

Interactive Exploration of Student Generated Content presented in Blogs

Iilir Jusufi¹, Marcelo Milrad¹ and Xurxo Legaspi¹

¹ Linnaeus University, Department of Media Technology, Växjö, Sweden

Abstract

Nowadays blogs are regarded as tools for communication as well as an important source for spreading information in almost every subject. In recent years, school teachers have started to take advantage of this technology in order to support their educational practices. In this paper we focus on the data generated by a project involving more than 50 Swedish schools where teachers and pupils are posting content related to their astronomy class activities in their blogs with the aims of improving the teaching process. The challenge here is to find suitable methods to explore all these blogs in an interactive and discovery fashion. Our proposed solution to this challenge is to provide a visual and interactive tool for the exploration of blog corpora by teachers, pupils, project managers and parents.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Computer Graphics]: User Interfaces—Graphical User Interfaces (GUI)

1. Introduction

The visualization of text documents is an ongoing challenge that continues to increase in popularity as the amount of text data is getting larger, in parts due to the widespread of social media content. Kucher et al. [KK15] present an extensive survey on the most recent literature review related to text visualization. In our current efforts, we are primarily interested in the visual exploration of the content of different text corpora; more precisely explore different blog entries. There are several related efforts presented for this purpose [PSJD15, SNR*14, AG16]. However, most of them do not offer specific metadata such as average length of a post within a blog, or are complicated to use without special training.

In this paper, we present a prototype visualization tool developed as a part of "Reaching for the Stars" project where more than 50 schools in the Kronoberg region in Sweden have competed in various challenges regarding the subject of physics, natural sciences and astronomy. Pupils and teachers described their work and activities about the challenges presented to them using school blogs allowing pupils, teachers and other parents to follow their work. The winners of the different challenges are decided based on the content and quality of the blog reports.

2. The Blog Visualization

There are two main requirements for our prototype as defined by the project managers and teachers during our requirement analysis session. The first one is to aid the selection of the most successful blog reports coming from the schools based on their reported activities for each one of the given challenges. The second one is to help teachers, children and their parents to navigate the blogs in order

to learn what other schools are doing, thus improving the overall quality of the activities in class. With these tasks in mind we have designed a prototype that addresses these requirements.

We were provided with links to all relevant blogs from the schools participating in the project. A simple crawler has been built to collect keywords and their frequency for each one of the posts. They are thereafter stored in a database. By keyword *frequency* we mean the number of the occurrences a specific keyword has in the given selection, i.e., frequency in all posts from all blogs or selected number of posts. This specific information is then visualized using the D3js [BOH11] library as shown in the Figure 1. Our prototype's GUI is divided into four main panels.

The Toolbar (Figure 1 (A)) features: a search box, a drop-down box to filter out blogs based on a specific metadata, and the frequency slider which will be described later. The Blogs View (Figure 1 (B)) shows a bar-chart of all the blogs ordered by the value of the specific metadata that can be selected from the dropdown control in the Toolbar. The users are able to select the following metadata: *number of posts*, *average post length*, *number of photos*, and *number of videos*. This particular view is targeted at project managers to help them identify active bloggers quickly. There are many teachers who publish a lot of posts with short messages, meaning they are lacking content and should not be considered for the award. Moreover, there are also teachers that post a lot of images and videos but little text. Therefore, there should be a way to identify these cases as well. Project managers appreciated having a quick and simple way to filter active posters. It should be noticed that this panel is helpful for teachers and pupils as well in order to

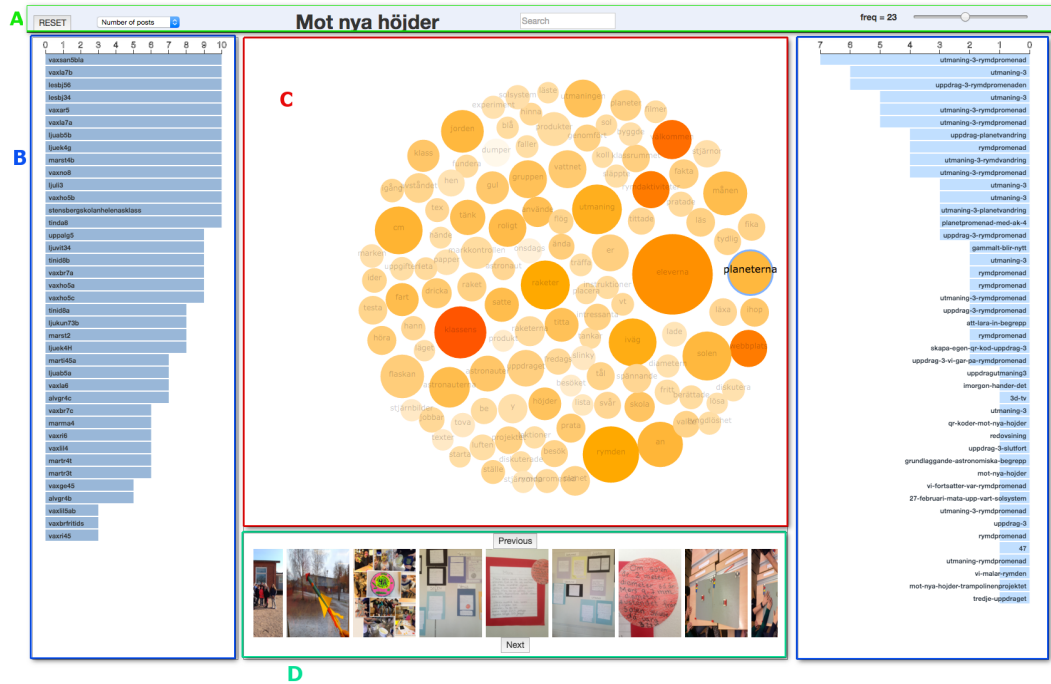


Figure 1: An overview of the GUI of our tool. The Toolbar panel is labeled with letter A. The Blogs View (B), shows the list of blogs using bar-chart metaphor. The Main View (C) visualizes the overall keywords. The size of the circle represents the relative frequency of the given keyword, while the color saturation represents the keyword popularity. The Posts View (E) shows the list of posts that contain the selected keywords visualized using a bar-chart metaphore. The Images View (D) lists the image thumbnails of the given selection.

help them identify blogs with the most relevant content and avoid browsing uninteresting posts.

As all the blogs are tackling a very well defined topic (astronomy lectures and challenges), we have decided to visualize the keyword frequency to aid the navigation and filtering of the underlying data. We used bubble clouds as seen in our Main View (Figure 1 (C)) instead of traditional tag clouds because longer words in tag clouds might appear as more frequent than shorter ones. The sizes of the circles represents the frequency of the given keyword based on the current posts selection. This metaphor allows for a quick identification of the most used keywords. The color saturation represents the *popularity* of the given keyword, i.e., the higher the number of posts the specific keyword is found the higher its popularity is. This makes it easy to identify popular keywords. The default view shows the bubble cloud of the entire corpora. The frequency slider in the Toolbar shows the value 23 (Figure 1 (A)) meaning all keywords with the frequency below this value are hidden. The user can change this threshold as desired and the Main View will reflect the changes accordingly. When the user selects one of the blogs from the Blogs View, the bubble cloud will change to show only the keywords found within the selected blog. The frequency slider will change its value to one, as the keyword frequency within one blog can be very low and user might be given a wrong impression that there are no keywords within a given selection (the user is notified about this change via a pop-up). Alternatively, a user can select one or more keywords from the bubble cloud and the Blogs View will filter out the blogs that do not contain the given keyword. This interaction will also display a bar-chart of posts containing the selected

keywords in the (Figure 1 (E)) ordered by the frequency of the selected keywords. When users are hovering the Posts View bars, the appropriate blog it belongs to will be highlighted. The reverse interaction is also possible when the user hovers the blogs. Users can also click on one of the posts triggering the bubble cloud to show only the tags that appear in the selected post. This feature provides the user with new insights about the topics covered within a given post or blog. A double click on the post or a blog in the corresponding panel will open the appropriate web page in the browser allowing the user to navigate to the desired post or blog quickly.

The outcomes of the work presented here shows how we tackled the identified requirements based on our informal usability sessions with the project managers and teachers. They found the proposed tool to be very interesting, as it allows them to gain new insights and quick ways to identify relevant content. The visualization and interaction approaches presented here could be easily used for visual exploration of other blogs as well. We are currently exploring the options for refining and improving the crawler mechanism and apply this approach on blogs containing posts about other domains. Temporal filtering is a feature that was requested by some of the teachers and project leaders and this will be part of our future work. We are also exploring options for adding some text mining capabilities that will equip our prototype with more analytic power. There have been some recent efforts in using information visualization in the field of technology enhanced learning (TEL) [SNR* 14], but most of the approaches have been targeting senior students. Here, we have presented a novel approach in the context of TEL that addresses younger students, teachers and family members.

References

- [AG16] ALEXANDER E., GLEICHER M.: Task-driven comparison of topic models. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan 2016), 320–329. doi:10.1109/TVCG.2015.2467618. 1
- [BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2301–2309. URL: <http://dx.doi.org/10.1109/TVCG.2011.185>, doi:10.1109/TVCG.2011.185. 1
- [KK15] KUCHER K., KERREN A.: Text visualization techniques: Taxonomy, visual survey, and community insights. In *Visualization Symposium (PacificVis), 2015 IEEE Pacific* (April 2015), pp. 117–121. doi:10.1109/PACIFICVIS.2015.7156366. 1
- [PSJD15] PETER J., SZIGETI S., JOFRE A., DIAMOND S.: Topicks: Visualizing complex topic models for user comprehension. In *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on* (Oct 2015), pp. 207–208. doi:10.1109/VAST.2015.7347681. 1
- [SNR*14] SCHEFFEL M., NIEMANN K., ROJAS S. L., DRACHSLER H., SPECHT M.: Spiral me to the core: Getting a visual grasp on text corpora through clusters and keywords. In *Proceedings of the Workshops at the LAK 2014 Conference co-located with 4th International Conference on Learning Analytics and Knowledge (LAK 2014), Indianapolis, Indiana, USA, March 24-28, 2014*. (2014), Yacef K., Drachsler H., (Eds.), vol. 1137 of *CEUR Workshop Proceedings*, CEUR-WS.org. URL: http://ceur-ws.org/Vol-1137/lakdatachallenge2014_submission_3.pdf. 1, 2