

Semantize: Visualizing the Sentiment of Individual Document

Alan J. Wecker, Joel Lanir, Osnat Mokryn, Einat Minkov, Tsvi Kuflik
The University of Haifa

ajwecker@gmail.com, ylanir@is.haifa.ac.il, ossi@mta.ac.il, einatm@is.haifa.ac.il,
tsvikak@is.haifa.ac.il

ABSTRACT

A plethora of tools exist for extracting and visualizing key sentiment information from a corpus of text documents. Often, however, there is a need for quickly assessing the sentiment and feelings that arise from an individual document. We describe an interactive tool that visualizes the sentiment of a specific document such as an online opinion, blog, or transcript, by visually highlighting the sentiment features while leaving the document text intact.

Categories and Subject Descriptors

H.5.m. Information interfaces and presentation (e.g., HCI):
Miscellaneous.

General Terms

Human factors

Keywords

Sentiment analysis, Visualization, User interface

1. INTRODUCTION

Over the last years, there has been tremendous growth in the volume of Web-based opinionated information in the form of blogs, user forums, and other types of social media. Sentiment analysis methods are typically applied to automatically identify the sentiment orientation of text [3]. Many visualization tools exist that compactly present the overall sentiment observed in large text collections, typically showing their aggregation, summary and underlying trends [1, 2, 4].

In this paper, we consider a different setting of sentiment visualization, where we wish to illuminate the sentiment discourse of an individual document. Visualizing a document inline can help draw out attention to positive, or negative, sentiment cues expressed in the text, which may otherwise be overlooked. For example, when reading an email message, it is often important to correctly decode the underlying sentiment: Is my boss really angry with me? Is this message trying to please me? Similarly, when composing a message, it may be useful to understand how

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

AVI '14, May 27 - 30, 2014, Como, Italy

ACM 978-1-4503-2775-6/14/05.

http://dx.doi.org/10.1145/2598153.2600056

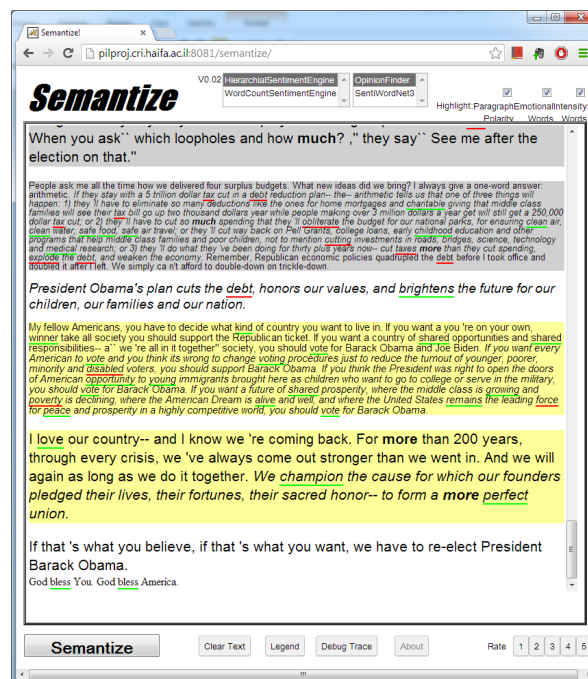


Figure 1: An example of Semantize.

the text may be perceived by its recipients. Furthermore, an inline sentiment visualization tool can support the reading of emotion-rich texts, such as reviews, or political texts. In particular, given long documents, such a visualization tool may allow one to skim through a received scientific review, or a literature critic, identifying sentiment intense text. To the best of our knowledge, such a visualization tool does not currently exist.

We describe *Semantize*, a document sentiment viewer prototype. Given text, *Semantize* provides visual indication of automatically assessed sentiment. For example, it highlights subjective paragraphs, along with their polarity. Figure 1 shows *Semantize* displaying the transcription of speech by former US President Bill Clinton at the 2012 Democratic Convention. Using *Semantize*, it is easy to see which paragraphs showed positive (yellow) or negative (gray) sentiment. We can also see other sentiment related cues such as subjective speaking (italic) and emotion words (color). *Semantize* was designed and built in a modular manner: a backend engine performs the sentiment analysis, and a frontend service is responsible for the user interface. The interface enables the user to choose among several sentiment features, and visualizes the text accordingly.

2. SYSTEM DESCRIPTION

Semantize aims at enabling the users to quickly grasp the emotional effect of the text by annotating the text itself while leaving its structure intact. Given a new text, the sentiment analysis values are first computed. Then, the computed values are displayed graphically on top of the text using CSS style sheets. This representation is highly flexible, where each sentiment feature type is mapped to some specified CSS formatting. Different formats can be formed using font properties, including size, color, spacing, animation, background color and so forth.

In our current implementation, five major sentiment features are displayed. We mapped polarity on a paragraph level using background color; subjective sentences were marked with italics; emotion words were marked with underline color (green and red); intensive or diminutive words used slightly larger or smaller bold font; and grade level was marked on a paragraph level using font size and spacing. Our main guideline in mapping the features to styles was maintaining a clear and non-cluttered presentation of the various sentiment indicators, which often overlap. In addition, intuitive options were used whenever possible; for example, a green font color to display positive emotional words and red font color for negative emotional words.

We acknowledge that different visualization options may be preferred by different users, as well as for different text genres. The system's UI is flexible, allowing the user to eliminate each of the sentiment visualization types from the display, using check boxes. This customization can be used for personalization, and is also useful in experimental setting, where one may wish to assess the utility of displaying different combinations of information types. In addition, upon pressing the "legend" button, a legend showing the mapping of features to styles as well as a short summary of the overall sentiment values is presented (e.g., the amount of positive and negative emotional words found). Finally, the selections of mapping of feature to style and level can be easily altered using variations of the CSS style sheet.

The backend engine provides the sentiment feature values. It generates a marked up text in XML format that is visualized by the front end UI. We implemented well-studied and efficient sentiment analysis techniques.

3. INITIAL EVALUATION

An informal qualitative evaluation of Semantize was performed in order to get an initial impression of the system's effectiveness. Eight participants (seven male, avg. age: 29.2), all non-native English speakers with a good level of English reading (self-defined) participated in the evaluation. After participants were shown how Semantize works, they were presented with four text documents of two domains and were asked to assess the document's general sentiment. The first domain included hotel reviews taken from the TripAdvisor website, one positive (5 stars) and one negative review (1 star). The second domain included two documents of academic peer-reviews, one positive (an accept) and one negative (a reject). In each domain one document was presented using Semantize, while the other was presented with no annotations. Participants were asked to state their opinions using several 7-point Likert scale questions, followed by open-ended questions.

In general, participants appreciated the use of Semantize and thought that it was very useful to quickly assess the sentiment of the text. Figure 2 presents the score distribution of several 7-point Likert scale statements. No scores were below Neutral (4 on the

Likert scale), and the averages to all questions were between 5.12 and 5.70.

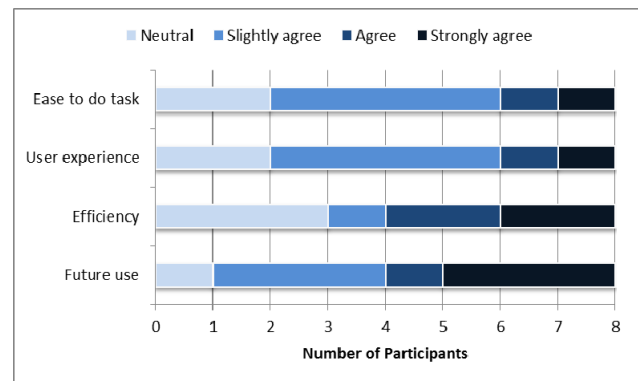


Figure 2: Distribution of scores per statement. For each statement, the distribution shows the number of participants who selected the same specific value.

In open-ended questions and interviews, participants commented that the main advantage of Semantize is its ability to quickly convey the sentiment and get the gist of the document ("[Semantize] provides help to characterize the text in a quick and efficient way"). In addition, participants commented that the system is advantageous for non-native speakers for whom it is often more difficult to assess a document's sentiment. Regarding the disadvantages of the system, most participants commented on inaccurate annotations. The system sometimes missed emotion words or wrongly marked other words. We note that while some of these problems stem from inherent difficulties in sentiment analysis, others can be fixed with a better (more precision-oriented) sentiment analysis engine.

4. CONCLUSIONS

We presented Semantize, a document sentiment viewer prototype that aids readers in quickly assessing various sentiment aspects of a single text document. Future work would examine how different UI mappings between the features and the levels and styles might affect users' perception of the service.

5. REFERENCES

- [1] Gregory, M. L., Chinchor, N., Whitney, P., Carter, R., Hetzler, E., & Turner, A. User-directed sentiment analysis: Visualizing the affective content of documents. *Workshop on Sentiment and Subjectivity in Text*. ACL (2006), 23-30.
- [2] Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., & Miller, R. C. Twitinfo: aggregating and visualizing microblogs for event exploration. *In proceedings of CHI'11*. ACM Press. (2011) 227-236.
- [3] Pang, B., & Lee L. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval 2.1-2* (2008) 1-135.
- [4] Zhang, J., Kawai, Y., Kumamoto, T., & Tanaka, K. A novel visualization method for distinction of web news sentiment. *In Web Information Systems Engineering-WISE*. Springer Berlin Heidelberg. (2009) 181-194.