VISTX: Experimenting with Interactive Stance and Sentiment Visualisation

Katherine Grayson & Jon Chamberlain
School of Computer Science and Electronic Engineering
University of Essex, Colchester, UK
kategrays676@gmail.com, jchamb@essex.ac.uk

***Abstract*— User-generated content online continues to grow rapidly, and with this it becomes more challenging to evaluate the general consensus on a trending topic. Even greater challenges are faced if one wishes to compare opinions across multiple topics, as it can be difficult to make comparisons and reach conclusions with large amounts of data in an intuitive manner. VISTX explores innovative methods of visualising opinion in terms of stance and sentiment using Twitter data from the SemEval-2016 Stance Dataset and presents three types of interactive visualisations. We discuss the method of testing the visualisations for effectiveness with a small group of users, where time taken and the number of incorrect interpretations were measured for each user. We discuss the results, which indicate that the 'circle-packing' visualisation type may lead to quicker interpretations for five or more topics. Finally, we discuss the work needed to further solidify these initial findings.**

***Keywords*— sentiment, stance, information visualisation, tweets, opinion, polarity, text classification.**

I. Introduction

The internet provides a platform for billions of people to share their thoughts with others from nearly every corner of the world. A large part of the reason this is able to happen has been due to the advent of social media, allowing us to condense our experiences into tweets, snaps, posts, and so on. There are countless approaches already present for visualising all types of online discussion boards. Social media visualisation has become a domain of great interest in particular; in 2017 it was found that 90% of data in the world had been generated in the previous 3 years [1]. With so many people now voicing these thoughts on the web, society has found that analysing these opinions can lead to making better decisions commercially, politically and financially, and so sentiment and stance has become a region of interest in recent years.

This paper discusses the current approaches to sentiment and stance visualisation, both separately and together, and briefly mentions their applications within society (section 2). We attempt to widen the area of research by contributing two interactive visualisations that have not yet been used in the field of sentiment and stance visualisation and discuss how this was achieved (section 3). We test properties of the visualisations with a small group of users (section 4) then discuss the results and therefore the effectiveness of the visualisations compared to current practices (section 5). In conclusion, we suggest further work that would need to be completed in order to gain greater insight on the topic (section 6).

II. Related Work

Sentiment analysis, also known as opinion mining or affect analysis, is a topic that has been extensively studied in the world of natural language processing and usually classifies pieces of text as 'positive' or 'negative' based on clues within the text. Other approaches classify sentiment into emotional categories like 'stressed' or 'relaxed' - an example of this in social media sentiment analysis is Healey and Ramaswamy's tweet visualizer [2] which uses several visualisation techniques to highlight different aspects of tweets [3]. Sentiment Sweep [4] analyses real-time Twitter data for sentiment and uses numerous visualisations to illustrate the results. Similarly, the work of Hao et al. visualises sentiment of twitter data streams but by using pixels on a geographical map and on a calendar [5]. The '#ImagineEurope' Twitter Demo [6] visualises the UK debate on the EU referendum in regards to sentiment, location and hashtag frequency by using 3 datasets and surprisingly suggests that official campaigns were not influencing the debate [7]. This leads us to believe that using visualisations to compare can lead to unprecedented insights. Sentiment analysis has been shown to have many applications in tandem with social media like tracking brand reputation [8], predicting election results [9], aiding in disaster management [10] and predicting the stock market [11].

The problem of visualising stance, on the other hand, is one that has not been sufficiently explored. One of few systems that addresses this is StanceXplore which visualises stance in regards to ten categories [12] in multiple views, including a timeline view and map view. uVSAT, created for use in the StaViCTA project [13], visualises what Kucher et al. refer to as "'multidimensional' sentiment analysis" [14], and uses bubble charts to look at distributions with regard to a particular type of stance. StanceVis Prime [15], which was also created for the StaViCTA project, allows for simple comparison of sentiment and stance for temporal Twitter and Reddit data using stream graphs and can convey peaks and patterns over time [16]. The interactive visualisation of the SemEval-2016 Stance dataset [17] uses multiple components to allow easy exploration of different aspects of the dataset. The visualisation features a treemap component that shows stance by target in a hierarchal manner. The dataset, which forms the basis of VISTX, is based on detecting whether an author of a tweet is in favour of or against a given target (stance), and the polarity of the language used in the tweet (sentiment) [18]. Stance analysis is related to sentiment analysis and so has similar applications, such as political debates [19] or online debates [20]. However, these domains have distinct differences as explained by Sobhani et al. [21]

The visualisation of both sentiment and stance alongside each other is an even lesser explored topic than stance alone; at the time of writing, the most relevant pieces of research are those of Kucher [13] and Mohammad et al. [18], along with Chamberlain et al. who presented a scalable visualisation of sentiment and stance [22] in planning application reviews, but with flexibility to be applied to other datasets such as the SemEval-2016 Stance Dataset. This method allows for accurate comparison regardless of any size difference between targets.

III. Architecture and Methodology

A. System Overview

Figure 1 gives an overview of the system architecture. The general idea of the system is that the dataset is read in and converted into a dictionary of objects, which is dumped to a JSON file and used by whichever visualisation is requested.

- SE-Test.csv: The SemEval-2016 task's 'test' dataset. See later in this section for more explanation.
- sentstan.py: Processes the dataset of tweets and converts them into a format for processing with Topic.py.
- Topic.py: Converts the dataset into a dictionary of 'Topic' objects and sends the dictionary back to sentstan.py.
- tweetcount.json: Stores the dictionary of 'Topic' objects dumped by sentstan.py.
- server.py: Facilitates the role of a web server by handling GET and POST requests. The desired HTML page requests the data, server.py runs sentstan.py with the desired parameters and the HTML page reloads with the updated data.

Originally, sentstan.py was run via a batch script and the visualisation was hosted using a local web server with Python. However, server.py has replaced this functionality.

The visualisations use the SemEval-2016 task's 'test' dataset (SE-Test.csv in our system) [17] containing 1956 tweets about six targets with stance gold labels and additional labels of 'sentiment' and 'opinion towards'. The targets provided in the dataset are:

- Atheism
- Climate Change is a Real Concern
- Feminist Movement
- Hillary Clinton
- Legalization of Abortion
- Donald Trump (test set only)

This dataset was used as there are currently very few datasets labelled for sentiment and stance, but the system could be easily adapted to use other datasets with a similar format. The targets are examples of heavily debated topics in 2016, so people generally feel very strongly about these topics and opinion can be clearly observed. Both the test and train datasets, along with further information, can be found on the SemEval-2016 Task 6 website [23].

The visualisations are hosted using PythonAnywhere. PythonAnywhere supports multiple technologies including Flask, a micro web framework for Python which we used to build the web application.

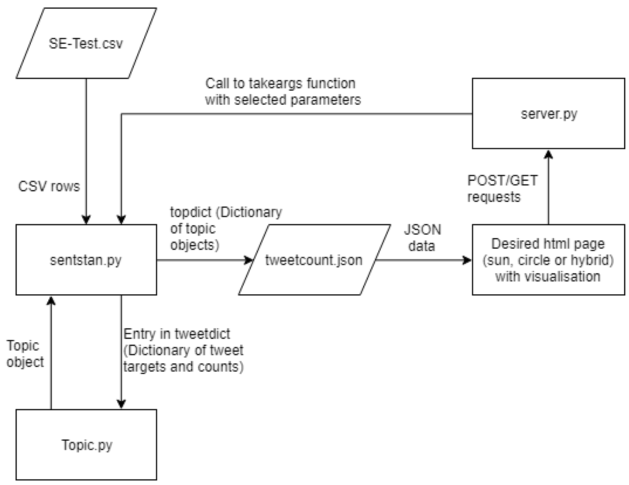

Fig. 1. Overview of the VISTX system, implemented as a web application.

B. Explanation of the Sunburst, Circle-packing and Hybrid Sunburst Visualisations

In VISTX, we adapted two existing visualisations - the well-known sunburst and circle-packing charts - to the field of sentiment and stance and propose a third modified visualisation: the 'hybrid sunburst'. Our **circle packing visualisation** (figure 2) is adapted from Mike Bostock's 'Zoomable Circle Packing' [24], and our sunburst visualisation (figure 3) is adapted from Kerry Rodden's 'Sequences Sunburst' [25]. The circle packing visualisation is explained as follows:

- The large grey outer node houses all the topics/targets.
- Inside this are yellow circles which represent each topic. The name of the topic can be seen when the focus is at the root node. The size of the topic is representative of how many tweets are in the topic.
- On the next level are the circles that represent stance, which are shown in light red, light green and white. Again, the size of the stance node represents how many tweets are in the topic.
- On the lowest level we can see the circles that represent sentiment. These are shown in dark red, dark green and white. Once more, the size of each

sentiment node represents how many tweets are in the topic.

- The sentiment and stance nodes are not shown if there are no tweets in that category. For example, the only topic with neutral/'none' stance was 'Donald Trump', which is why the white stance node is visible here.

The topic, stance and sentiment nodes are interactive. One can hover a node to display its' name and size, and click the node to switch the focus to the desired node, zooming in. Topic, stance and sentiment labels are shown in figure 2.

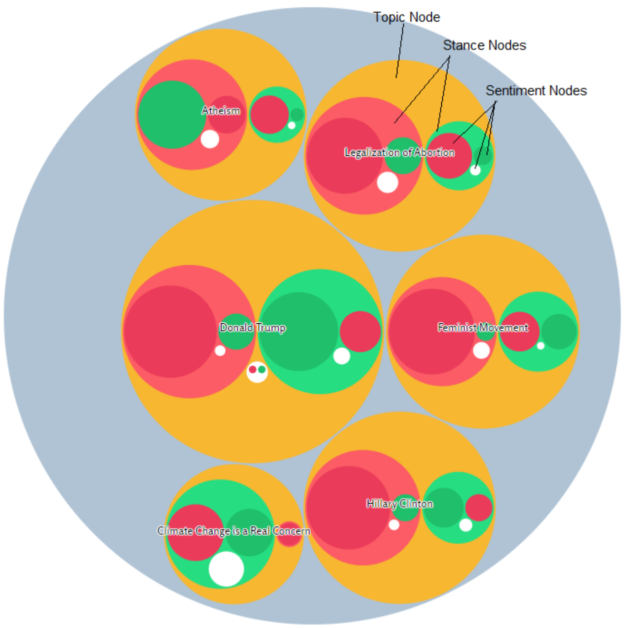

Fig. 2. An example of circle packing visualisation.

We present the **sunburst visualisation** as a different approach to the problem. It was designed after some very small-scale user testing, which implied that the circle-packing visualisation was unintuitive for some to understand. This chart is also based on a hierarchal system where the innermost arc represents target/topic, the middle arc represents stance, and the outermost arc represents sentiment. The components labelled in figure 4 (see next page) are used to aid in understanding the data:

- The legend: A key for the sunburst to show what each colour represents.
- The breadcrumb: Explicitly displays the current sequence shown to help a user understand the levels of the sunburst.
- The example tweet: Shown when hovering any of the sentiment nodes, it will display a randomly chosen tweet (from up to ten tweets) in that category. This would help a user understand what constitutes different stances and sentiments, as per the feedback from the first testing.
- The sunburst: The centrepiece of the visualisation. Displays the data on levels, where the inner (yellow) level represents the topic, the middle layer represents stance, and the outer layer represents sentiment.
- The tweet counter: Shows how many tweets are in the current section. This is to aid the distortion that can happen when using radial layouts.

To use the visualisation, one should hover over different sections using the mouse. Hovering over the topic, stance or sentiment nodes will change the breadcrumb and tweet counter and highlight the selected segment whilst fading segments that are not ancestors of the selected segment. Hovering sentiment nodes will additionally display an example tweet for the section. In this chart, the neutral stance and sentiment nodes are white and appear as gaps since we are more interested in non-neutral stance and sentiment, and because the visualisation is slightly easier to understand with gaps - overcrowded data can be incomprehensible.

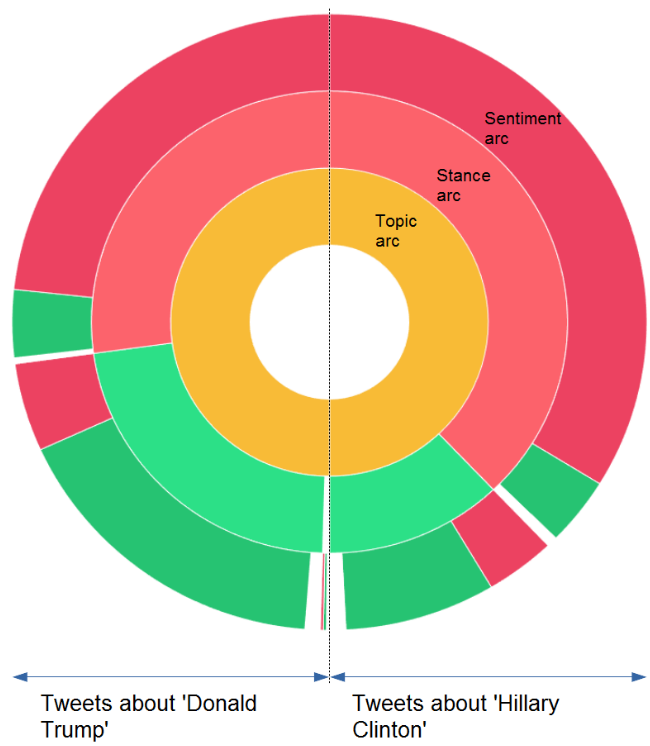

Fig. 3. The hybrid sunburst visualisation with the 'hilvstrump' dataset.

The **hybrid visualisation** for the 'hilvstrump' dataset is shown in figure 3. It is interpreted in the same way as the sunburst, and contains the same components: the breadcrumb, the legend, the example tweet, the chart and the tweet counter. In the example, 'Donald Trump' is the topic represented by the left half of the sunburst, and 'Hillary Clinton' is represented by the right half. The idea behind the hybrid sunburst is to scale two topics proportionately so they can be compared. Although the two datasets might have different

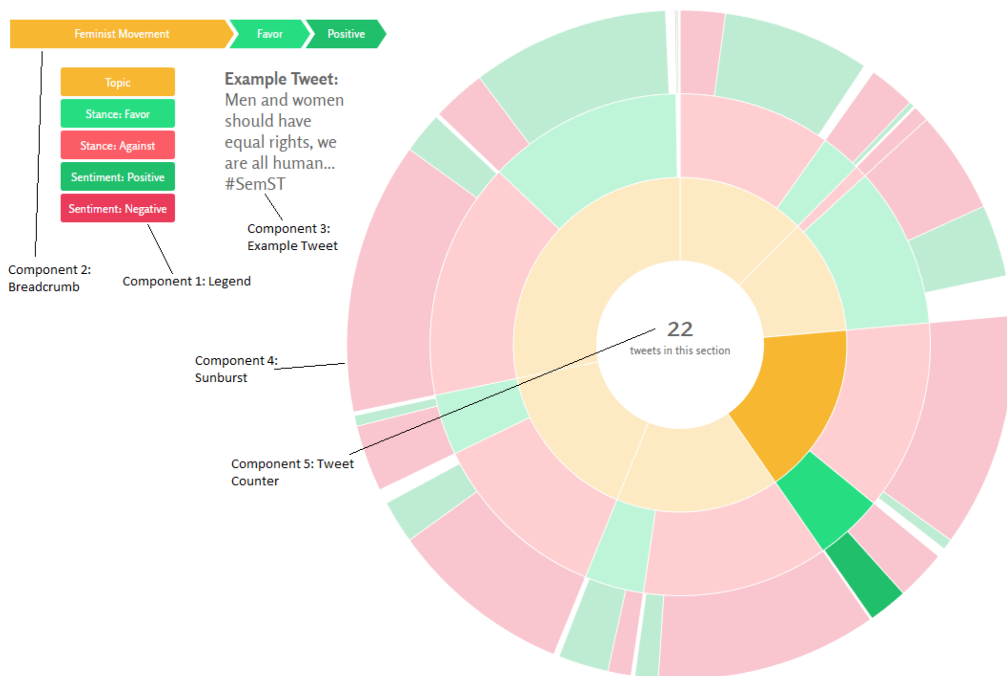

Fig. 4. The sunburst visualisation viewed with the 'full' dataset, labelled with components and hovered over a segment.

amounts of tweets, such as how ‘Donald Trump’ has more tweets than ‘Hillary Clinton’, the hybrid means that the two topics can be compared equally for stance and sentiment. The true number of tweets in each topic is still shown when hovering over a node to help users understand the true numbers. Through the hybrid, certain interpretations become more obvious; in the ‘hilvstrump’ dataset, we clearly see that ‘Hillary Clinton’ had a higher proportion of ‘against’ stance and ‘negative’ sentiment tweets compared to ‘Donald Trump’. This conclusion is hard to reach through using the original sunburst or the circle-packing visualisations due to the topic sizes being vastly different, so it looks as though ‘Donald Trump’ had a higher proportion of ‘against’ stance and ‘negative’ sentiment’.

The circle-packing and sunburst charts are very different approaches to typical visualisations of sentiment and stance as we chose to represent the data hierarchically; a tweet is categorised firstly by target, then by stance and lastly sentiment. This approach can show a completely different angle of the data when compared to the interactive visualisation of the SemEval-2016, since stance and sentiment are represented separately here. We can therefore see the sentiment within a particular stance, and areas where stance and sentiment are oppositely aligned e.g. in the 'Atheism' target where stance is 'against' and sentiment is 'positive', showing people against atheism but using positive language. The hybrid sunburst initially appears to be effective for comparing two related topics for sentiment and stance. Although the concept of the two topics not being the same size but proportionally the same might be visually confusing at first, this visualisation has the potential to help a user interpret data in a completely new way.

All the visualisations are written in JavaScript using the D3.js library, and viewable online [26].

IV. User Testing

The testing was completed in two separate sessions. One small round of iterative testing was completed about the circle-packing visualisation, and more comprehensive testing was completed later in the project about all three visualisations. The second testing was performed on a group of twelve users. User testing was performed via a web application made in PHP. Each subject was shown a short YouTube video [27] to explain what sentiment and stance is, how they are shown in tweets, and how to interpret the visualisations that are seen in the testing questions. A test user was allocated a subject number according to what order they completed the questionnaire in, which is selected from a dropdown menu on the index page. The testing was performed to evaluate two hypotheses:

1. The sunburst hybrid visualisation will lead to quicker interpretation and fewer incorrect interpretations in two topics when compared to a histogram visualisation.
2. The circle packing visualisation will lead to quicker interpretation and fewer incorrect interpretations in five topics when compared to a histogram visualisation.

Each hypothesis was tested by using five datasets, making up ten datasets total. A user was not shown a dataset more than once. Datasets one through five tested the first hypothesis, so the test user was randomly shown the sunburst or histogram for the current dataset, whereas datasets six through ten tested the second hypothesis and showed the circle-packing or a histogram for the dataset. A Python program was created to decide which visualisation was seen for each dataset (in order) for each test user. These values were placed into two CSV files which were read by the PHP when the questionnaire was taking place. Semi-random artificial datasets were created by use of another program, which take inputs indicating how many tweets are desired in total for a dataset and how many tweets are desired in each topic in the dataset. Based on these values, data were generated for each topic's stances and sentiments. If two topics were being generated, then the second topic was given 'size' and 'tweets' values; both values were needed for the hybrid sunburst visualisation. The testing site works as follows:

1. After the user has watched the video and selected their subject number, the test begins. The user was timed from the start of a question until they choose the correct answer and submit it – the timer runs in the background of the page.
2. Based on the test user's subject number and the visualisation that has been randomly chosen for that user, the question and visualisation were displayed.
3. There were four possible answers for a question, usually in a true/false statement format. If the user selected the wrong answer to the question, they were alerted of it after trying to submit the answer. The user cannot advance to the next question until they have chosen the correct answer, and the number of incorrect attempts is logged.
4. When the correct answer was submitted, the following was logged to answers.csv: the subject number, the visualisation seen, the correct answer, the number of incorrect attempts, and the time taken in seconds.
5. After the ten datasets have been seen, the user was taken to the end page and closed the browser window.

V. RESULTS AND DISCUSSION

The results of testing are displayed in tables 1 and 2 and were tested for statistical significance using paired t-tests. An alpha value of $p<0.01$ was used.

TABLE I. A COMPARISION OF THE SUNBURST HYBRID VISUALISATION WITH A HISTOGRAM ON TWO TOPICS (HYPOTHESIS 1).

Hyp. 1 Results	**Dependent variables: incorrect answers and time (seconds)**			
	Hist. Incorrect Ans.	***Sun. Incorrect Ans.***	***Hist. time***	***Sun. time***
Mean	0.43	0.60	64.28	62.87
Standard Deviation	0.73	0.77	34.41	38.03
Standard Error	0.13	0.14	6.39	6.94

TABLE II. A COMPARISON OF THE CIRCLE PACKING VISUALISATION WITH A HISTOGRAM ON FIVE TOPICS (HYPOTHESIS 2).

Hyp. 2 Results	**Dependent variables: incorrect answers and time (seconds)**			
	Hist. Incorrect Ans.	***Circ. Incorrect Ans.***	***Hist. time***	***Circ. time***
Mean	1.00	0.90	73.10	56.17
Standard Deviation	1.44	0.99	41.08	27.17
Standard Error	0.26	0.18	7.50	4.96

The results for hypothesis 1 (comparing the sunburst hybrid visualisation with a histogram on two topics) indicate that there was no statistically significant difference between the number of incorrect interpretations per dataset (0.60 vs 0.43, $n=30$ $p=0.3925$, paired t-test) and the time taken to make an interpretation (62.87s vs 64.28s, $n=12$, $p=0.8820$, paired t-test). The standard deviation of the results also indicates that the distribution of response (or the ability of the user to make an interpretation) was highly varied and may suggest a non-parametric distribution.

The results for hypothesis 2 (comparing the circle packing visualisation with a histogram on five topics) also indicate that there was no statistically significant difference between the number of incorrect interpretations per dataset (0.90 vs 1.00, $n=30$, $p=0.7553$, paired t-test) and the time taken to make an interpretation (73.10s vs 56.17s, $n=12$, $p=0.0647$, paired t-test). The standard deviation of the results also suggests a varied, non-parametric distribution. Although the results of the time taken was not significant it is a marked decrease in time taken which, with further testing, may be shown to have a significant effect.

When comparing between the two datasets we can see that the test users produced more errors in interpretation (with both the novel and histogram visualisations), supporting the intuitive belief that understanding concepts with a greater number of variables are more challenging. This is also reflected in the time to process the histograms; however, we notice a marked decrease in processing time for the circle packing visualisation compared to the hybrid sunburst.

Due to the small numbers of test users we cannot make a definitive conclusion about whether the circle-packing or sunburst visualisations were more effective than the histograms. The p-values calculated show there is not strong enough evidence to reject the null hypotheses: that there is no significant difference in interpretation times and incorrect interpretations between the sunburst hybrid and histogram for two topics, and that there is no significant difference in interpretation times and incorrect interpretations between circle-packing and histogram for five topics. To achieve more accurate results more test users are needed. A larger sample size would provide a more accurate evaluation and allow us to identify outliers. The data did reveal one outlier, but with a larger sample size we might have found more which could have made the results more insightful.

The possibility of a learning effect present in the results must also be considered, meaning that as users see a particular representation more times, the interpretation time taken decreases and the number of incorrect interpretations decrease. To gauge whether this effect is present in the results collected, the first and last responses for each user could be plotted on a scatter graph for time taken and incorrect interpretations and tested in an unpaired t-test for significance.

VI. Conclusion and Further Work

These results prove as a starting point for further research into effective ways of visualising sentiment and stance. In this research we developed three novel visualizations for Tweet datasets and demonstrated how they could be tested in order to support human evaluation and interpretation. Putting our results into perspective is difficult to do at present, since there are no known user evaluations of sentiment and stance visualisations. However, that continued research into sentiment and stance from other authors will produce such results for comparison.

The approach taken in the project could be expanded into a geographical representation to show tweets based on location. This approach was considered but the dataset used did not include locational data or the means to retrieve it using tweet IDs. Using a different dataset or gathering tweets directly through the Twitter API could achieve this.

This research focused on the visualisation of text analytics rather than the analytics themselves; however, stance and sentiment analysis is a very active research field in itself that would benefit from synthesized visualization such as those proposed by the VISTX system.

Acknowledgment

We acknowledge support from the School of Computer Science and Electronic Engineering at The University of Essex. We thank Dr. Orland Hoeber for his suggestions, and everyone who participated in the online survey.

References

[1] L. Kharb, 'Exploration of Social Networks with Visualization Tools', *American Journal of Engineering Research (AJER)*, vol. 6, no. 3, pp. 90-93, 2017.

[2] C. Healey and S. Ramaswamy, 'Tweet Sentiment Visualisation App', 2018. [Online]. Available: https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/. [Accessed: 8- Apr- 2019].

[3] C. Healey and S. Ramaswamy, 'Tweet Sentiment Visualisation', 2018. [Online]. Available: https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/. [Accessed: 8- Apr- 2019].

[4] A. Sykes, 'Twitter Sentiment Visualisation', 2012. [Online]. Available at: https://github.com/Lissy93/twitter-sentiment-visualisation. [Accessed: 7- Apr- 2019].

[5] M. Hao et al., 'Visual Sentiment Analysis on Twitter Data Streams', in IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, RI, USA, 2011.

[6] Neuropolitics Research Lab, 'Neuropolitics Research #ImagineEurope Twitter Demo', 2018. [Online]. Available at: http://www.pol.ed.ac.uk/neuropoliticsresearch/sections/remote_content. [Accessed: 7- Apr- 2019].

[7] C. Llewellyn and L. Cram, 'Brexit? Analyzing Opinion on the UK-EU Referendum within Twitter', in International Conference on Weblogs and Social Media, Cologne, Germany, 2016.

[8] N. A. Vidya et al., 'Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers', in Third Information Systems International Conference (ISICO), Surabaya, Indonesia, 2015.

[9] L. Oikonomoi and C. Tjortjis, 'A Method for Predicting the Winner of the USA Presidential Elections Using Data Extracted from Twitter', in South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media Conference (SEEDA_CECNSM), Kastoria, Greece, 2018.

[10] D. Buscaldi and Irazú Hernandez-Farias, 'Sentiment Analysis on Microblogs for Natural Disasters Management: a Study on the 2014 Genoa Floodings', in Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 2015.

[11] StockFluence, 'StockFluence | Stock Sentiment Analysis', 2017. [Online]. Available: https://www.stockfluence.com/. [Accessed: 8- Apr-2019].

[12] R. M. Martins et al., 'StanceXplore: Visualization for the Interactive Exploration of Stance in Social Media', in 2nd Workshop on Visualization for the Digital Humanities, Phoenix, Arizona, 2012.

[13] K. Kucher, 'Sentiment and Stance Visualization of Textual Data for Social Media', Doctoral Thesis, Linnaeus University, 2019.

[14] K. Kucher et al., 'Visual analysis of online social media to open up the investigation of stance phenomena', Information Visualization, vol. 15, no. 2, pp. 93 - 116, 2015.

[15] K. Kucher et al., 'Visual Analysis of Sentiment and Stance in Social Media Texts', in Eurographics/IEEE VGTC Conference on Visualization (EuroVis), Brno, Czech Republic, 2018.

[16] E. Cuenca et al., 'MultiStream: A Multiresolution Streamgraph Approach to Explore Hierarchical Time Series', IEEE Transactions on Visualization and Computer Graphics, vol. 24, no. 12, pp. 2160 - 3173, 2018.

[17] S. M. Mohammad et al., 'SemEval-2016 Stance Dataset', 2016. [Online]. Available at: http://www.saifmohammad.com/WebPages/StanceDataset.htm. [Accessed: 7- Apr- 2019].

[18] S. M. Mohammad et al., 'Stance and Sentiment in Tweets', ACM Transactions on Internet Technology (TOIT), vol. 17, issue 3, no. 26, 2017.

[19] M. Thomas et al., 'Get out the vote: Determining Support or Opposition from Congressional Floor-Debate Transcripts', in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), Sydney, Australia, 2006.

[20] S. Somasundaran and J. Wiebe, 'Recognizing Stances in Online Debates', in Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Suntec, Singapore, 2009.

[21] P. Sobhani et al., 'Detecting Stance in Tweets and Analyzing its Interaction with Sentiment', in Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics, Berlin, Germany, 2016.

[22] J. Chamberlain et al., 'Scalable Visualisation of Sentiment and Stance', presented at Proceedings of the Eleventh International Conference on Language Resources and Evaluation, Miyazaki, Japan, 2018.

[23] S. M. Mohammed et al., 'Data, Evaluation, and Results < SemEval-2016 Task 6', 2016. [Online]. Available: http://alt.qcri.org/semeval2016/task6/index.php?id=data-and-tools. [Accessed: 9- Apr- 2019].

[24] M. Bostock, 'Zoomable Circle Packing / D3 / Observable', 2018. [Online]. Available: https://observablehq.com/@d3/zoomable-circle-packing. [Accessed: 10- Apr- 2019].

[25] K. Rodden, 'Sequences sunburst (d3 v4) – bl.ocks.org', 2019. [Online]. Available: https://bl.ocks.org/kerryrodden/766f8f6d31f645c39f488a0befa1e3c8. [Accessed: 14- Apr- 2019].

[26] K. Grayson, 'kategray.pythonanywhere.com', 2019. [Online]. Available: http://kategray.pythonanywhere.com/. [Accessed: 9- Apr- 2019].

[27] K. Grayson, 'Understanding Visualisations with Sentiment and Stance', 2019. [Online]. Available: https://www.youtube.com/watch?v=us-V9hE6vWI. [Accessed: 20- Jun- 2019].