Visualizing Sentiment: Do You See What I Mean?

Alan J. Wecker
Information Systems
University of Haifa
Haifa, Israel
ajwecker@gmail.com

Joel Lanir
Information Systems
University of Haifa
Haifa, Israel
joel.lanir@gmail.com

Einat Minkov
Information Systems
University of Haifa
Haifa, Israel
einatm@is.haifa.ac.il

Tsvi Kuflik
Information Systems
University of Haifa
Haifa, Israel
tsvikak@is.haifa.ac.il

Osnat Mokryn
Computer Science
Tel Aviv-Yafo College
Tel Aviv, Israel
ossi@mta.ac.il

IUI'14, Feb 24-27 2014, Haifa, Israel
ACM 978-1-4503-2729-9/14/02.
http://dx.doi.org/10.1145/2559184.2559204

Abstract
Many tools exist for extracting and visualizing key information from a corpus of text documents. However often, one would like to assess the sentiment and feelings that arise from a single document. This paper describes an interactive service that visualizes the sentiment of a specific document. The service enables the user to visualize the sentimental polarity of each paragraph to get a detailed impression; to quickly detect the polarity of emotional words; to identify subjective sentences within the text, and the grade level of language used in each sentence. Participants in an initial qualitative evaluation found the service fast and useful.

Author Keywords
Sentiment analysis; visualize; emotion polarity; subjectivity; user interface;

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

General Terms
Human Factors; Design;

General Background Material

Information Visualization

Fortuna, B., Mladenić, D., & Grobelnik, M. *Visualization of text document corpus*. 2005

Jia, L., Yu, C., and Meng, W. The effect of negation on sentiment analysis and retrieval effectiveness. *In Proceedings of the 18th ACM conference on Information and knowledge management* (2009) 1827-1830.

Šilić, A., & Bašić, B. D. Visualization of text streams: a survey. *In Knowledge-Based and Intelligent Information and Engineering Systems.* Springer Berlin Heidelberg. (2010), 31-43

Ware, C. *Information visualization* (Vol. 2). San Francisco: Morgan Kaufmann. 2000

Sentiment Analysis

Feldman R. Techniques and applications for sentiment analysis. Commun. ACM 56, 4 (2013), 82-89.

Pang, B., & Lee L. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval 2.1-2* (2008) 1-135.

Tang, H., Tan, S., & Cheng, X. A survey on sentiment detection of reviews. *Expert Systems with Applications*, *36*, 7 (2009), 10760-10773.

Introduction

In today's web environment, we express our opinions in blogs, forums and in social media. This revolution has resulted in an enormous growth in Web-based information. To cope with this information overload, a variety of tools for extracting, tagging and visualizing information digests emerged. Documents are summarized, labeled and aggregated, and visualization tools are developed for a smart presentation of these results [e.g., 2, 11]. Recently, sentiment visualization, in the context of document summarization, draws attention as a tool that can ease the understanding of sentiments expressed in text [4, 5]. However, most of the research focuses on compact presentation of sentiments of a corpus of text documents, and does not provide a view of the sentiment of a single document.

Often the sentiment expressed within the text of a document is important. For example: a political journalist receiving the transcript of a speech about to be given, might want to quickly analyze the pattern of emotions and sentiment polarity throughout the speech; or a consumer browsing product reviews might be interested in rapidly determining the polarity of different aspects in the review. Alternatively, the consumer might be interested in finding the sentiment expressed in the description of a certain product aspect that appears in a particular paragraph. Clearly, a service that visually highlights the sentiment polarity of each paragraph can aid in the above scenarios.

Related Work

Information visualization generally aims at providing new insights and helps understand some structure of the data. Text visualization is a subfield of it that deals with the presentation and visual analysis of texts. Thus, our work of highlighting sentiment patterns in a text document can be seen as related to the general area of text visualization. There are many ways to analyze and visualize text. Much of the work in this area has focused on analyzing large corpora of texts such as news articles [3] and other types of collections [10]. In Sentiment map [11], sentiment tendencies of news articles are aggregated based on time and geographic scales. In [2], an interactive visualization that examines the general sentiment of a large document collection is explored. TwitInfo [4] summarizes events on Twitter. Among its other features, it also colors tweets in red if they show negative sentiment or in blue if they show positive sentiment.

Focusing on visualizing a single text document, tag clouds [7] show frequency of words in a document laying out prominent works using size and color. WordTree [9] also visualizes word frequencies using size, but shows the words in the context of their sentences using a tree-like structure. Other works on single document visualization, such as TextArc [6], focus on the connection between various terms within the document.

Common to these works on single document visualization, is that they all visualize single terms, connections between terms, or other features that characterize the text. Few works leave the text itself as is and simply change the typeset for highlighting underlying concepts. An exception is Jigsaw [8] that included highlighting of different entity types of the text using a color code. Most related to our work is ScentHighlights [1], which uses automatic highlighting to help users skim and detect relevant sentences over text documents. Topic interests are obtained from

Feature	Level	Style
Polarity	P	Background Color (yellow=positive; gray=negative)
Subjective	S	Italics (*cursive*=subjective; normal=objective)
Emotion	W	Word Underline Color (green=positive; red=negative)
Intensive / Diminutive	W	Intensive= **slightly larger size + bold font**, Diminutive= **slightly smaller size + bold font**
Grade Level	P	Font size and spacing (larger font size and spacing show lower grade level)

Levels
P=Paragraph,
W=Word,
S=Sentance

Table 1: Sentiment Features Representation

user-entered keywords, and relevant words, sentences and paragraphs are highlighted accordingly. Similarly, we guide the user's attention during the reading of text; however, in our work the focus is on sentiment analysis and the aim is to aid users in recognition of the text's sentiment and the feeling expressed in it.

System description

Semantize consists of a UI frontend and a sentiment analyzer backend.

User Interface

The UI presentation method aims at enabling the users to quickly grasp the emotional effect of the text by annotating the text itself while leaving the structure intact. Five major sentiment features were chosen; each one is annotated on a different level (word, sentence, paragraph), and in a different style. The list of features and our choices of levels and styles for each feature are shown in Table 1.

Given a new text, the sentiment analysis values are computed as described in the next section. Text stylization displays the computed values graphically 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 style, size, color, spacing, animation, background color and so forth. 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 of the goals may be to assess the utility of displaying different combinations of information types. In addition, our selections of mapping can be easily altered. Alternative mapping may be given by using variations of the CSS style sheet, and their level can also be altered. An optimization of the mapping of feature to level and style is out of the scope of this work and is left as future work.

Sentiment analysis backend

In this exploratory work, well-studied sentiment analysis techniques, for which the implementation and computation costs are low, were implemented. Given a body of text, it is represented as a non-structured `bag-of-words'. The text is parsed, splitting it into words, sentences and paragraphs, and words are annotated with part-of-speech tags. Words are then assigned a sentiment polarity score (positive vs. negative), using available sentiment lexicons. In addition, words are associated with the emotions that they evoke. Another text processing step identifies negation words like "not" or "never", using a small handcrafted word list. Specifically, scores can be aggregate at document, paragraph, and sentence levels. The focus of this work is on sentiment visualization rather than on optimizing sentiment analysis performance. Importantly, however, the service can readily support alternative sentiment inference methods, if desired.

Lexicons used

Lexicon	Purpose	**Note**
Opinion Finder	Polarity	1
SentiWord Net	Polarity	2
NRC	Emotion	3

[1] Baccianella, S., Esuli, A., and Sebastiani, F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *LREC*, *10* (2010), 2200-2204

[2] Baccianella, S., Esuli, A., and Sebastiani, F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *LREC*, *10* (2010), 2200-2204

[3] Mohammad S. M., and Turney P. D, Crowdsourcing a Word-Emotion Association Lexicon, *Computational Intelligence*, *29*, 3 (2013)

Future Work

Future work would involve an extensive survey that will measure the impact of different UI mappings between the features and the levels and styles on users' perception of the service. Additionally, one could use the annotated text backend as input for a Text to Speech system, in order to enable it to give an improved output emphasizing sentiment and emotion.

Demonstration

This work demonstrates the ability of visualization techniques to enable readers to get at a glance an overall sentimental representation of a document as well as the sentiments expressed by its components. The chosen architecture for the Semantize shows how to enable different choices for the UI as well as for the backend engine. We demo the system with a wide variety of relevant texts (such as: hotel reviews, conference reviews, etc.) showing how the system visualizes the sentiments of each of these texts. Users can play with the parameters of the system, visualizing different aspects of the texts. We anticipate that this demo will show the potential of this technology, and generate new ideas for possible applications.

Acknowledgements

We thank Moran Nave and Lena Kozhuhov for work on earlier version of the project.

References

[1] Chi, E. H., Hong, L., Gumbrecht, M., & Card, S. K. ScentHighlights: highlighting conceptually-related sentences during reading. *In Proceedings of the 10th international conference on Intelligent user interfaces,* ACM Press (2005), 272-274.

[2] Gregory, M. L., Chinchor, N., Whitney, P., Carter, R., Hetzler, E., & Turner, A. User-directed sentiment analysis: Visualizing the affective content of documents. *In Proceedings of the Workshop on Sentiment and Subjectivity in Text.* Association for Computational Linguistics (2006), 23-30.

[3] Grobelnik, M., Mladenic, D. (2004). Visualization of news articles. *Informatica journal*, *28,* 4 (2004).

[4] 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 the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press. (2011) 227-236.

[5] Miao, Q., Li, Q., & Dai, R. AMAZING: A sentiment mining and retrieval system. *Expert Systems with Applications, 36*, 3 (2009) 7192-7198.

[6] Paley, W. B. TextArc: Showing word frequency and distribution in text. *In Poster presented at IEEE Symposium on Information Visualization* (2002).

[7] Sinclair, J., & Cardew-Hall, M. The folksonomy tag cloud: when is it useful?. *Journal of Information Science*, *34*, 1 (2008), 15-29.

[8] Stasko, J., Görg, C., & Liu, Z. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, *7*, 2 (2008) 118-132.

[9] Wattenberg, M., & Viégas, F. B. The word tree, an interactive visual concordance. *Visualization and Computer Graphics, IEEE Transactions on*,*14*, 6 (2008), 1221-1228.

[10] Wei, F., et al. Tiara: a visual exploratory text analytic system. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM Press, (2010). 153-162

[11] 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.