

Integrating Interactivity into Visualising Sentiment Analysis of Blogs

Hyowon Lee, Paul Ferguson, Neil O'Hare, Cathal Gurrin and Alan F. Smeaton

CLARITY: Centre for Sensor Web Technologies

Dublin City University, Ireland

hlee@computing.dcu.ie

ABSTRACT

With an increased amount of freely available online resources and strong interest in automatic crawling and analysis on such resources, suitable visualisation techniques to present the results of such analysis present an important agenda for the visualisation research community. Interactivity on the Web has also become much more commonplace and acceptable as today's Web technologies such as Web 2.0 and Flash become more widespread. While conventional graphs and charts augmented with interactivity are one way to present the output of analysis, an interaction strategy that leverages the interactivity style of the Web should be more suitable than what we see today. We present a novel interactive visualisation technique designed and implemented on top of text-based sentiment analysis for financial blog posts where a user can easily search and browse bloggers' aggregated opinions on commercial companies in a way that helps understand the levels of online opinion in a summarised as well as a detailed manner.

Author Keywords

Interactive Visualisation, Sentiment Analysis, Financial Blogs

INTRODUCTION

Studies on crawling Websites such as blogs, news articles, product reviews and political columns and then automatically extracting useful information or determining their meaning, has become an increasingly active area of research. This helps realise the great potential for leveraging the rich online resources available today. Sentiment Analysis of blog posts is one of these efforts: trying to determine bloggers' opinions in terms of positive or negative sentiment, by analysing the text contained in the blog posts. As with any other similar text analyses, an important issue derived from such an analysis is how to present the results to users in a way that facilitates easy understanding of the overall trend of blog sentiment, as well as specific instances of such blogs.

The question we ask (and the solution we present) in this pa-

per is how we could leverage Web interactivity in visualising the sentiment of blogs as a result of crawling and analysing a large number of such blogs. Conventional charts and graphs are an effective static visualisation tool but how can we better incorporate the dynamic and higher levels of interactivity that suit the style of today's Web interaction ?

Considering the increasing interactivity on today's Web services and its' popularity, leveraging the style of Web interactivity and exploring visualisation strategies in that direction is well worth an investigation. In this paper we present a novel interactive visualisation strategy and the resultant Web interface allowing its users to interactively search and browse the output of sentiment analysis, similar to the way Web users search and browse Web pages using a Web search service.

VISUALISING SENTIMENT ANALYSIS — ALTERNATIVES TO PIE CHARTS AND BAR GRAPHS ?

Visualising the blogosphere has become an increasingly popular research challenge, and a number of graphical representations have been adapted to visualise the results of blog text analysis, although most of these are still at an experimental or planned, rather than deployed, stage. Examples include the use of different sizes of rectangular areas and colours as positive/negative opinion indicators [4], [1], spatial segmentation of Google News stories with colour-coded story categorisation¹, a pie-chart-like petal visualisation of different facets of sentiment such as Positive/Negative, Cooperative/Conflict, Pleasure/Pain and Virtue/Vice [6], the use of a large area to plot months of US presidential election events by colour-coding between Republican (Red) and Democrat (Blue) [13], multiple mini bar charts to visualise different facets of home electronics such as LCD, battery and speaker [11], blog visualisations where positive/negative emotions are colour-coded in blue/red and presented as a stack of horizontal bars [5], and where blog entry length, comment length, and the number of posts by the same bloggers are mapped to visual properties (colour, circle size and distance from timeline) [10]. More conventional graph/chart-based timeline visualisations of consumer-generated data include MoodViews [3], ThemeRiver [8], and a method that uses a Time Series Data Processing technique [2].

Other innovative blog visualisations include Twingly Blog-

Permission to make digital or hard copies of all or part 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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2009, April 3 - 9, 2009, Boston, MA, USA.

Copyright 2009 ACM 978-1-60558-246-7/07/0004...\$5.00.

¹Newsmap: <http://marumushi.com/apps/newsmap/index.cfm>

stream² which analyses and visualises the linkage status of bloggers in a time graph, where such linkage status amongst blog messages can be visualised as 3-D shapes [9]. “We Feel Fine” [7] crawls blog websites and identifies where a blogger’s feelings are mentioned, and visualises these as animated keyword clouds with a montage of images taken from the blog sites on a given day. We can imagine how these could be extended to the style of well-known 3-D visualisation techniques such as Cone-Trees, Document Lens and Perspective Wall where large, high-resolution monitors are usually assumed. In this paper we address the question of what would be an effective visualisation strategy, other than variations of conventional graphs and charts above, which is at the same time Web-friendly yet integrates even more user interactivity.

INTEGRATING INTERACTIVITY INTO SENTIMENT VISUALISATION

In this section, we briefly describe the sentiment analysis system we have developed, then present the novel interaction strategy and detailed design considerations and decisions we have taken to realise the strategy.

System for Sentiment Analysis on Financial Blogs

The sentiment analysis system at the back-end of the interface described in this paper was developed from a collaboration between Dublin City University and Zignals³, a company working in online stock trading. The aim of the system is to automatically extract subjective opinions found on blogs and to track the changing sentiment from the blogosphere towards individual stocks and the market in general. The system has been crawling financial weblogs from over 170 sources since May 2009, and has to date crawled over 44,000 relevant articles, namely those relevant to any company in the S&P 500 list (currently our system has analysed over 34,000 article-company matches). These are then analysed for sentiment (positive, neutral, negative) towards that company, using topic-based sentiment analysis approaches described in [12]. The results of this sentiment analysis is then aggregated for the interactive visualisation described in this paper.

Unit Representation

The core of our interactive visualisation is the concept of “Unit Representation”, a visual representation of the result of sentiment analysis on a particular object (a company in our case), serving as the building block of the overall interaction our visualisation uses for searching and browsing.

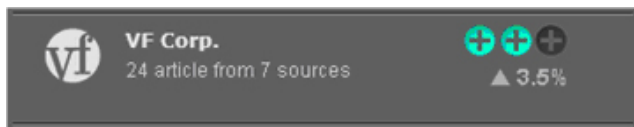


Figure 1. Unit representation - defining a visual representation for interactive querying and browsing

²Twingly Blogstream: <http://www.twingly.com/enterprise>

³Zignals: <http://www.zignals.com>

Figure 1 shows such a Unit Representation depicting (1) the name and logo of a company (VF Corp in this case), (2) the number of blog sources used for analysis (24 articles from 7 Websites in this case) and (3) the aggregated level of opinion (+2 on a scale between -3 and +3, and amounting to 3.5% increase compared to a previous time period in this case).

Searching and Browsing

Once the Unit Representation is defined as in Figure 1, then it can be used as a ‘virtual document surrogate’ that can be presented as a unit of retrieval on the user-interface. In other words, a user can conduct a search and the result is a list of Unit Representations, ranked initially by the order of the analysed level of opinions. The user can further select an entry in the search result, browse more details, check where the sources of this level of opinion came from, etc. Figure 2 shows a screenshot of the overall system interface.

On the top left of Figure 2, a user starts by selecting a category of companies (Finance, Medicine, Insurance, etc. currently Technology is selected in the figure) and the overall sentiment on all companies in the selected category is indicated on the right of the category selection. Below this, the user can further specify a time interval by clicking on mini calendar icons and selecting a date, or by clicking on common time interval types (week, month, and year) upon which ‘from/to’ boxes will be adjusted to the selected interval. When the user clicks on the ‘GO’ button, the result will be presented below, as a list of Unit Representations. The search result can be sorted by the levels of positive/negative sentiment, the rate of opinion change, company name, or the number of articles that mentioned the company, by clicking on the sorting buttons (below the ‘GO’ button). Clicking on any of the entries then presents all the articles that were used to derive the opinion rating in the Articles panel in the middle of the screen (in Figure 2 the user selected Microsoft and the Articles panel presents the articles that refer to Microsoft). At the top of this panel a summary of the opinions from all articles that refer to the selected company is presented, and below it is a list of articles in a summarised format. Each article entry also shows whether it has a positive, negative or neutral opinion about the company with a small colour circle beside the article’s title. The article entries can be sorted by the level of positiveness/negativeness, title of the article or by the date of post, similar to the sorting feature of the search result panel on the left. Clicking on the title of an article then opens up a web browser window and brings the user to the original online article so that the user can read the full article from the source Website. Finally on the right side of the screen is the Sources panel that shows a list of source Websites and the number of articles used in the analysis of the selected company, changing as the user selects a different company.

As can be seen in this interaction, the strategy employed turns the conventional concept of static graphical representation (which most financial information and company profile charts use) into an inherently interactive, search-like interface where the user starts with querying followed by sorting the search result to see the results in different orders and then

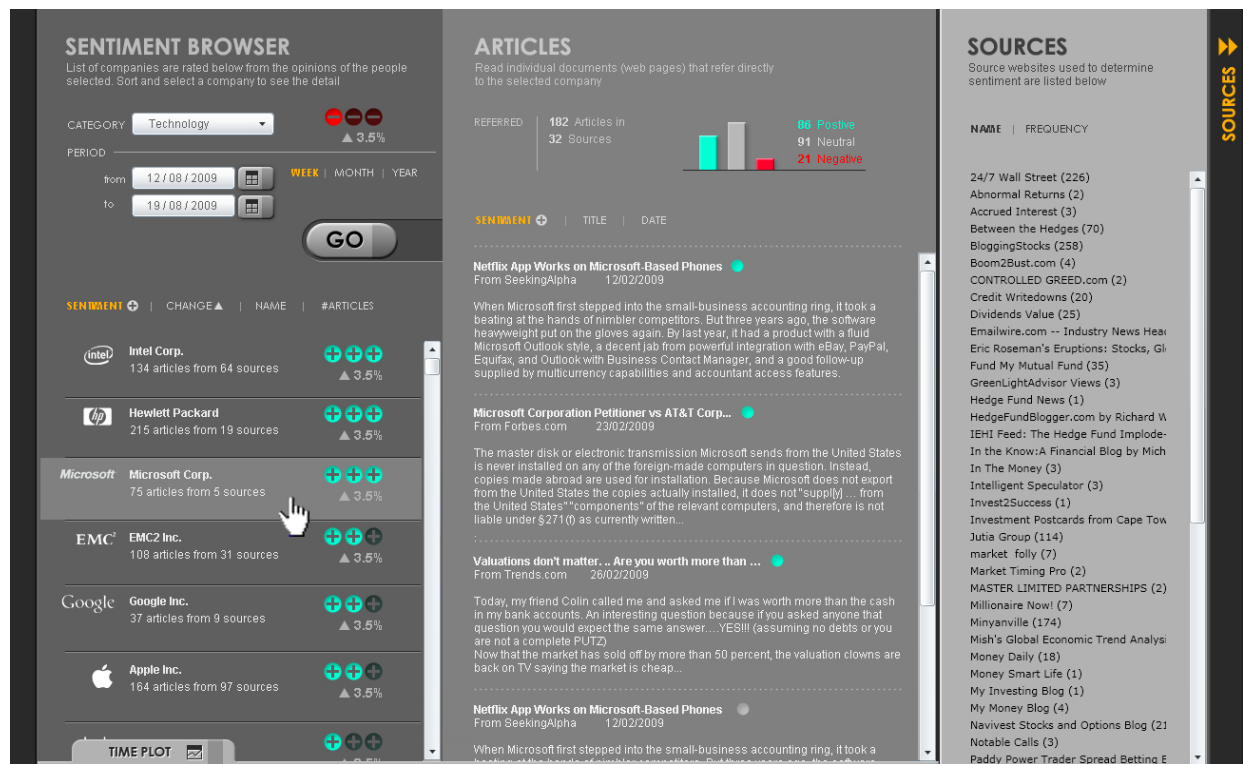


Figure 2. Putting together - stacking Unit Representation as a result of searching

browsing for more detail on the entries in the results.

Opinion Changes Over Time

As changes of opinion over time can be effectively presented and easily understood with a conventional time-based graph, such a graph can be incorporated into our overall design simply as a panel at the bottom of the screen which can slide up or down as the user wishes. If a user wants to view the temporal changes of an opinion about a particular company, s/he selects a company on the search result panel and drags it down to the bottom of the screen where the Time Plot panel tab is located (see Figure 2). This will then slide up the Time Plot panel and present the selected company's profile change over the user-specified period (see Figure 3). The user can drag in more than one company into this slide-up panel to compare the profile changes of multiple companies.

In Figure 3 the user has dragged two companies (Intel Corp. and Hewlett Packard) into the Time Plot panel, and each is assigned a unique colour (orange and red). Bringing the mouse cursor on the entry on the left of the Time Plot panel highlights the line in the graph area on the right, with vertical dotted lines at each of the data analysis points indicating the variance of opinions at that point in time. In Figure 3, the Intel Corp. time plot shows that the opinions improved over the past 1 week with the variance of opinions decreasing (i.e. opinions converging) as the orange line and its vertical dotted lines indicate. At any time the user can click on the Time Plot panel heading to slide it down or up, trading off the area with the list of companies presented above. On a

very large computer screen we could facilitate a permanent area for Time Plot panel without sacrificing other information areas, but in the current implementation we assumed a computer monitor in a typical office, and thus we adopted such a slide-in and -out panel solution.

CONCLUSION

Facilitating interactivity in visualisation does not necessarily mean conventional graphs and charts augmented with a few animated or mouse-over effects. Our contribution in this paper is to explore and introduce a new interactive visualisation where an individual retrieval unit is visually defined then it is used as the unit of searching and browsing as if one searches Web pages, rather than attempting a more conventional high-density visualisation schemes often tried in the visualisation community. Thus the issue here is not so much on how much information density one screen can accommodate (e.g. how many companies the left panel can display) similar to the fact that a scalability is not an issue on most Web search engines' search result display where only top few entries are of concern to the user and he/she can easily sort or filter the order of entries. The system is fully implemented with its Web interface running on Silverlight. While we have had a series of informal user tests during the prototype development stage with the interface, we are now planning to conduct a more formal evaluation with the complete system in order to gain a better understanding of the ways this strategy can support a specific set of tasks in the financial domain.

ACKNOWLEDGMENT

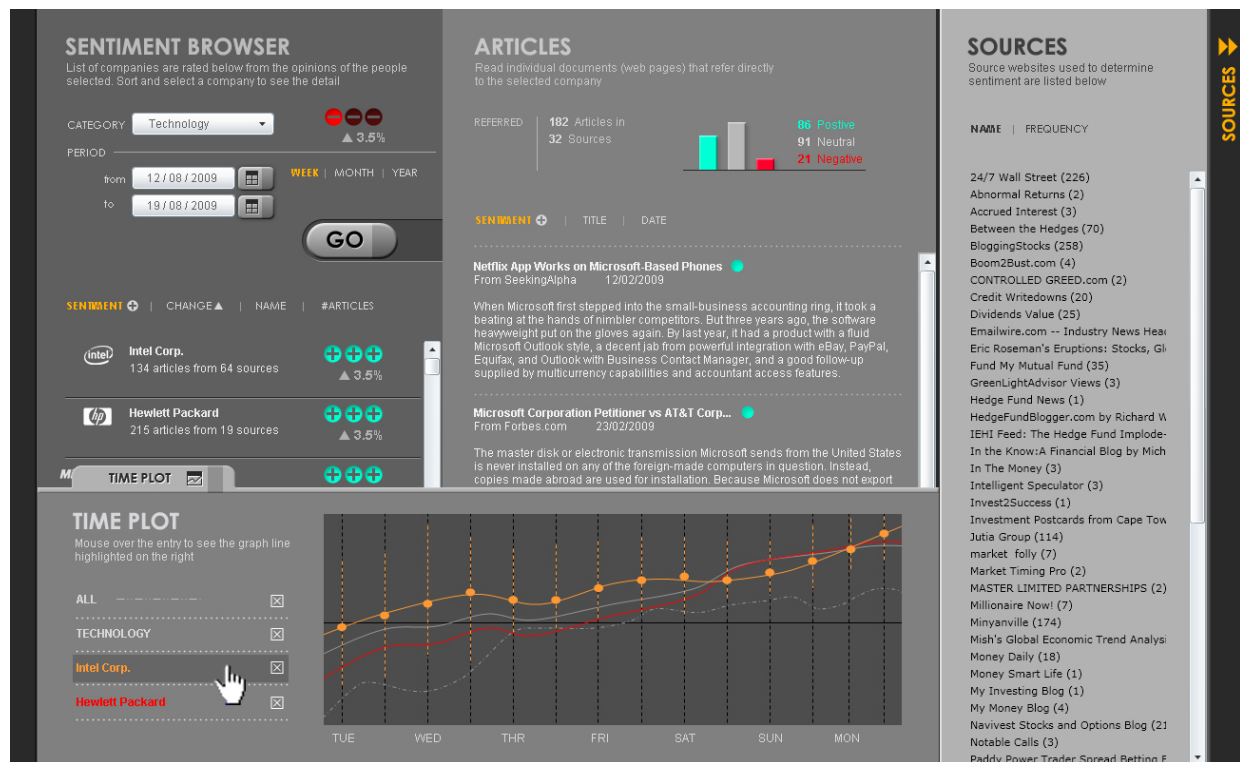


Figure 3. Time Plot - looking at the opinion changes over time and comparing those for multiple companies

This work is supported by Science Foundation Ireland under grant 07/CE/11147, and by Enterprise Ireland under grant IP/2008/0549.

REFERENCES

1. G. Carenini, R. T. Ng, and A. Pauls. Interactive multimedia summaries of evaluative text. In *IUI '06*, pages 124–131, New York, NY, USA, 2006. ACM.
2. T.-c. Fu, D. C. M. Sze, P. K. C. Leung, K.-y. Hung, and F.-I. Chung. Analysis and visualization of time series data from consumer-generated media and news archives. In *WI-IATW '07*, pages 259–262, 2007.
3. T. Fukuhara, H. Nakagawa, and T. Nishida. Understanding sentiment of people from news articles: temporal sentiment analysis of social events. In *ICWSM 2007*, 2007.
4. 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, Berlin/Heidelberg, 2005. Springer.
5. M. Gamon, S. Basu, D. Belenko, D. Fisher, and M. Hurst. Blews: Using blogs to provide context for news articles. In *ICWSM 2008*. Association for the Advancement of Artificial Intelligence, 2008.
6. 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 *SST '06*, pages 23–30, 2006.
7. J. Harris and S. Kamvar. We feel fine. an exploration of human emotion, in six movements. available at: <http://www.wefeelfine.org/>, 2009.
8. 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.
9. M. Hurst. Data mining: Mapping the blogosphere, from text minding, visualization and social media, 2009. <http://datamining.typepad.com/gallery/blog-map-gallery.html>.
10. Indratmo, J. Vassileva, and C. Gutwin. Exploring blog archives with interactive visualization. In *AVI '08*, pages 39–46, New York, NY, USA, 2008. ACM.
11. B. Liu, M. Hu, and J. Cheng. Opinion observer: analyzing and comparing opinions on the web. In *WWW '05*, pages 342–351, New York, NY, 2005.
12. N. O'Hare, M. Davy, A. Bermingham, P. Ferguson, P. Sheridan, C. Gurrin, and A. F. Smeaton. Topic-dependent sentiment analysis of financial blogs. In *TSA '09*, Nov 2009.
13. F. Wanner, C. Rohrdantz, F. Mansmann, D. Oelke, and D. Keim. Visual sentiment analysis of rss news feeds featuring the us presidential election in 2008. In *VISSW 2009*, February 2009.