangular glyphs located above utterances to represent the categories detected by the classifier (see Figure 1(h)). These glyphs share



(a) Tableau Software 2015 annual report.



(b) Yahoo Inc. 2015 annual report.



(c) “The Hound of the Baskervilles” by Arthur Conan Doyle.

Figure 3: Overviews of stance categories detected in several documents with the LR classifier at 66% classification confidence.

the color coding, opacity mapping, and filtering behaviour with the overview plots. They are also connected with linking and brushing—see the elements highlighted in yellow in Figure 1(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 1(i). Such marks are displayed for the classification results with confidence values lower than the global threshold controlled by the corresponding slider.

4 USE CASES

As mentioned in Section 1, 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 (Kearney and Liu, 2014; Nopp and Hanbury, 2015). 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 3(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 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 & 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 (Schreibman et al., 2016) could make use of the support for distant and close reading provided by DoSVis. Figure 3(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 (Keim and Oelke, 2007). 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 (Simaki et al., 2017a; Skeppstedt et al., 2017b) discuss co-occurrences of such stance categories as *prediction* with *uncertainty* and *hypotheticals* with *uncertainty*, respectively, in political blog data. Figure 4 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.

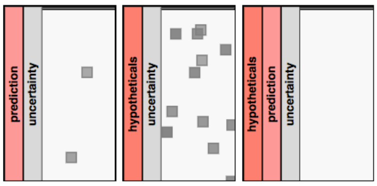


Figure 4: Overviews of several stance category combinations detected for the data in Figure 3(c).

5 DISCUSSION

Stance classification The existing methods of automatic stance classification do not reach the same levels of precision/accuracy (Mohammad et al., 2016) as, for instance, sentiment classification methods, especially for topic-independent tasks (Skeppstedt et al., 2016a). 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.

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 (Constantin et al., 2013). 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 (Mao et al., 2003; Strobelt et al., 2009).

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 (Asokarajan et al., 2016; Asokarajan et al., 2017). 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.

6 CONCLUSIONS AND FUTURE WORK

In this paper, 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.

ACKNOWLEDGEMENTS

This research was funded by the framework grant "The Digitized Society—Past, Present, and Future" with No. 2012-5659 from the Swedish Research Council.

REFERENCES

Abbasi, A. and Chen, H. (2007). 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.

Alexander, E., Kohlmann, J., Valenza, R., Witmore, M., and Gleicher, M. (2014). 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.

Asokarajan, B., Etemadpour, R., Abbas, J., Huskey, S., and Weaver, C. (2016). 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.

Asokarajan, B., Etemadpour, R., Abbas, J., Huskey, S., and Weaver, C. (2017). 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.

Chandrasegaran, S., Badam, S. K., Kisselburgh, L., Ramani, K., and Elmqvist, N. (2017). Integrating visual analytics support for grounded theory practice in qualitative text analysis. *Computer Graphics Forum*, 36(3):201–212.

ColorBrewer (2009). ColorBrewer 2.0 — color advice for cartography. http://colorbrewer2.org/. Accessed October 31, 2017.

Constantin, A., Pettifer, S., and Voronkov, A. (2013). PDFX: Fully-automated PDF-to-XML conversion of scientific literature. In *Proceedings of the ACM Symposium on Document Engineering*, DocEng '13, pages 177–180, New York, NY, USA. ACM.

D3 (2011). D3 — data-driven documents. http://d3js.org/. Accessed October 31, 2017.

Drucker, J. (2016). Graphical approaches to the digital humanities. In Schreibman, S., Siemens, R., and Unsworth, J., editors, *A New Companion to Digital Humanities*, pages 238–250. John Wiley & Sons.

Eick, S. G., Steffen, J. L., and Sumner, E. E. (1992). Seesoft—A tool for visualizing line oriented software statistics. *IEEE Transactions on Software Engineering*, 18(11):957–968.

El-Assady, M., Gold, V., Acevedo, C., Collins, C., and Keim, D. (2016). ConToVi: Multi-party conversation exploration using topic-space views. *Computer Graphics Forum*, 35(3):431–440.

Gold, V., Rohrdantz, C., and El-Assady, M. (2015). Exploratory text analysis using lexical episode plots. In *Short Papers of the EG/VGTC Conference on Visualization*, EuroVis '15. The Eurographics Association.

Hearst, M. A. (1995). 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.

Hosmer, Jr., D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons, Inc.

Jänicke, S., Franzini, G., Cheema, M. F., and Scheuermann, G. (2015). 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.

Kearney, C. and Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33:171–185.

Keim, D. A. and Oelke, D. (2007). 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.

Koch, S., John, M., Wörner, M., Müller, A., and Ertl, T. (2014). VarifocalReader — In-depth visual analysis of large text documents. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1723–1732.

Kucher, K. and Kerren, A. (2015). Text visualization techniques: Taxonomy, visual survey, and community insights. In *Proceedings of the 8th IEEE Pacific Visualization Symposium*, PacificVis '15, pages 117–121. IEEE.

Kucher, K., Kerren, A., Paradis, C., and Sahlgren, M. (2016a). 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.

Kucher, K., Paradis, C., and Kerren, A. (2017a). The state of the art in sentiment visualization. *Computer Graphics Forum*.

Kucher, K., Paradis, C., Sahlgren, M., and Kerren, A. (2017b). Active learning and visual analytics for stance classification with ALVA. *ACM Transactions on Interactive Intelligent Systems*, 7(3):14:1–14:31.

Kucher, K., Schamp-Bjerede, T., Kerren, A., Paradis, C., and Sahlgren, M. (2016b). Visual analysis of online social media to open up the investigation of stance phenomena. *Information Visualization*, 15(2):93–116.

Liu, H., Selker, T., and Lieberman, H. (2003). 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.

Mao, S., Rosenfeld, A., and Kanungo, T. (2003). Document structure analysis algorithms: A literature survey. *SPIE Proceedings*, 5010.

Martins, R. M., Simaki, V., Kucher, K., Paradis, C., and Kerren, A. (2017). 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.

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In Meiselman, H. L., editor, *Emotion Measurement*, pages 201–237. Woodhead Publishing.

Mohammad, S. M., Kiritchenko, S., Sobhani, P., Zhu, X., and Cherry, C. (2016). SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the International Workshop on Semantic Evaluation*, SemEval '16.

Nopp, C. and Hanbury, A. (2015). 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. Association for Computational Linguistics.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In Smola, A. J., Bartlett, P. L., Schölkopf, B., and Schuurmans, D., editors, *Advances in Large Margin Classifiers*, pages 61–74. MIT Press.

Schreibman, S., Siemens, R., and Unsworth, J. (2016). *A New Companion to Digital Humanities*. John Wiley & Sons.

Simaki, V., Paradis, C., and Kerren, A. (2017a). 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.

Simaki, V., Paradis, C., Skeppstedt, M., Sahlgren, M., Kucher, K., and Kerren, A. (2017b). Annotating speaker stance in discourse: The Brexit Blog Corpus. *Corpus Linguistics and Linguistic Theory*.

Sinclair, S. and Rockwell, G. (2016). Text analysis and visualization. In Schreibman, S., Siemens, R., and Unsworth, J., editors, *A New Companion to Digital Humanities*, pages 274–290. John Wiley & Sons.

Skeppstedt, M., Kucher, K., Paradis, C., and Kerren, A. (2017a). Language processing components of the StaViCTA project. In *Proceedings of the Workshop on Logic and Algorithms in Computational Linguistics*, LACompLing '17, pages 137–138.

Skeppstedt, M., Sahlgren, M., Paradis, C., and Kerren, A. (2016a). 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. Association for Computational Linguistics.

Skeppstedt, M., Sahlgren, M., Paradis, C., and Kerren, A. (2016b). 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. Association for Computational Linguistics.

Skeppstedt, M., Simaki, V., Paradis, C., and Kerren, A. (2017b). 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.

Strobelt, H., Oelke, D., Kwon, B. C., Schreck, T., and Pfister, H. (2016). Guidelines for effective usage of text highlighting techniques. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):489–498.

Strobelt, H., Oelke, D., Rohrdantz, C., Stoffel, A., Keim, D. A., and Deussen, O. (2009). Document Cards: A top trumps visualization for documents. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1145–1152.