Visual Sentiment Analysis of RSS News Feeds Featuring the US Presidential Election in 2008

Franz Wanner, Christian Rohrdantz, Florian Mansmann, Daniela Oelke, Daniel A. Keim
University of Konstanz, Germany
firstname.lastname@uni-konstanz.de

ABSTRACT
The technology behind RSS feeds offers great possibilities to retrieve more news items than ever. In contrast to these technical developments, human capabilities to read all these news items have not increased likewise. To bridge this gap, this paper presents a visual analytics tool for conducting semi-automatic sentiment analysis of large news feeds. While the tool automatically retrieves and analyzes RSS feeds with respect to positive and negative opinion words, the more demanding news analysis of finding trends, spotting peculiarities and putting events into context is left to the human expert. For a solid analysis the news similarity filter enables highlighting of similar or redundant news items. A case study about news related to the US presidential election in 2008 shows how the visual interface of the tool empowers the analyst to draw meaningful conclusions without the effort of reading all news postings.

Author Keywords
sentiment analysis, opinion mining, information visualization, visual analytics

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: Miscellaneous

INTRODUCTION
The web is the largest information source in the world. One major aspect of the web is to bring news from all over the world via RSS feeds instantaneously on your screen. Apart from passive usage of the web as a media, web 2.0 technology helps more and more people to actively contribute to this valuable information source by creating content in an easy way. There are many possibilities to take an active part in the web: blogs, reviews and other ways to state comments.

Analyzing news stories and user generated content is of huge importance for many people and organizations. Economic analysts, for example, would like to find consumer and public opinions on their products and services. Likewise, potential consumers seek experiences of existing users before making a purchase decision or afterwards to cope with the product's shortcomings or praise its functionality. Furthermore, politicians want to find out their public reputation, the manner the news write about them, and the reaction of the public on these articles.

Since public opinion polls are an expensive undertaking, our goal is to offer a semi-automatic approach by mining the web for particular key words, conducting sentiment analysis on the text to assess how positive or negative a particular news postings is, and then to present the information in a visual exploration tool. While our approach is not suitable to completely replace a thoroughly conducted opinion poll due to the lack of accuracy, it has also some unique advantages, namely low costs and the possibility to continuously monitor a particular subject in real-time. Knowing at an early stage that consumers have a problem with a sub-component of a product gives the company more time to react appropriately and to avoid damage to valuable trade marks.

In this paper, we demonstrate a novel way of using text analysis methods in combination with a visual representation. On the one hand, this system automatically evaluates the emotional content of a news posting. On the other hand, the visual interface empowers the human expert to draw meaningful conclusions, to selectively read a few news postings with strong emotional content, to discover trends, and to gain an overview of the development of chosen topic in the media.

To exemplify our tool we have a closer look at the news coverage in the web of the 2008 US presidential election. Out of 50 chosen political RSS newstickers, we retrieved all RSS articles containing at least one of the following key words: "Obama", "McCain", "Biden" and "Palin" as well as "Democrat" and "Republican". Thereupon, the articles are automatically evaluated with respect to the contained positive and negative opinion words, resulting in a normalized sentiment score for each article.

For presentation purposes, these articles are then visualized on a daily timeline using symbols to encode the contained key words. The vertical position of each symbol is defined by the article's sentiment score, which makes strong emotional news more visible. Furthermore, we demonstrate an interactive feature to show relations between the news items to track the development of a specific topic.

The rest of this paper is structured as follows: In section

Related Work text and sentiment analysis methods and visual interfaces for them are discussed. The next section *Visual Sentiment Analysis* then presents our processing, visualization, and interaction approaches for analyzing the news coverage of the 2008 US presidential election. Afterwards, section *Results* shows how some interesting topics about the candidates and their parties manifest in our visualization. By summarizing our contributions we draw our conclusions in the last section.

RELATED WORK

Text Analysis

The visualization and visual analysis of textual data is increasingly attracting interest in different application domains. Many of the early approaches in that area dealt with the visualization of retrieval results (see e.g., VIBE [22] or InfoCrystal [27]). Furthermore, a variety of techniques concentrate on the visualization of large document collections, most of which are based on dimensionality-reduction methods (see e.g. WebSOM [23], Galaxies and ThemeScape of IN-SPIRE^{TM} [30], or [9]). In contrast to this, text feature visualization techniques visualize single documents in detail and show the distribution of specific text features across the text. Prominent examples among these are e.g. TileBars [16], Seesoft [3], the FeatureLens [6], and Literature Fingerprinting [19]. But also [1] and the Compus system of Fekete and Dufournaud [7] are worth being mentioned: As opposed to the other techniques they offer the possibility to visualize several text features at once.

Relatively few approaches tackle the problem of visualizing temporal variations across a set of documents as we do in this paper. One example for such an approach is the well-known ThemeRiver visualization [15] that reveals the development of topics over time in a river-like graphic. According to the metaphor each topic is represented as one colored "current" in the "river" that flows in the direction of the timeline from left to right. To allow for several different themes to be displayed at once the currents are stacked on top of each other. The thickness of a current at a specific point in time represents the strength of the topic in the associated documents. TimeMines [28] and Narratives [8] are examples for visualizations that are based on standard line charts. TimeMines automatically determines keywords and judges those keywords with respect to their temporal significance in the context of the corpus. Furthermore, keywords that show to have a similar development over time are grouped to form a topic. Narratives presents the development of a specific topic over time and searches for correlated terms.

A similar concept is reported in [12]. The system BlogPulse (that can be found at www.blogpulse.com) monitors blogs and displays timelines that show how many blogs talk about a specific topic at a specific point in time. In addition, hot topics are detected automatically. All of the mentioned time-oriented approaches have a common limitation: They merely display the development of the significance of keywords or topics over time. Our approach goes beyond that by means of additionally revealing the sentiment of the documents.

Two further approaches being related to our work are [2] and [11]. Both of them analyze blogs and / or newspaper articles with respect to their political orientation. However, none of the approaches explores the development over time as we do. Instead they both focus on analyzing the link structure between the different blogs respectively the citation patterns for newspaper articles. In addition, [11] takes into account how emotionally charged a post is.

Sentiment Analysis

Within the abundant literature that exists in the context of sentiment analysis and opinion mining, some major tasks can be identified:

- Classification of the statements of a document (or a sentence) as subjective or objective. (e.g. [29, 14])
- Classification of a document (or a sentence) as expressing a negative or positive sentiment (or opinion). (e.g. [25, 5])
- Feature-based opinion mining made up by two successive steps: First, the features (or attributes), that have been commented on, are identified. Secondly, the respective opinion that has been expressed on them is detected. (e.g. [17, 18, 26, 21, 20])

Note that our approach is not contributing to the area of automatic sentiment analysis but makes use of some of its standard techniques. However, we contribute to the development of visualizations for sentiment analysis. Related work in this respect includes [10, 24, 13]. The visualization, which shows to have the highest resemblance to our work, can be found in [24]. The authors suggest to use bars to visualize how many positive respectively negative statements – that comment on one of the analyzed attributes of a product – exist within the document corpus. Our work is similar in that we also use the vertical deflection of bars to encode the opinion that is expressed. In contrast to [24] however, in our case one bar represents one document instead of the summary of all sentences talking about a specific attribute of a product. Moreover, in our visualization the development over time is central, something that is completely omitted in all of the above mentioned approaches for sentiment analysis / opinion mining. In [10] customer reviews are visualized, too, but a Treemap representation is used to display the result of the analysis. Finally, [13] presents an adaptation of the Rose Plot visualizations to illustrate the affective content of a document. In addition to positive and negative sentiments, the documents are also analyzed with respect to the categories *virtue, vice, pleasure, pain, power cooperative, and power conflict*.

VISUAL SENTIMENT ANALYSIS

Data Processing

The data we used was gathered from 50 different RSS news feeds, that mainly dealt with the 2008 US presidential elections. The RSS feeds were retrieved every 30 minutes during a time interval of one month (10/09/2008 - 11/10/2008). For every news item in each feed we saved date, title and description, as well as the id of the feed. Next, noise was eliminated out of the title and description. With noise we refer to strings that do not carry any content, such as URLs or strings consisting of special characters. The concatenation of title and description was then considered to be the content of the news item. Finally, we filtered out those documents that contained none of the following signal words: "Obama", "McCain", "Biden", "Palin", "Democrat" and "Republican". More than 23,000 news items contained at least one of the six strings.

Pairwise similarities between news items were calculated by applying a similarity measure, which counts the number of non-stopwords that two items have in common (normalized by the length of the larger item). Although this is a relatively simple measure it works quite well for the short descriptive texts in the RSS news feeds.

Another aspect of interest is the sentiment context of a news item, which is done by enriching each item with a sentiment score. For this purpose we make use of a freely available list of words that evoke positive or negative associations [4]. We count the number of positive and negative words and evaluate the whole news item as rather positive if it contains in total more positive than negative words. Likewise, the item is evaluated as rather negative if it contains more negative than positive words. The absolute relation of positive against negative words normalized by the item's length, provides our sentiment score. One important point to mention here is that the appearance of a candidate, e.g., in a negative context, does not necessarily mean, that the item contains negative publicity for the candidate, but simply that he appears in a negatively connoted context. This becomes clear when we consider the example of news telling that racists planned to assassinate Obama (see section "Results"). This was bad news for Obama not about Obama, with a visibly negative connotation.

Data Visualization

The visualization on the one hand aims to give a meaningful representation of the data and on the other hand is intended to be an appropriate starting point for the interactive exploration and discovery of interesting patterns. Figure 4 shows a screenshot of the visualization. Each line represents one day and each colored object depicts one news item. The news item's emotional score is encoded by a vertical displacement of the news item. Colors encode whether the text mentions the Democratic party, the Republican party or both. Additionally, the shape of the news objects visualizes whether the first candidate, the second candidate or only the name of the party itself was mentioned. The following passages describe each of those aspects in detail.

Placement

Every news item is represented by an object in a 2D plane. The position of the object within the plane depends on the date the news was published. Thereby, the day it was published accounts for the line it will be placed in (as each line represents one day) and the time of day determines its horizontal position within the line. The exact vertical position depends on the sentiment score of the object. According to this value an object is slightly shifted up (positive) or down (negative). Horizontal lines mark the position that a news item would have that is neither positive nor negative.

Coloring

Everything that is solely related to the conservatives (Republican party) is colored in red and everything purely related to the liberals (Democratic party) in blue. Gray news objects relate both to the liberals and the conservatives, which basically means that both camps are mentioned within the news' content.

Shape

The use of different shapes for the object allows us to make a distinction between news items in which the first candidate of a party was mentioned, the second candidate but not the first candidate or none of them but only the name of the party. Figure 1 shows the visual appearance of the different shapes. Please note that we keep the horizontal interruptions that are utilized to mark news items that talk about the second candidate always at the same vertical position of each line (regardless of the vertical shift of the object that encodes the emotional score). This leads to a clear visual pattern of continuous white horizontal lines, if several neighboring objects refer to the second candidates only.

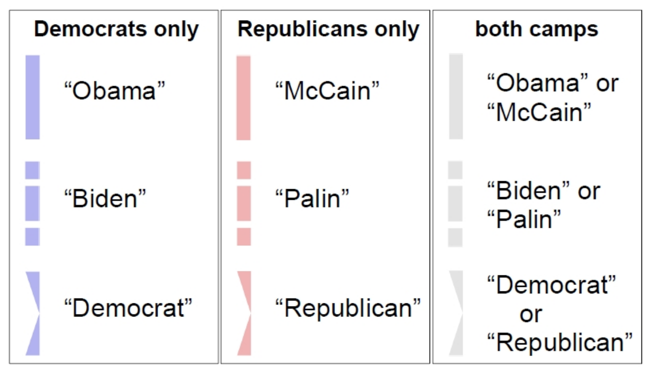

Figure 1. Symbols used to represent news items according to the appearance of certain keywords.

Opacity

We paint our news objects with a relatively low opacity. That means they are partly transparent, which comes with two advantages: First, the problem of overlapping news objects is reduced. In most cases every object is visible and can be differentiated clearly from its overlapping neighbors. Secondly, if multiple news items are put on top of each other, the overall opacity at this position increases, resulting in an object that is less opaque and can therefore be distinguished from objects that represent just one news item. The situation that several feeds bring the same news nearly at the same moment in time is often the case when the news is very important. That means that the less opaque news objects often represent news that are more important and surely more widely spread. Figure 2 visually illustrates the above mentioned design decisions.

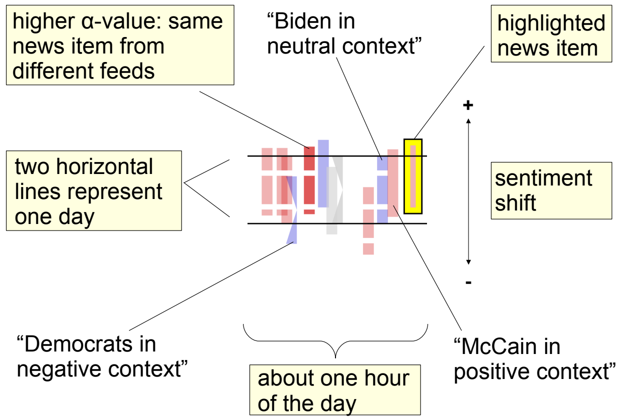

Figure 2. Semantics of the visualization

Interactive Visual Analytics

The visualization is designed for an interactive data exploration. There are several possibilities to interact with the tool:

- Zooming: Continuous zooming allows to analyze certain parts at a greater level of detail.
- Details on demand: When the mouse is dragged over a news object, a tooltip appears containing date, time, feed id, and content of the item.
- Similarity search: With a mouse click on a news object, the search for similar news items is started. The news item itself and every other news object that is related to it is highlighted (please refer to section "Data Processing" for our definition of similarity). Figure 3 shows an example.
- Filtering: The user can select the different candidates / parties he is interested in. Another possibility to reduce the number of items that are displayed is to select one specific RSS feed. Both filtering mechanisms can be used to analyze in detail the behaviour of one specific news provider respectively the development of news for a subset of candidates and/or parties.

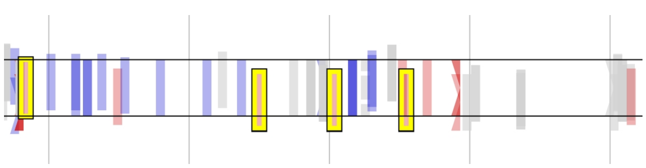

Figure 3. After selecting one news item, similar items are highlighted in yellow enabling the user to track specific topics (low threshold) or redundant postings (high threshold).

RESULTS

First of all, we present an overview of all 50 monitored RSS feeds over a time period of 31 days in Fig. 4. A predefined filter displays all news postings containing at least one of the terms Obama, McCain, Biden, Palin, Democrat, and Republican. To exemplify our Visual Analytics technique, we picked five interesting discussions in the monitored RSS feeds.

Palin abused power in Alaska

On Saturday, 10th October, many negative news postings occurred about Sarah Palin. Almost all articles deal with the topic whether Sarah Palin had abused her power in Alaska or not. As demonstrated in Fig. 5 there is a high density of red shapes with two white bars symbolizing news postings about Palin. Their positions below the baseline denote that mainly negative emotion words were used in these postings. Only one exceptionally positive red news item sticks out in the visualization. A closer look at this posting reveals that it is a response from the McCain-Palin presidential campaign: "Sarah Palin acted 'within proper and lawful authority' in removing the state's public safety commissioner".

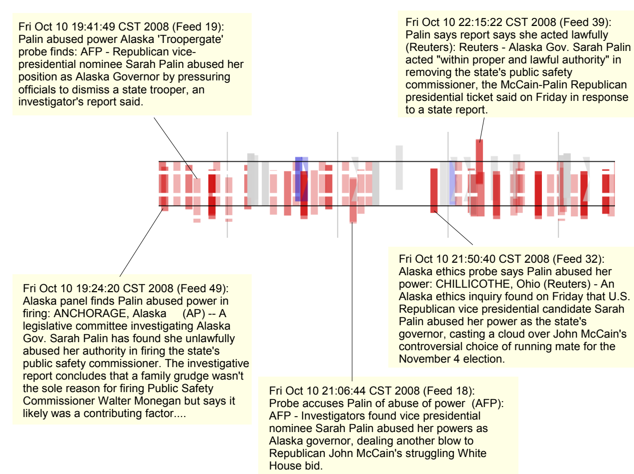

Figure 5. Media coverage dealing with the topic of Sarah Palin's abuse of power as a governor of Alaska.

Bad news for the Democrats

Approximately one week before the US presidential election we detected a high appearance of news which included "Obama" (see Fig. 6). The sentiment scores of these postings were mainly negative and dealt with a plot to assassinate Barack Obama and 102 blacks. Note that the news are bad for him but not about him, meaning that a negative event is related to him in the news postings although the negative opinion words do not refer to him as a person.

The used emotion words were so strong, that even in the overview it is possible to recognize the emergence of the negative news of that event on 28th of October (see Fig. 4). Note that although each RSS posting only consist of a few sentences, the few contained positive or negative opinion words are sufficient to provide clear results. Further headlines of that day discuss the corruption scandal of a Democratic senator and result in negative headlines for the Democrats.

TV debate Obama vs. McCain

In the middle of October the final TV debate between the Democrat candidate Barack Obama and the Republican can-

Figure 4. 31 days of the 2008 US presidential election showing a scandal of power abuse by Palin (A), the TV debate McCain vs. Obama (B), assassination plans against Obama (C), the election day (D), and a debate about Palin's election wardrobe (E).

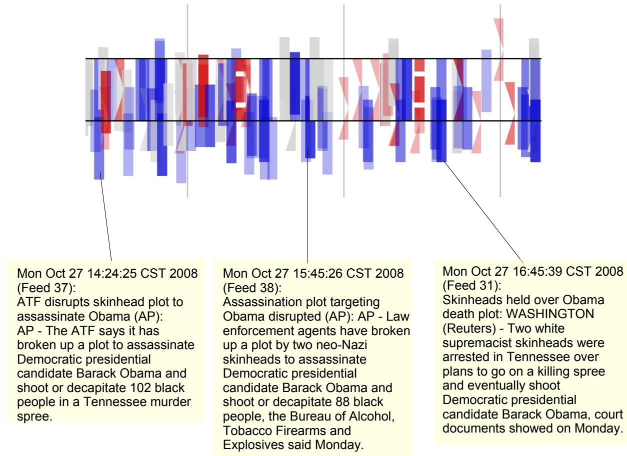

Figure 6. Democrats appears in "negative context". Bad news for Obama, but not about him.

didate John McCain was held. As shown in Fig. 7, news postings of the event cover both candidates (gray) and generally have low sentiment scores due to the criticism of both candidates against each other. The debate revealed little novelty with respect to each candidate's political plans after the election. Therefore, there were no strong positive statements about the event in the monitored feeds.

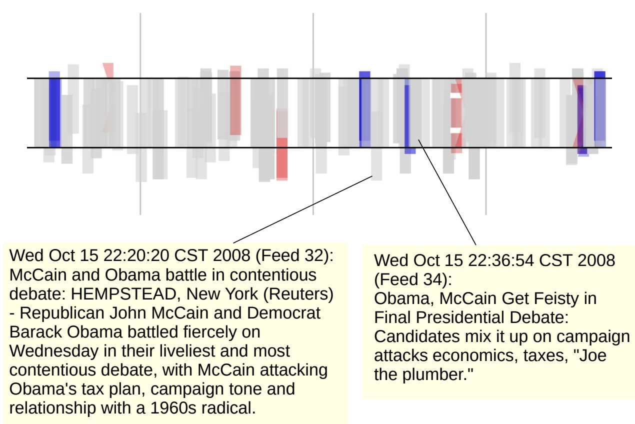

Figure 7. TV debate

Obama wins the election

As you can see in annotation D in Fig. 4 the election day is dominated by gray bars. This is due to the fact that these news postings reported about election results in particular states, featuring scores of both candidates. In the evening of the election day lots of news postings were received about the winner Barack Obama. The density of news about the Democrats increased rapidly after the result was known and dominate the news for several days.

Palin's wardrobe

Although after the election the blue shapes increased immensely, some red negatively rated items stick out (see Fig. 8). These outliers deal with some critical notes about the expensive wardrobe, which was bought by Sarah Palin for her campaign, and her inappropriate use of language describing her critics.

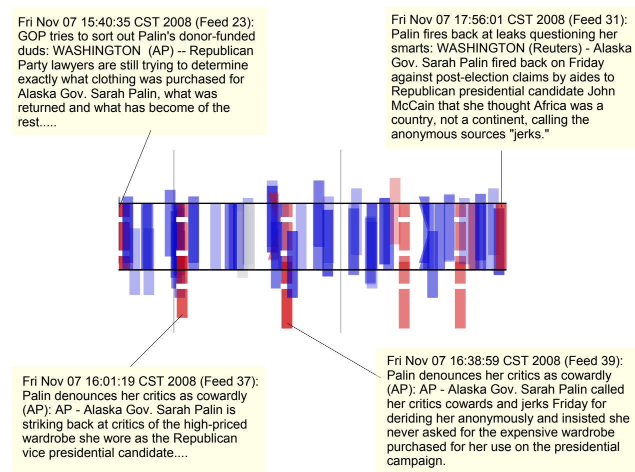

Figure 8. Palin under attack after the elections.

Further trends

The Democratic vice presidential candidate Joe Biden, who is represented by blue bars with two interruptions, was not referenced often. As it can be seen in Fig. 4, he appears very rarely compared to the Republican vice presidential candidate Sarah Palin.

A further discovery was that some feeds show daily patterns. For example, one RSS-feed only sent messages in the morning at about 7AM, others broadcast their news during working hours and some feeds even switched the coverage of political events within daily patterns, which is probably due to two editors each preferring news about one party and taking turns in writing news postings.

Often, the same news story is broadcasted in many different feeds (e.g., the above mentioned news about Palin's wardrobe). This is mainly due to the fact that some feeds immediately broadcast the news copied from a particular news agency, whereas other feeds broadcasted this information later. Another feed resent the same news posting several times as shown in Fig. 9.

CONCLUSIONS

The main contribution of this paper is the combination of a sentiment analysis method with a visualization technique revealing the emotional content of RSS news feeds over time. Through textual filters, we focused our analysis on the 2008 US presidential election featuring positive and negative news items about the presidential candidates Obama and McCain, the vice president candidates Biden and Palin and the two major parties. The timeline visualization builds upon three basic elements, first the attribute color denotes the political party featured in the news article, second, different shapes are used to distinguish between the discussed persons, and third, the emotional score of each RSS news article resulted

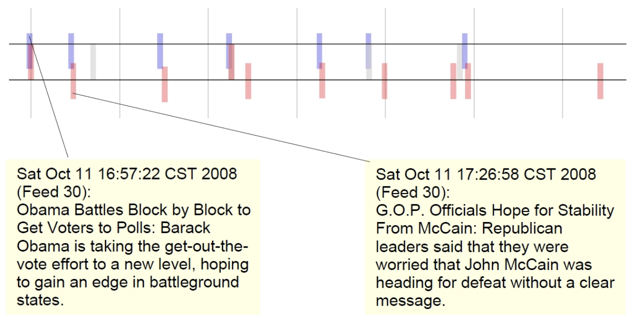

Figure 9. Technical failure or search engine optimization resulting in resending the same news postings over and over again.

in the vertical position of the representative symbol on the time line.

Within the result section, we showed how some emotional discussions manifested in our news visualization: 1) Palin abused power in Alaska, which resulted in many negative news items and her own version sticking out as a highly positive article. 2) The story about assassination plans against Obama dominated the news for several hours with highly negative sentiment scores. 3) The final TV debate consisted of mainly gray elements since reports featured both candidates. In general, the accusations of both candidates against each other resulted in more negative than positive sentiment scores. 4) Obama wins the elections, which is documented by the vast dominance of blue news elements on the eve of the election day and the following days. 5) Even after the election a discussion about the expensive wardrobe of Palin fills negative headlines.

The tool's interaction concept shows the corresponding RSS news articles when the mouse is moved over a symbol on the timeline. To find redundant or similar news items in the process of analyzing particular events, we furthermore implemented a simple document similarity filter, which after selecting a particular news item highlights all related news postings surpassing a certain threshold of similarity.

We believe that the presented analysis tool can not only be used to monitor public emotional discussions, but is also capable of evaluating product reviews, public opinions on a particular subject, or to get hints about the reputation an enterprise. By offering sentiment analysis functionality of a multitude of large RSS feeds in real-time, users of this technique can take early action, such as reacting before a topic dominates news coverage. This strategic dimension of our application is very valuable for public relation specialists and could be implemented in early warning systems. Furthermore, we expect the tool to be useful for monitoring the evolution of popularity of certain products, persons, or views, ultimately answering the question about why a positive public image turned into a negative one.

Future Work

For computing the similarity between news items we used a simple word matching method. Due to the fact that many news items are copied from other news tickers, related RSS postings are often based on the text of the same announcement of a newswire and therefore often contain almost identical vocabulary. For the analysis of other content, such as product reviews or the full articles linked in the RSS tickers, more complex document similarity measures could be employed. Furthermore, we believe that more sophisticated sentiment analysis methods can be integrated into the presented analysis tool.

Acknowledgement

This work has been funded by the research center "Computational Analysis of Linguistic Development" at the University of Konstanz and by the German Research Society (DFG) under the grant GK-1042, Explorative Analysis and Visualization of Large Information Spaces, Konstanz.
We thank the anonymous reviewers of the VISSW 2009 for their valuable comments.

REFERENCES

1. A. Abbasi and H. Chen. Categorization and analysis of text in computer mediated communication archives using visualization. In *JCDL '07: Proceedings of the 2007 conference on Digital libraries*, pages 11–18, New York, NY, USA, 2007. ACM.

2. L. A. Adamic and N. Glance. The political blogosphere and the 2004 U.S. election: divided they blog. In *LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM, 2005.

3. T. Ball and S. G. Eick. Software Visualization in the Large. *IEEE Computer*, 29(4):33–43, 1996.

4. V. Buvac. Internet General Inquirer, 2008. http://www.webuse.umd.edu:9090/ as retrieved on Nov. 14, 2008.

5. K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 519–528. ACM, 2003.

6. A. Don, E. Zheleva, M. Gregory, S. Tarkan, L. Auvil, T. Clement, B. Shneiderman, and C. Plaisant. Discovering interesting usage patterns in text collections: integrating text mining with visualization. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 213–222. ACM, 2007.

7. J.-D. Fekete and N. Dufournaud. Compus: visualization and analysis of structured documents for understanding social life in the 16th century. In *DL '00: Proceedings of the fifth ACM conference on Digital libraries*, pages 47–55, New York, NY, USA, 2000. ACM.

8. D. Fisher, A. Hoff, G. Robertson, and M. Hurst. Narratives: A Visualization to Track Narrative Events as they Develop. In *IEEE Symposium on Visual Analytics and Technology (VAST 2007)*, pages 115–122, 2008.

9. B. Fortuna, D. Mladenic, and M. Grobelnik. Visualization of Text Document Corpus. *Informatica Journal*, 29(4):497–502, 2005.

10. M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger. Pulse: Mining Customer Opinions from Free Text. In *Advances in Intelligent Data Analysis VI*, pages 121–132. Springer, 2005.

11. M. Gamon, S. Basu, D. Belenko, D. Fisher, M. Hurst, and A. C. König. BLEWS: Using Blogs to Provide Context for News Articles. In *ICWSM*, 2008.

12. N. Glance, M. Hurst, and T. Tomokiyo. BlogPulse: Automated Trend Discovery for Weblogs. In *WWW 2004 Workshop on the Weblogging Ecosystem*. ACM, May 2004.

13. M. L. Gregory, N. Chinchor, P. Whitney, R. Carter, E. Hetzler, and A. Turner. User-directed Sentiment Analysis: Visualizing the Affective Content of Documents. In *Workshop on Sentiment and Subjectivity in Text*, pages 23–30, 2006.

14. V. Hatzivassiloglou and J. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity, 2000.

15. S. Havre, E. Hetzler, P. Whitney, and L. Nowell. ThemeRiver: Visualizing Thematic Changes in Large Document Collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):9–20, 2002.

16. M. A. Hearst. TileBars: Visualization of Term Distribution Information in Full Text Information Access. In *Proceedings of the Conference on Human Factors in Computing Systems, CHI'95*, 1995.

17. M. Hu and B. Liu. Mining and summarizing customer reviews. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

18. M. Hu and B. Liu. Mining Opinion Features in Customer Reviews. In *AAAI*, pages 755–760, 2004.

19. D. A. Keim and D. Oelke. Literature Fingerprinting: A New Method for Visual Literary Analysis. In *EEE Symposium on Visual Analytics and Technology (VAST 2007)*, pages 115–122, 2007.

20. S.-M. Kim and E. Hovy. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text*, pages 1–8, 2006.

21. N. Kobayashi, K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima. Collecting Evaluative Expressions for Opinion Extraction. In *IJCNLP*, pages 596–605, 2004.

22. R. R. Korfhage. To see, or not to see – is That the query? In *SIGIR '91: Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 134–141. ACM Press, 1991.

23. K. Lagus, T. Honkela, S. Kaski, and T. Kohonen. Self-organizing maps of document collections: A new approach to interactive exploration. In E. Simoudis, J. Han, and U. Fayyad, editors, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 238–243. AAAI Press, 1996.

24. B. Liu, M. Hu, and J. Cheng. Opinion observer: analyzing and comparing opinions on the Web. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 342–351. ACM, 2005.

25. B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pages 79–86. Association for Computational Linguistics, 2002.

26. A.-M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 339–346. Association for Computational Linguistics, 2005.

27. A. Spoerri. InfoCrystal: a visual tool for information retrieval & management. In *CIKM '93: Proceedings of the second international conference on Information and knowledge management*, pages 11–20. ACM, 1993.

28. R. Swan and D. Jensen. TimeMines: Constructing Timelines with Statistical Models of Word Usage, 2000.

29. B. Wang, B. Spencer, C. X. Ling, and H. Zhang. *Semi-supervised Self-training for Sentence Subjectivity Classification*, pages 344–355. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2008.

30. J. A. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the non-visual: spatial analysis and interaction with information from text documents. In *INFOVIS '95: Proceedings of the 1995 IEEE Symposium on Information Visualization*, pages 51–58, 1995.