# Visual Analysis of Sentiment and Stance in Social Media Texts

K. Kucher[1], C. Paradis[2], and A. Kerren[1]

[1]Department of Computer Science and Media Technology, Linnaeus University, Sweden
[2]Centre for Languages and Literature, Lund University, Sweden

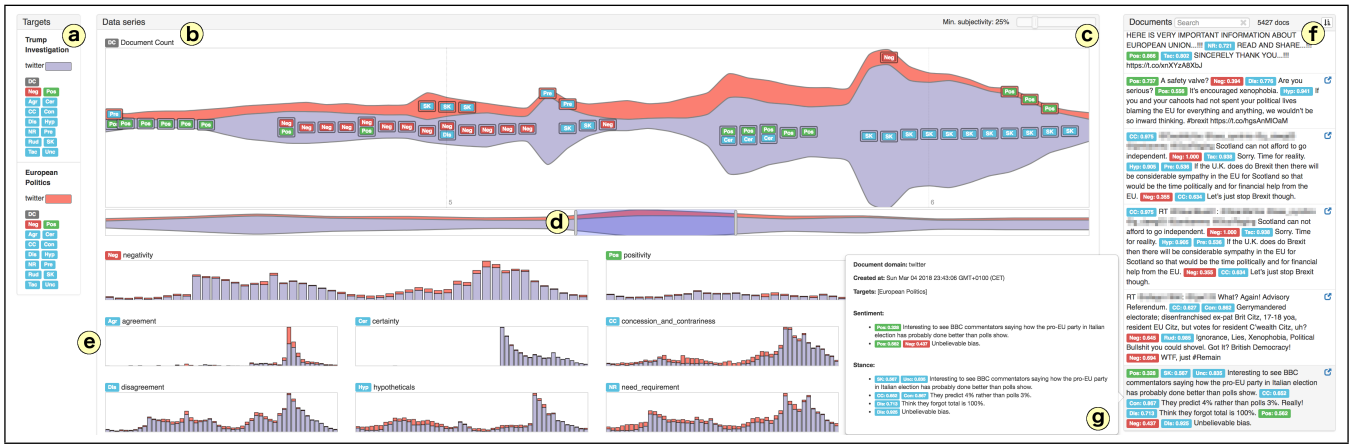


**Figure 1:** *Visualization of sentiment and stance data series discovered in social media data in our VA system StanceVis Prime: (a) a list of tracked targets of interest, corresponding data domains, and detected subjectivity categories; (b) a stacked graph representing the processed document counts and cues about detected subjectivity levels; (c) a slider controlling the threshold for displaying subjectivity cues in (b); (d) a range slider providing an overview about the complete data set; (e) bar charts representing sentiment and stance data series; (f) a document view panel; and (g) a tooltip with detailed sentiment and stance classification results for a specific document.*

**Abstract**

*Despite the growing interest for visualization of sentiments and emotions in textual data, the task of detecting and visualizing various stances is not addressed well by the existing approaches. The challenges associated with this task include development of the underlying computational methods and visualization of the corresponding multi-label stance classification results. In this poster abstract, we describe the ongoing work on a visual analytics platform, called StanceVis Prime, which is designed for analysis of sentiment and stance in temporal text data from various social media data sources. Our approach consumes documents from several text stream sources, applies sentiment and stance classification, and provides end users with both an overview of the resulting data series and a detailed view for close reading and examination of the classifiers' output. The intended use case scenarios for StanceVis Prime include social media monitoring and research in sociolinguistics.*

**CCS Concepts**

•***Human-centered computing*** → *Visual analytics;* •***Computing methodologies*** → *Discourse, dialogue and pragmatics;* •***Information systems*** → *Sentiment analysis;*

## 1. Introduction

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, such as sentiments, opinions, and emotions [PL08]. The analysis of stancetaking in textual data [Eng07, SPS*17] can provide even further insights about the



DOI: 10.2312/eurp.20181127

subjective position of the speaker, for instance, agreement or disagreement with a certain topic [MKS*16, MSK17b], or expression of certainty and prediction [SPK17, SSPK17].

However, manual analysis of texts and examination of raw output of computational text analyses do not scale up to amounts of data produced by social media. Information visualization and visual analytics approaches have been therefore applied successfully to address this challenge [CLY17]. More specifically, multiple existing sentiment visualization techniques [KPK18] address the tasks of visualizing polarity [CLS*12] and emotions [ZGWZ14] detected in temporal text data. Several existing approaches are also relevant to the stance visualization task: for instance, Lingoscope [DZES14] visualizes the language use of acceptors and sceptics of a certain topic in blogs. ConToVi [EAGA*16] supports visualization of debate transcripts using categories beyond sentiment, such as certainty, eloquence, and politeness. uVSAT [KSBK*16] focuses on data series based on markers of sentiment, emotions, and stance categories such as certainty and uncertainty in blogs and forums. Finally, StanceXplore [MSK*17a] supports visualization of multiple stance categories detected in social media data from Twitter.

In this poster abstract, we present our ongoing 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. Compared to the existing works related to stance visualization, our approach aims to support the following: (1) consumption of data from multiple sources (currently, streams of Twitter posts and Reddit comments), (2) classification of both sentiment polarity and multiple non-exclusive stance categories, and (3) visualization of multi-label classification results both at the overview (derived time series for various sentiment and stance categories at multiple levels of granularity) and detail (individual document contents injected with classification results) levels.

## 2. Architecture

The backend of StanceVis Prime is designed around a data collection service in Python which consumes text data streams from Twitter and Reddit. There are multiple targets of interest to be tracked, each defined by a list of key terms (for Reddit, also a list of subforums). The retrieved text documents are saved into MongoDB and put on the queue for classification. Our system uses the VADER sentiment classifier [HG14] and a logistic regression-based stance classifier [SSPK16, SSPK17] developed with scikit-learn [PVG*11] for twelve stance categories. The classification results are used to create the corresponding time series at the granularity of one second and several levels of aggregation (minute, hour, day). Our visualization frontend is served with Flask, and it is implemented in JavaScript with D3 and Rickshaw.

## 3. Visualization methodology

After selecting the list of interesting targets and domains and specifying the overall time interval, the user is presented with a visualization interface demonstrated in Fig. 1. Here, the user has loaded a Twitter data set on two targets of interest: the Trump investigation and European politics. The list marked in Fig. 1(a) acts as a legend for colors associated with each target/domain and enumerates the corresponding data series: the overall document count (incl. documents with no detected occurrences of subjectivity), negative and positive sentiment polarity, and twelve stance categories (incl. agreement, certainty, contrast, prediction, etc.). Category titles are represented with abbreviated labels. These labels are also used in other parts of the interface to avoid the introduction of complicated glyphs for the corresponding 2+12 categories of subjectivity.

A stacked graph [HHN00, BW08] is used in the central part of the interface (see Fig. 1(b)) 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 Fig. 1(d), which might cause the change of data granularity (e.g., from days to hours). Here, the user decided to focus on an area with an increased number of documents on European politics, which happened to be the dates around March 4, 2018, when a general election took place in Italy. The main 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 (e.g., negative sentiment being detected in 75% of documents for a specific political target during one day). By controlling the slider depicted in Fig. 1(c), the user can adjust the minimal threshold for the relative level of subjectivity. The detailed information about temporal distributions of sentiment and stance categories with regard to targets/domains is available from multiple small bar charts positioned under the main stacked graph (see Fig. 1(e)).

Finally, the visual interface includes a document list view depicted in Fig. 1(f), which supports text search and multiple sorting options (e.g., sorting by timestamp or by overall number of detected stance occurrences). 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 Fig. 1(g) which includes additional details about the document timestamp, data domain, associated target(s) of interest, and detailed classification results.

## 4. Conclusions and future work

In this poster abstract, we have briefly introduced the ongoing work on StanceVis Prime, a visual analytics system for social media text streams supporting sentiment and stance analysis. The future work includes the design of additional visual representations to help the user navigate to interesting reasonably-sized data subsets and corresponding topics, documents, and stance & sentiment classification results. Our refinements of the main representation might also involve the approach recently described by Cuenca et al. [CSWP18]. Finally, the next steps of the development will include close collaboration with experts in sociolinguistics and evaluation of the system with a user study.

**Acknowledgments:** This work was partially supported by the framework grant "The Digitized Society—Past, Present, and Future" from the Swedish Research Council (Vetenskapsrådet) [grant number 2012–5659].

## References


[BW08] BYRON L., WATTENBERG M.: Stacked graphs — geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (Nov. 2008), 1245–1252. doi:10.1109/TVCG.2008.166. 2

[CLS*12] CAO N., LIN Y.-R., SUN X., LAZER D., LIU S., QU H.: Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (Dec. 2012), 2649–2658. doi:10.1109/TVCG.2012.291. 2

[CLY17] CHEN S., LIN L., YUAN X.: Social media visual analytics. *Computer Graphics Forum 36*, 3 (June 2017), 563–587. doi:10.1111/cgf.13211. 2

[CSWP18] CUENCA E., SALLABERRY A., WANG F. Y., PONCELET P.: MultiStream: A multiresolution streamgraph approach to explore hierarchical time series. *IEEE Transactions on Visualization and Computer Graphics* (2018). doi:10.1109/TVCG.2018.2796591. 2

[DZES14] DIAKOPOULOS N., ZHANG A. X., ELGESEM D., SALWAY A.: Identifying and analyzing moral evaluation frames in climate change blog discourse. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media* (2014), ICWSM '14, pp. 583–586. URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8064. 2

[EAGA*16] EL-ASSADY M., GOLD V., ACEVEDO C., COLLINS C., KEIM D.: ConToVi: Multi-party conversation exploration using topic-space views. *Computer Graphics Forum 35*, 3 (June 2016), 431–440. doi:10.1111/cgf.12919. 2

[Eng07] ENGLEBRETSON R. (Ed.): *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, vol. 164 of *Pragmatics & Beyond New Series*. John Benjamins, 2007. doi:10.1075/pbns.164. 1

[HG14] HUTTO C., GILBERT E.: 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* (2014), ICWSM '14. URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109. 2

[HHN00] HAVRE S., HETZLER B., NOWELL L.: ThemeRiver: Visualizing theme changes over time. In *Proceedings of the IEEE Symposium on Information Visualization* (2000), InfoVis '00, IEEE, pp. 115–123. doi:10.1109/INFVIS.2000.885098. 2

[KPK18] KUCHER K., PARADIS C., KERREN A.: The state of the art in sentiment visualization. *Computer Graphics Forum 37*, 1 (Feb. 2018), 71–96. doi:10.1111/cgf.13217. 2

[KSBK*16] KUCHER K., SCHAMP-BJERede T., KERREN A., PARADIS C., SAHLGREN M.: Visual analysis of online social media to open up the investigation of stance phenomena. *Information Visualization 15*, 2 (Apr. 2016), 93–116. doi:10.1177/1473871615575079. 2

[MKS*16] MOHAMMAD S. M., KIRITCHENKO S., SOBHANI P., ZHU X., CHERRY C.: SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the International Workshop on Semantic Evaluation* (2016), SemEval '16. URL: http://aclweb.org/anthology/S16-1003. 2

[MSK*17a] MARTINS R. M., SIMAKI V., KUCHER K., PARADIS C., KERREN A.: StanceXplore: Visualization for the interactive exploration of stance in social media. In *Proceedings of the 2nd Workshop on Visualization for the Digital Humanities* (2017), VIS4DH '17. URL: http://vis4dh.dbvis.de/papers/2017/. 2

[MSK17b] MOHAMMAD S. M., SOBHANI P., KIRITCHENKO S.: Stance and sentiment in tweets. *ACM Transactions on Internet Technology 17*, 3 (June 2017), 26:1–26:23. doi:10.1145/3003433. 2

[PL08] PANG B., LEE L.: Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval 2*, 1–2 (2008), 1–135. doi:10.1561/1500000011. 1

[PVG*11] PEDREGOSA F., VAROQUAUX G., GRAMFORT A., MICHEL V., THIRION B., GRISEL O., BLONDEL M., PRETTENHOFER P., WEISS R., DUBOURG V., VANDERPLAS J., PASSOS A., COURNAPEAU D., BRUCHER M., PERROT M., DUCHESNAY Ë.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (Oct. 2011), 2825–2830. URL: http://jmlr.org/papers/v12/pedregosa11a.html. 2

[SPK17] SIMAKI V., PARADIS C., KERREN A.: Stance classification in texts from blogs on the 2016 British referendum. In *Proceedings of the International Conference on Speech and Computer* (2017), SPECOM '17, Springer, pp. 700–709. doi:10.1007/978-3-319-66429-3_70. 2

[SPS*17] SIMAKI V., PARADIS C., SKEPPSTEDT M., SAHLGREN M., KUCHER K., KERREN A.: Annotating speaker stance in discourse: The Brexit Blog Corpus. *Corpus Linguistics and Linguistic Theory* (2017). doi:https://doi.org/10.1515/cllt-2016-0060. 1

[SSPK16] SKEPPSTEDT M., SAHLGREN M., PARADIS C., KERREN A.: 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* (2016), PEOPLES '16, Association for Computational Linguistics, pp. 50–59. URL: http://aclweb.org/anthology/W16-4306. 2

[SSPK17] SKEPPSTEDT M., SIMAKI V., PARADIS C., KERREN A.: Detection of stance and sentiment modifiers in political blogs. In *Proceedings of the International Conference on Speech and Computer* (2017), SPECOM '17, Springer, pp. 302–311. doi:10.1007/978-3-319-66429-3_29. 2

[ZGWZ14] ZHAO J., GOU L., WANG F., ZHOU M.: 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* (2014), VAST '14, IEEE, pp. 203–212. doi:10.1109/VAST.2014.7042496. 2