

VISUALIZING MULTIPLE NEWS STORIES USING THE NEW GALACTIC AND CONCURRENT STREAM METAPHORS

Alena KOVAROVA

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
Alena.Kovarova@stuba.sk*

Abstract. We propose two new ways of visualization, both employing an exponential timeline. The first one uses concurrent streams metaphor and the second one galactic streams metaphor. Both are highly suitable for visualizing multiple developing stories represented by sets of articles ordered in time. Our visualizations are designed to suit readers of news portals. They can help them quickly find the most wanted article according to their opinion.

1. News visualization

There are several different approaches to news article visualization. They differ mainly because of their different purpose. Since we have focused on a common end user – a reader of daily news, our purpose of visualization is to offer the user the latest news in the context of previous ones, i.e. to embed them in a particular story [1, 2]. The key here is to see, which article is the most recent (e.g. from the last time the user saw it). From the research literature on article visualization, the following two ones have been most inspiring for our research:

Rennison [3] presented an approach that creates edges between articles and categorizes them into several levels (each category is represented by a keyword). Rennison then visualized the articles using progressive zoom and thus created the information pyramid. It is interactive and allows user to zoom to different levels. The author named the approach Galaxy of News.

Havre et al. [4] have presented ThemeRiverTM which uses the river metaphor. In this approach to visualization the x-axis denotes time in linear scale and the y axis denotes the number of documents relevant to themes. Unlike the previous one, this visualization re-

flects quantity of articles. However, to keep all themes in one “river flow” can be confusing, since recipient can obtain false sense of quantity. Such representation seems to be more appropriate for professionals.

2. Abstract model of news stories visualization

Some visualizations do not exhibit appropriately the most important aspect of news, i.e. time dependency. Our abstract visualization model (story map) is oriented on a common reader of daily news. There is few typical use cases associated with the visit of such user on a daily news portal. These use cases can be prioritized and formed as questions:

- 1) What’s new?
- 2) Is it interesting to me? (Does it belong to the story I’m interested in?)
- 3) Are there any other articles related to the one I’m currently reading? Did I read them already?
- 4) How long is the news story? Can I view it all at once? Can I zoom in on a shorter period of the story?

Other questions are also viable, but for us these are of the highest priority and thus our visualization model covers them in the first place. Our model can be enriched by other characteristics to cover additional needs of the user. When necessary to emphasize what is new, we can do it in different ways (by using size, colour, etc.), but more often the time dimension is visualized using a timeline oriented from left to right (as with ThemeRiver™ [4] and many others). Such timeline forms the base of our visualization model as well. We use the fact that the forgetting as well as an attention curve has exponential character [5]. Because of this it, we have decided to adopt an exponential timeline. Therefore the most recent articles get the most space, which gives them visual emphasis. On the other hand the time information is less obvious for the oldest articles. Inspired by the work of Renison [3], we address this problem by adding the possibility to zoom in on less recent articles, allowing for better resolution on time information. Further details are described in our paper [1].

3. Galactic and Concurrent Stream Metaphors

There are multiple visualizations that meet the four criteria of the defined abstract model, however not all of them are equally user friendly. We describe here two visualizations that have been gradually developed and can be considered as user friendly. The first one is based upon traditional visualizations, such as ThemeRiver™ [4]. We chose depiction of the stories as streams that flow from left to right, following the timeline, although with exponential scale in our case. As we consider the ability to distinguish between stories to be crucial with our model, we created multiple concurrent streams (see Figure 1), instead of a single multi-coloured stream. This visualization exhibits all the characteristics outlined in previous section, some of which show upon mouse events (more details in our paper [1]). Time slots in Figure 1 amount to one day.

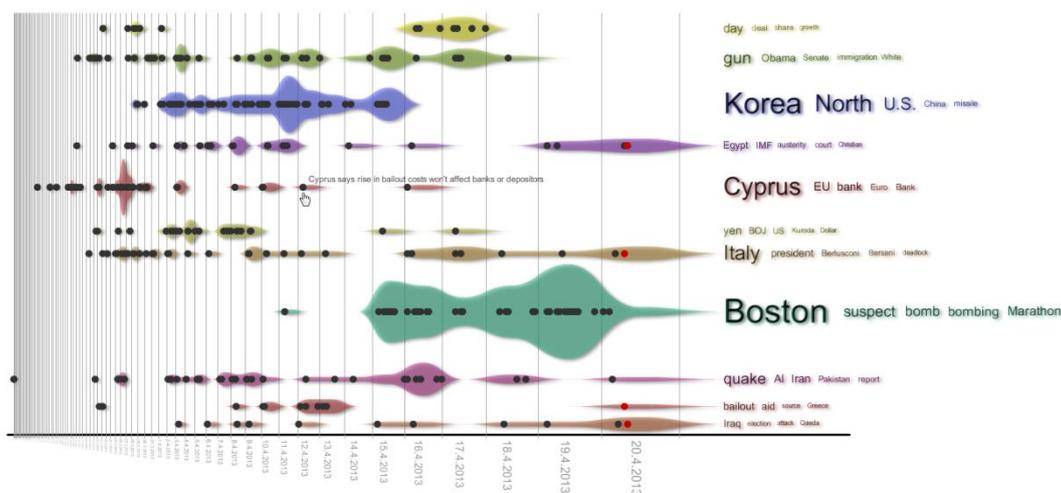


Figure 1. Visualization of news stories using the metaphor of *concurrent streams* [1].

The second visualization approach arranges the story streams in a circular pattern, where the time flows from the edge to the center of the circle. Streams are also separated and coloured using random colours. This visualization is also interactive the same way as the former one. We call this visualization mode the *galactic streams*, as the radial shape, when zooming, reminds us of space travel depiction, with stories representing galaxies. The approach is elaborated in more detail in our paper [1].

4. Visualization of an alternative dataset

Our news data were taken from Reuters news portal. In order to test the versatility of our visualization approaches, we used them to visualize data from an alternative source, i.e. the website of BBC news. There are notable differences, which may be attributed to a different structure of both portals, but most notably to a different pattern of connections between the articles.

5. Discussion

The number of articles, their density and the final shape of areas representing news stories may vary depending on the dataset being visualized. Default visualization settings may yield either too dense or too sparse clusters of articles. Therefore it would be desirable to calculate the optimal number of stories per each time slot and adjust the length of time slots accordingly. If necessary, visual density of the articles in the past may be adjusted as well. However these are currently unexplored attributes of visualization that may enrich the model in the future. When considering the original four requirements outlined in the second section, both concurrent and galactic streams meet them. But despite the common model, they are still two different visualizations.

6. Future work

We intend to explore other visualization attributes that we could incorporate into our abstract model. First we will focus on automatic sorting of the stories based on their topic proximity. We would also like to develop a method for automatic adjustment of the time step, which is currently fixed to one day. This would allow us to adjust the number of articles fitting into a given space so that they are never too thick, or too spaced out. The time step would be also explicitly configurable by the user. Another interesting option would be to allow for a possible modification of a view in a way similar to a fish-eye effect. It would enable the user to zoom in on the past articles, dynamically adjusting the resolution only on a single portion of the scale, without having to zoom in on the whole scale and thus hiding the most recent articles.

Acknowledgement: This work was partially supported by the Slovak Research and Development Agency under contract APVV-0208-10 and by the Scientific Grant Agency of Slovak Republic under grants No. VG1/0671/11 and VG 1/0752/14.

References

to other papers publishing the results that are summarized here

- [1] Sabo, S., Kovarova, A., Navrat, P.: Multiple developing news stories identified and tracked by social insects and visualized using the new galactic streams and concurrent streams metaphors. *International Journal of Hybrid Intelligent Systems*, [to appear].
- [2] Navrat, P., Sabo, S.: Beehive Based Machine to Give Snapshot of the Ongoing Stories on the Web. *Transactions on Computational Science XXI*, Springer, LNCS 8160, (2013), pp. 296-314

Other references

- [3] Rennison, E.: Galaxy of news: an approach to visualizing and understanding expansive news landscapes. In: *Proceedings of the 7th annual ACM symposium on User interface software and technology (UIST '94)*. New York, NY, USA: ACM, (1994), pp. 3–12.
- [4] Havre, S., Hetzler, B., Nowell, L.: ThemeRiver: visualizing theme changes over time. In: *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, IEEE, (2000), pp. 115–123.
- [5] Wu, F., Huberman, B. A.: Novelty and collective attention. In: *Proceedings of the National Academy of Sciences*, (2007), vol. 104, no. 45, pp. 17 599–17 601.