Interactive Visualization of Emerging Topics in Multiple Social Media Streams

Adrian M.P. Braşoveanu, Alexander Hubmann-Haidvogel, Arno Scharl

MODUL University Vienna, Department of New Media Technology

Am Kahlenberg 1, 1190 Vienna, Austria

(t) +43-1-3203555-500; (f) +43-1-3203555-903

www.modul.ac.at/nmt

{adrian.brasoveanu, alexander.hubmann, arno.scharl}@modul.ac.at

ABSTRACT

This paper introduces an interactive *news flow* visualization that reveals emerging topics in dynamic digital content archives. The presented approach combines several visual metaphors and can be easily adapted to present multi-source social media datasets. In the context of this work, we discuss various methods for improving visual interfaces for accessing aggregated media representations. We combine falling blocks with bar graphs and arcs, but keep these elements clearly separated in different areas of the display. The arc metaphor is adapted and enriched with interactive controls to help users understand the dataset's underlining meaning. The paper describes the implementation of the prototype and discusses design issues with a particular emphasis on visual metaphors to highlight hidden relations in digital content. We conclude with a summary of the lessons learnt and the integration of the visualization component into the *Media Watch on Climate Change* (www.ecoresearch.net/climate), a public Web portal that aggregates environmental information from a variety of online sources including news media, blogs and other social media such as Twitter, YouTube and Facebook.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – interaction styles. I.3.6. [Computer Graphics]: Methodology and Techniques – interaction technique

General Terms

Design, Human Factors.

Keywords

Information visualization, news flow, thread arc, associated term, emerging topic, interface metaphor

1. INTRODUCTION

Social media allow analysts to interactively track the perceptions and activities of users during and after breaking news [1]. Because of the large number of status updates, tweets and videos,

visualizations often play an important part in understanding the flow of events. When designing new visualizations, more emphasis should be placed upon the emergence and decay of topics in tweets or status updates (which will be referred to as 'documents' in the remainder of this paper). Topical clusters and longitudinal patterns are often overlooked, either being lost between complex transitions, or not taken into consideration at all in the early design phase. Even the most aesthetically pleasing visualizations, for example, can suffer from these symptoms.

The underlying motivation and use case for this work is the *Media Watch on Climate Change* (www.ecoresearch.net/climate), a public web portal that aggregates environmental data from multiple sources including social media, environmental NGOs, popular science magazines, and corporate Web sites [2]. Emerging topics and author sentiment expressed towards these topics are calculated automatically and shown in interactive trend charts. Aiming to conceptualize and implement a more powerful metaphor, this paper introduces a topic-centered visualization method that is geared towards displaying search results or quickly changing document archives – e.g., tweets, YouTube videos or Facebook status updates. Keywords can be shown through classic bars (each bar represents the number of occurrences of a keyword from all social networks) or stacked bars (each bar represents the number of occurrences of a keyword in one social network).

One of the solutions to the problem of understanding the underlining semantics of the social media streams is to combine multiple visual metaphors into novel visualizations. To integrate multiple metadata dimensions into a single display, the method introduced in this paper deploys three visual metaphors, which will be discussed in the following related work section:

1. **Falling bar graphs**, similar to the popular stack visualization of the Digg social news service (www.digg.com).

2. **Stacked and color-coded sentiment bars.** These two classic visualization types [3] are used to represent keywords on the lower end of the falling bar graphs.

3. **Threaded arcs**. An adaptation of the thread arcs display, visually similar but lacking message nodes or threads as in the conversation visualizations [4;5].

Combining these metaphors allows identifying key topics, describing relations, and showing the evolution of social media coverage – the co-occurrence of terms in a document (falling bars) or in multiple documents (arcs), for example, or dynamic changes in the keyword distribution across various social media sources.

2. RELATED WORK

Only recently the specific ways in which visual metaphors shape the information that is presented to the user have received increased attention [6]. Various approaches have been developed for visualizing temporal data [7]. In the information visualization literature, well-known metaphors for presenting data with a pronounced temporal component are *ThemeRiver* [8;9] and the *Perspective Wall* [10]. While such temporal activity and temporal intensity views are suitable for discovering temporal patterns and temporal behavior of entities, they generally cannot express manifold relations and complex patterns in heterogeneous datasets.
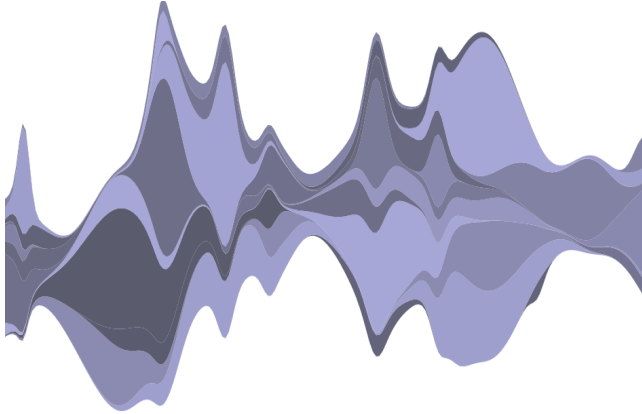

Figure 1. Open-source stream graph implementation [12], inspired by ThemeRiver [9] and Byron and Wattenberg's work on stacked graphs [13].

The *falling bar graphs* metaphor first appeared in a Flash visualization called the "Digg Stack" (www.digg.com). Squares falling from the top of the screen to the upper end of the bars typically represent news items read by users. Each news item will grow when hit by a new square. When visualizing single-source datasets, representing news items as falling squares is sufficient, but when dealing with multiple sources the same stories will be represented by different items of similar content. Therefore, we focused on keywords that summarize the major news stories.

Arc diagrams [4] and *thread arcs* [5] share similar visuals, but represent different situations. Arc diagrams show the connections between repetitive sequences of strings in different arrangements (text, music, DNA sequencing, etc.). Arc threads, by contrast, are similar to tree structures and convey the chronology of the messages they depict. Arc diagrams present arcs only above the message nodes, while arc threads show arcs both above and below the message nodes. In our visualization, we present arcs below the keywords, with the added advantage of being able to adapt the upper part of the visualization depending on the specific context.

Chan et al. [11] analyze the semantic structure of classical music using *layer threads*. Similar threads are clustered together and then intertwined with other groups to express the relations between themes. *Rainbow arcs* encode the information about layers which would typically not be presented with arc diagrams through color. A color represents a specific track, while a color band can accommodate the information from multiple tracks. *Bundled styles* offer modalities through which the repetition of the same theme is abstracted in a butterfly stitching way to not overload the graphic. It is a derivative of the pattern from knitting. *Theme fabric* is a type of layout that accommodates the relations between different themes. It uses multiple visual metaphors such as rainbow arcs or bundled styles to highlight semantic structures that can be extracted from musical compositions.

3. NEWS FLOW VISUALIZATION

The design and implementation of the *news flow* visualization followed six key principles:

1. **Chronology**. Show the order of the documents and of the most important 100 words that appeared in the last 30 days.

2. **Falling words**. Concurrently introduce terms that co-occur within a single document.

3. **Connected terms**. Provide an option to highlight relations between terms that co-occur within multiple documents.

4. **Interaction**. Start or stop the animation on different conditions, and display graphic elements or additional information when hovering over specific bars.

5. **Scalability**. The solution has to be computed in real time and support a large number of documents.

6. **Ease of integration**. Provide straightforward configuration of datasets and controls – e.g., changing speed, adjusting the dimensions of various components, or modifying their style.

Following these principles, the dynamic *news flow* visualization shown in Figure 2 is split into three distinct areas:

- **Falling Keyword Bars.** The upper part of the visualization uses the falling bars metaphor. The blocks that are falling from the top stand for the keywords extracted from a document. As they approach to the height of the column to which they belong, the words fade. The vertical size of the bars increases when they are "hit" by new terms.

- **Stacked and Color-Coded Bar Charts.** The middle part contains a classic bar chart for individual sources or specific attributes (e.g. average sentiment of a keyword), or a stacked bar chart for categorized multi-source data. The color scheme reflects either the average sentiment of the respective keywords or the familiar colors of the represented social networks. For sentiment data, for example, the color scheme is an interpolation between red and black for the negative values, and between green and black for positive values (Figure 2). The stacked bar chart (Figure 3) represents per-keyword frequency counts for different social media (currently Twitter, Facebook, and YouTube).

- **Temporal Controls and Semantic Associations.** The lower part of the visualization contains data about the temporal sequence (the start date and the current time), a list of the most recent documents, and mouse-over arcs that reflect semantic associations between the terms.

While each visual metaphor occupies its own area, the dynamics of the animation and user interactions can result in collisions – a word that falls into the middle area, for example, or a block growing into the upper area. During the development of the prototype, therefore, several restrictions have been added. Even if the visualization extends beyond the currently available screen space, all the important components (blocks, words, bars and arcs) have to remain visible in the three areas of the visualization. We show a limited number of arcs and we restrict the amount of information displayed in the document area to titles.
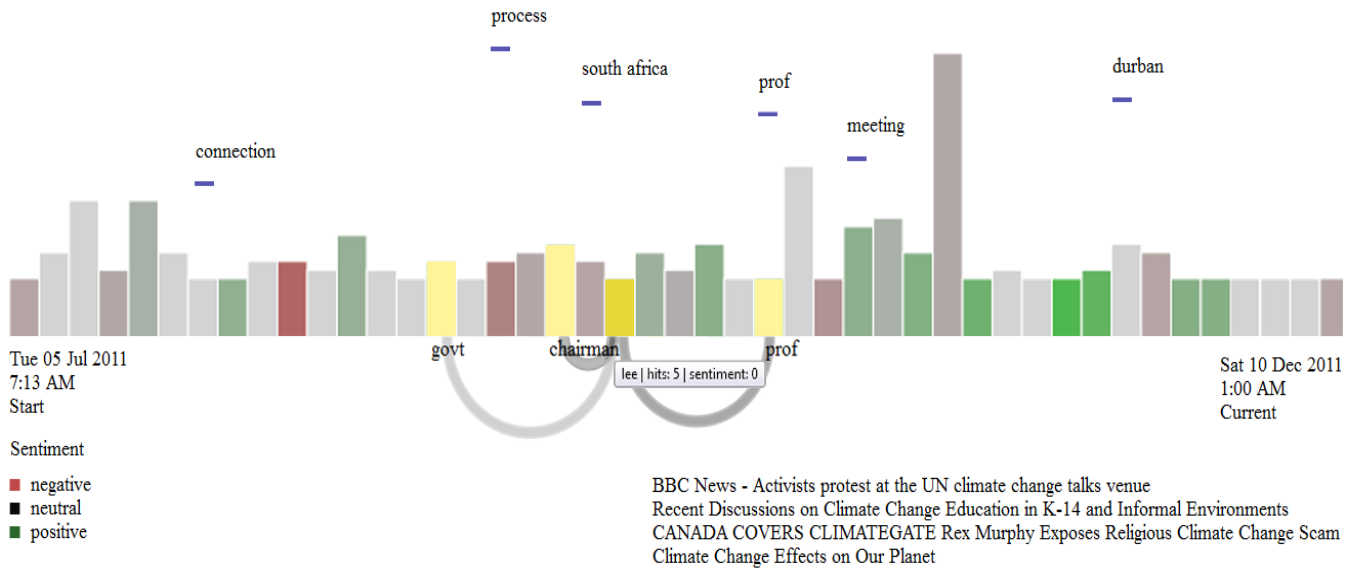
Figure 2. News flow visualization (color-coded sentiment)

Most visualizations use the arc metaphor adopting one of its popular variants such as arc diagrams and thread arcs. In the *news flow* visualization, the arcs do not connect message nodes but major keywords identified in the online coverage (represented by color-coded bars). The arcs are being displayed directly underneath the bars once users start interacting with the chart. Arcs are only shown for the selected bars, since connecting all related content elements would result in occlusion and a crowded visualization difficult to interpret.

4. INTERACTION MECHANISMS

The *news flow* visualization highlights connected terms, which are represented through falling words and through arcs that connect the bars which represent them. The falling words fade out as they get closer to the bar graph. On mouse-over, the animation is stopped and the connected terms are highlighted while simplified arc threads structures are being displayed underneath them.
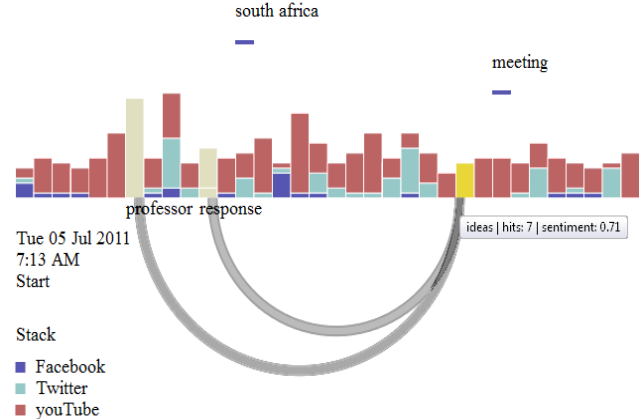

Figure 3. Stacked news flow visualization (color-coded source)

Options to run the visualization at various speeds, or to change the color scheme to highlight different attributed were added to provide a better user experience. To improve readability, we highlighted the arcs for a short period of time when keyword blocks from a document would hit the associated bars, slightly decreasing the speed to capture the user's attention. The *news flow* visualization makes use of color coding in three distinct ways:

- **Source coloring.** In the multi-source stacked bar display, the color indicates the respective source (e.g., blue for Facebook, red for YouTube, etc.) to show provenance information.

- **Sentiment coloring.** The color of a bar in the non-stacked version of the display reflects the average sentiment of the respective keyword.

- **Arc coloring (co-occurrence shading).** Lighter colors express lower co-occurrence values. This type of coloring gives us at a glance some information about strong or weak semantic relations between the highlighted co-occurring terms.

Optionally, source coloring and sentiment coloring are also applied to the falling keywords. Information about connectedness can also be encoded through different color schemes that involve shading. The basic color scheme involves lighter shades of gray. In the case of the stacked bar chart, the occurrence of terms in documents from the same social media source can be shown using variations of the social network's popular color for the shading of the arcs, for example.

5. IMPLEMENTATION

Among the key requirements of the *news flow* visualization was the use of technologies that run directly in modern browsers, without the need for additional plug-ins. To quickly prototype the desired visualization, we used the *Data-Driven Documents* (D3) JavaScript library [14]. As opposed to similar libraries, D3 is not focused on a new graphics grammar, but on the integration of existing standards to create visualizations. D3 uses HTML for presentation, CSS for styling, JavaScript for animation and SVG for vector drawing. By providing a scene graph and acting as a bridge between common standards, D3 has the potential to replace common JavaScript visualization frameworks such as *Processing* (www.processing.org) and *Raphaël* (www.raphaeljs.com).

In the context of the *news flow* visualization, the most important features offered by D3 were the rendering of fast transitions, the ability to debug the visualization within the browser, and the interaction capabilities built into the library.

Scrolling and zooming posed a number of challenges because of the lack of features like z-order from the SVG and D3 specifications. In many cases, only parts of the arcs would have been displayed on the screen if the number of blocks exceeded a certain threshold. Limiting the height of the arcs addressed this issue and makes sure that the arcs are displayed in their entirety. This resulted in adapted values for arc center and radius, and the intended usage of smaller arc dimensions.

When integrating the *news flow* visualization into the multiple coordinated view ensemble of the *Media Watch on Climate Change*, several features had to be added including methods to replace the datasets and color schemes on the fly (usually the colors represent the respective media brands), to set different sizes (at least a small widget size and a full screen size for the beginning), and to integrate with other controls such as time selection mechanisms. For visualizing search results, an Application Programming Interface (API) had to be provided that would deliver results using JSON (JavaScript Object Notation). In order to show multiple social media sources via stacked bars, we transformed the *news flow* visualization into a generic component and redesigned the title objects so that they are easier to understand for first-time users. Finally, we had to make sure that the interface elements are properly styled and that the new visualization could be integrated into the existing view synchronization framework [2] in a scalable manner.

6. FUTURE WORK
The *news flow* visualization introduced in this paper combines (i) falling keywords, (ii) sentiment-colored or stacked bar charts, and (iii) thread arcs to visualize the emergence and decay of topics in social media coverage. Hidden relations in the depicted datasets are revealed through the use of color or size. The JavaScript prototype was integrated into the *Media Watch on Climate Change* (www.ecoresearch.net/climate), a public news and social media aggregator on climate change and related environmental issues.

Future research will continue to experiment with new visual metaphors and labeling strategies, and apply them to datasets of various size, complexity, and origin – e.g. extracting common themes and displaying them instead of the individual document titles, and connecting those themes to the associated keywords through interactive layer threads.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Leskovec, J., Lackstrom, B., and Kleinberg, J. 2009. Meme-tracking and the Dynamics of the News Cycle. In *KDD '09. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. (Paris, France, June 28 – July 01, 2009). 497-506.

[2] Hubmann-Haidvogel, A., Scharl, A. und Weichselbraun, A. 2009. Multiple Coordinated Views for Searching and Navigating Web Content Repositories. In *Information Sciences,*179, 12 (2009), 1813-1821.

[3] Tufte, E. 2001. *The Visual Display of Quantitative Information*. 2nd Edition. Graphics Press, Cheshire, CT.

[4] Wattenberg, M. 2002. Arc diagrams: Visualizing structure in strings. In *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002*. (Boston, MA, USA, Oct 2002). 110-116.

[5] Kerr, B. Thread arcs. 2003. In *IEEE Symposium on Information Visualization, 2003. INFOVIS 2003*. (Seattle, WA, USA, Oct 2003). 211-218.

[6] Zienkiewics, C. and Kosara, R: 2008. The Shaping of Information by Visual Metaphors. In *IEEE Trans. Visualization & Comp. Graphics*, 14, 6 (Nov-Dec 2008), 1269-1276.

[7] Müller, W. and Schumann, H. 2003. Visualization Methods for Time-dependent Data - An Overview. In *Proceedings of the 2003 Winter Simulation Conference,* New Orleans, USA: IEEE Press. 737- 745.

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

[9] Imrich, P., Mueller, K., Imre, D., Zelenyuk, A. And Zhu, W. 2003. Interactive Poster: 3D ThemeRiver. In *IEEE Symposium on Information Visualization, 2003. INFOVIS 2003*. (Seattle, WA, USA, Oct 2003).

[10] Mackinlay, J.D., Robertson, G.G. and Card, S.K. 1991. The Perspective Wall: Detail and Context Smoothly Integrated. In *SIGCHI Conference on Human Factors in Computing Systems*. New Orleans, USA. 173-176.

[11] Chan, W.Y., Huamin, Q. and Mak, W.-H: 2010. Visualizing the Semantic Structure in Classical Music Works. In *IEEE Trans. Visualization & Comp. Graphics*, 16, 1 (Jan-Feb 2010), 161-173.

[12] Bostock, M. 2012. Streamgraph Algorithm, Colors, and Data Generation. http://mbostock.github.com/d3/ex/stream.html

[13] L. Byron and M. Wattenberg. Stacked graphs – geometry & aesthetics. *IEEE Trans. Vis. and Comp. Graphics*, 14(6) : 1245-1252, 200.

[14] Bostock, M., Ogievetsky,V. and Heer, J: 2011. D3: Data-driven documents. In *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 17, 12 (Dec 2011), 2301-2309.