VISUALIZATION OF SOCIAL MEDIA CONTENT

Jaroslav Ráček
IBA CZ, s.r.o., jaroslav.racek@ibacz.eu

Tereza Pařilová
Fakulta informatiky, Masarykova univerzita, 146314@mail.muni.cz

Dalibor Toth
Fakulta informatiky, Masarykova univerzita, xtoth2@fi.muni.cz

ABSTRACT:
The paper describes possible ways of web content discussion visualizations. It focuses in particular on ways to show relationships between behavior and number of users, topic discussion and sentiment over time. The development of analytical and visualization portlets for Liferay Enterprise Portal is also mentioned.

KEYWORDS:
Web Discussions, Visualization, Enterprise Portals, Liferay, Data Analysis.

1 Introduction

21st century brings fast pace of life and its logically associated rapid development of technology. Some technologies make life in social society easier, others more complicated. The product-oriented market became the market for services, which also brings a new perspective on competitiveness. The marketing came into prominence, but it was often understood incorrectly or incompletely. Social relations between people were significantly shifted to the level of relationships on social networks.

The term "social media" is quite misleading. In relation to the social function of evoking a sense of the social impact of media in general, unilaterally. To a certain level it is so. However it supports an idea that television or press can be also classified among social media. Relationship of media to society is not supposed to be one-sided. All today known as social media has to have one specific feature - interactivity. Social media is not responsible just for spreading information among society, but it should also use the activity of their customers.

Social media is based on digitally coded data transfer, Web 2.0 platform, priority of real-time communication and many-to-many communication model. It is aimed at large communication interaction and the development of social relationships. Creating a community networks of users and the virtual world enable to integrate people into groups. Virtuality figures here as something intangible that does not exist here and now, something you can't really live on the outside, but only inside of an individual - emotionally. Communities in virtual space are described as communication between groups with common interests and values [1].

Social networks are the most popular social media today. By Nicole B. Ellison, professor of the University of Michigan, the social network can be defined as a web service that allows individuals [2]:

- to create a public or semi-public profile within a limited system,

- to specify a list of other users with whom we want to be in a relationship,
- to see a list of individuals who interact with individuals from our list within that system.

Social networks are a typical representative of Web 2.0. Building block is the user himself. He, as mentioned, participates in content creation and also forms a network with other individuals, who ultimately create community. The market of social networks is huge, but almost all benefit will get the few platforms in the head with Facebook, which is confirmed by claims of Roger McNamee, technology investor, who put a lot of money into social networking [3].

2 Social media as a marketing tool

Social media, especially the already mentioned social networks and discussion forums are very powerful marketing tool completely, used most in the last few years and getting more and more popular. Success is primarily supported by huge mass of people who visit virtual spaces. The voice of the individual has tremendous power in this environment - draws others to action, creating a psychological effect. And also this is the kind of marketing, which is less expensive than conventional marketing campaigns and also provides the possibility to analyze the environment and position of the company on the current market. Marketing in social media, however, must be dealt with an experienced marketer. Needs analysis, human approach and correctly worded responses of marketers are important prerequisites for successful marketing in social media [4].

In discussion forums and also in discussions in social networks, there is formed a natural sequence of comments to posts and a discussion thread is also being created. This thread is sorted chronologically, thus creating a kind of time series to which individual comments are recorded. Comments may also react on the original post or on an existing comment. A discussion thread, therefore, arises gradually as a cluster of responsive comments and users [5]. These clusters are very valuable to marketing strategy. They use visible acquaintances between users, their common interests, sympathies and antipathies, and just on the basis of this information can then be targeted propagation.

Marketing departments typically select the largest group of users of social media discussing the chosen topic, and here it starts its strategy. Instead of flooding advertising to users which lurks in the email or banner on each web page, advertising gets through friends and it is sufficient to send such notice or advertisement was applied on just one thread, one channel. No spam.

The primary need of every company and marketer is to find out as much as possible about social networking and user forums, especially for these clusters and user comments. For effective analysis of customer satisfaction and it is necessary to find out:

- who is discussing,
- what is being discussed,
- how many contributions of users the discussion contains and for what time,
- what is it's sentiment and it's color.

In the following text are presented visualizations dedicated for marketing purposes, which are integrated by the creative team into an analytical portlets for an enterprise portal Liferay.

3 Visualization of time series

The most commonly used view of the discussions is currently visualization of their development over time using 2D graph where the x-axis represents time and y axis the observed quantity. It is increasingly moving away from the original point and bar graphs, which are currently being replaced by functions. One of the possible uses of this type of graph is shown in Figure 1. This is visualization of the number of discussion visitors showing the total number of visitors and the number of discussing visitors.
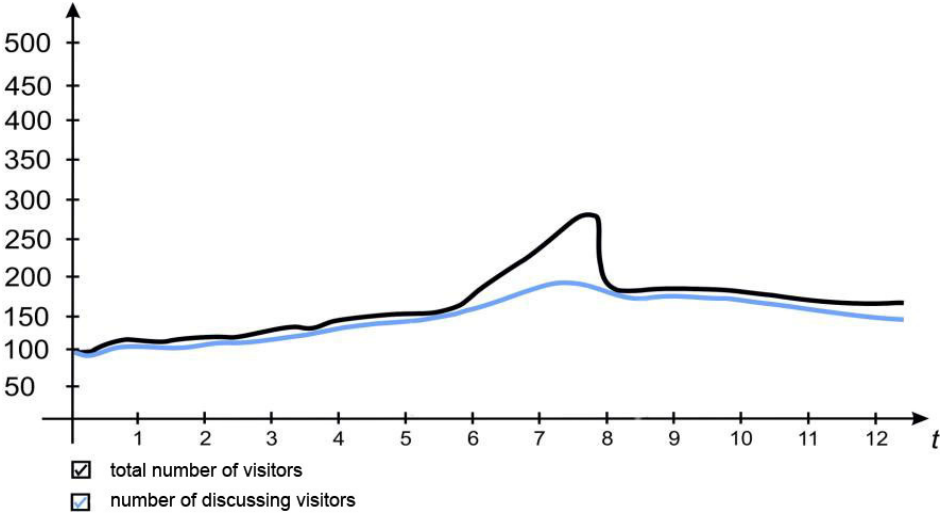
Figure 1: Visualization of the number of visitors at the time of discussion.

Another possibility is the use of visualization of time series variance of some variable which occurs above the posts. For example, the variable of how is the thread over time close, respectively isn't close to a particular topic, but in the similar way can be represented a sentiment in the discussion or amount of positively and negatively evaluated contributions.
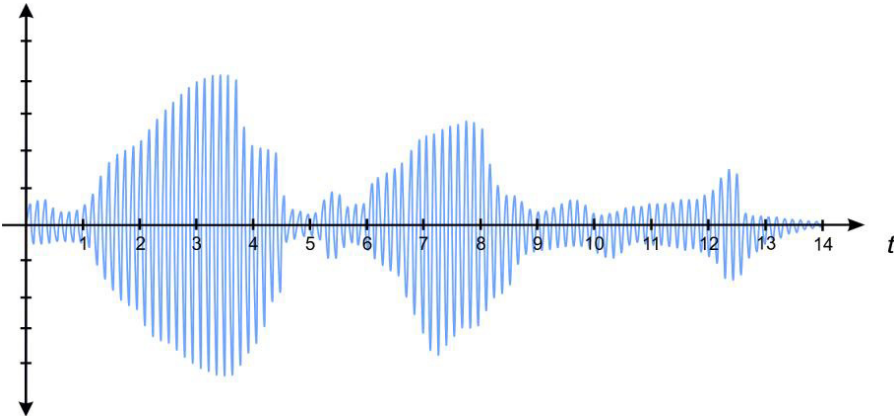

Figure 2: Visualization of the range of the discussion within the variable in time.

Figure 3 shows time series visualization of two different variables. In this case, it shows the number of visitors of the discussion and its sentiment. At the same time it can show interval when was the selected user present in the discussion, which is particularly important to identify so-called "strong players" who affect attendance, theme and sentiment.
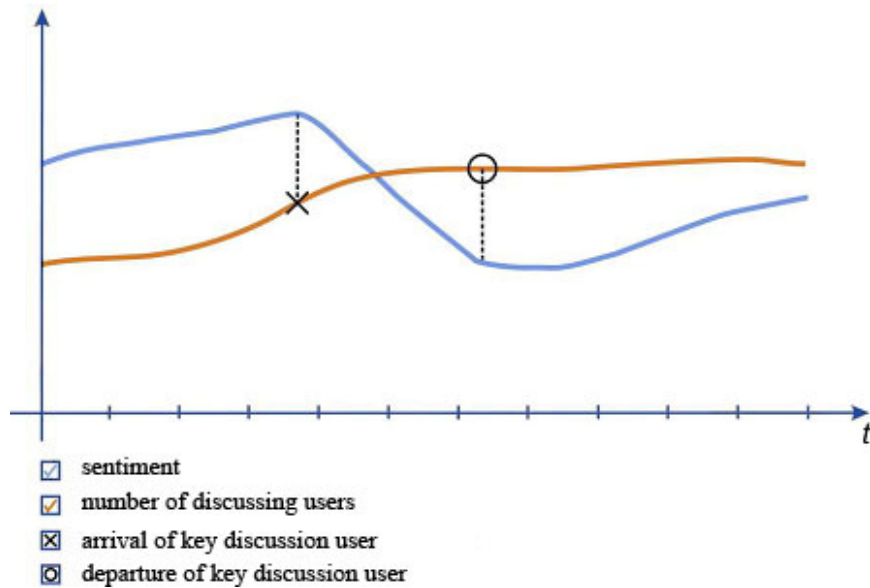
Figure 3: Illustration of arrival and departure of key user.

4 Branching visualization

A separate issue in the area of visualizing discussions is illustrating of branching in time. For these purposes are used "spectrographs". This is a representation of discussion branching, while individual branches or parts have a different color from the color spectrum. Different colors represent different values of monitored variables in the discussion. Typically, the color is used to represent the number of users, number of reactions on posts, proximity to the topic or sentiment. Time on this chart goes from left to right.
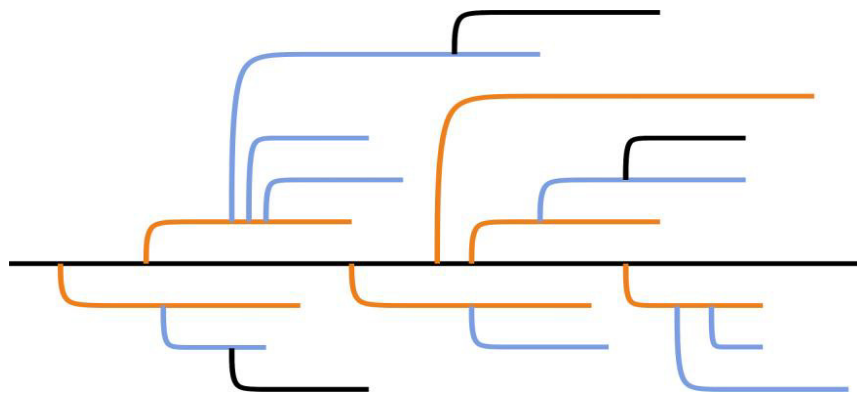

Figure 4: Spectrograph.

Similar capabilities as a meaningful spectrograph have a diagram in Figure 5. This is also an illustration of discussion branching. Time is represented by concentric circles ("rings"). Beginning of the discussion is in the middle, for example it can be an introductory article. Other contributions are represented by nodes. Unlike the spectrographs it can simultaneously affect multiple variables. For example the direction of thread in the direction of x and y axis can represent how the contributions relate to any two topics (each axis represents one topic). It is also possible to use color and size of knots.
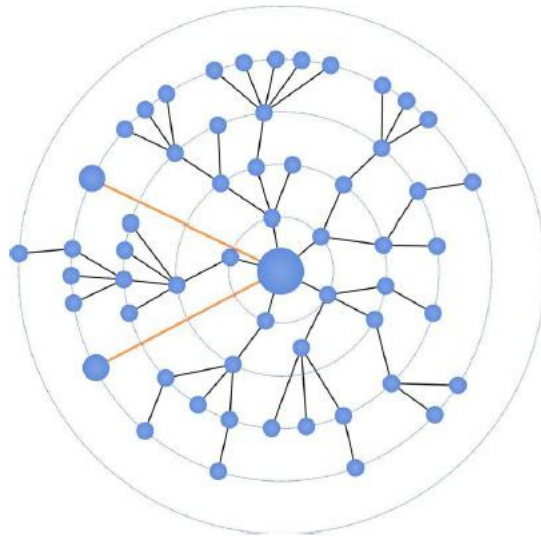
Figure 5: Visualization of branching and topic in time.

5 Topic visualization

Another visualization, which is used to illustrate the topic of the contributions, is use of the thematic n-gons. They don't capture time, but individual contributions are placed in the form of points in the regular polygons, where the particular sides represent individual topics of discussion. The closer to the edge the post is, the more it is related to a given topic. This access allows in a larger number of contributions and suitably chosen sequence of edges illustration of the thematic clusters of contributions.
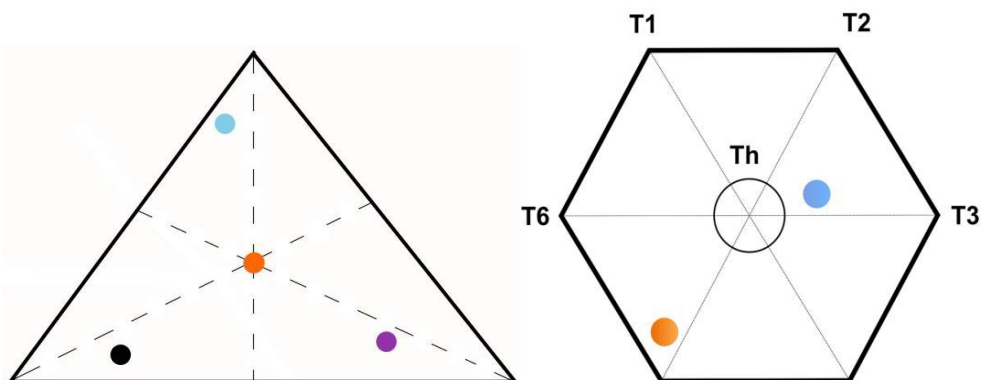

Figure 6: Visualization of the topic in the n-gon.

6 Analytical portlets

The industrial partnership of the Faculty of Informatics of Masaryk University and the IBA CZ is based in collaboration with other partners on development of new analytical and visualization tools in the field of unstructured data. These tools work primarily over the data coming from the Internet social networking, expert forums, FAQ documents and archives. [6]

As a carrier platform is used Liferay, which is an enterprise portal based on Java EE, which allows easy integration of information, applications and processes. Portal presents the information to its visitors through a web interface that can be adapted to the momentary needs of individual users.

The basic building blocks of a portal are Java portlets, which are web components designed to integrate Web applications and portals. Portlets are used as interchangeable components of

the user interface that provides the presentation layer. Interaction of portals and portlets is provided by the API.

Portlets are based on Java technology and in its nature are similar to servlets. Like servlets and portlets are managed by the container. Portlet receives and processes requests by which it changes its content.

Figure 7 shows a newly developed analytical portlet with a spectrograph that displays the sentiment in the discussion. The distance of contributions is represented by points (circles) is in proportion to the time period that elapsed between the posts. For clarity, the contributions are numbered according to order. The color spectrum in this case serves to visualize the sentiment. Thickness of the lines between nodes (contributions) is in this case attendance (number of views) discussion in a given time period.
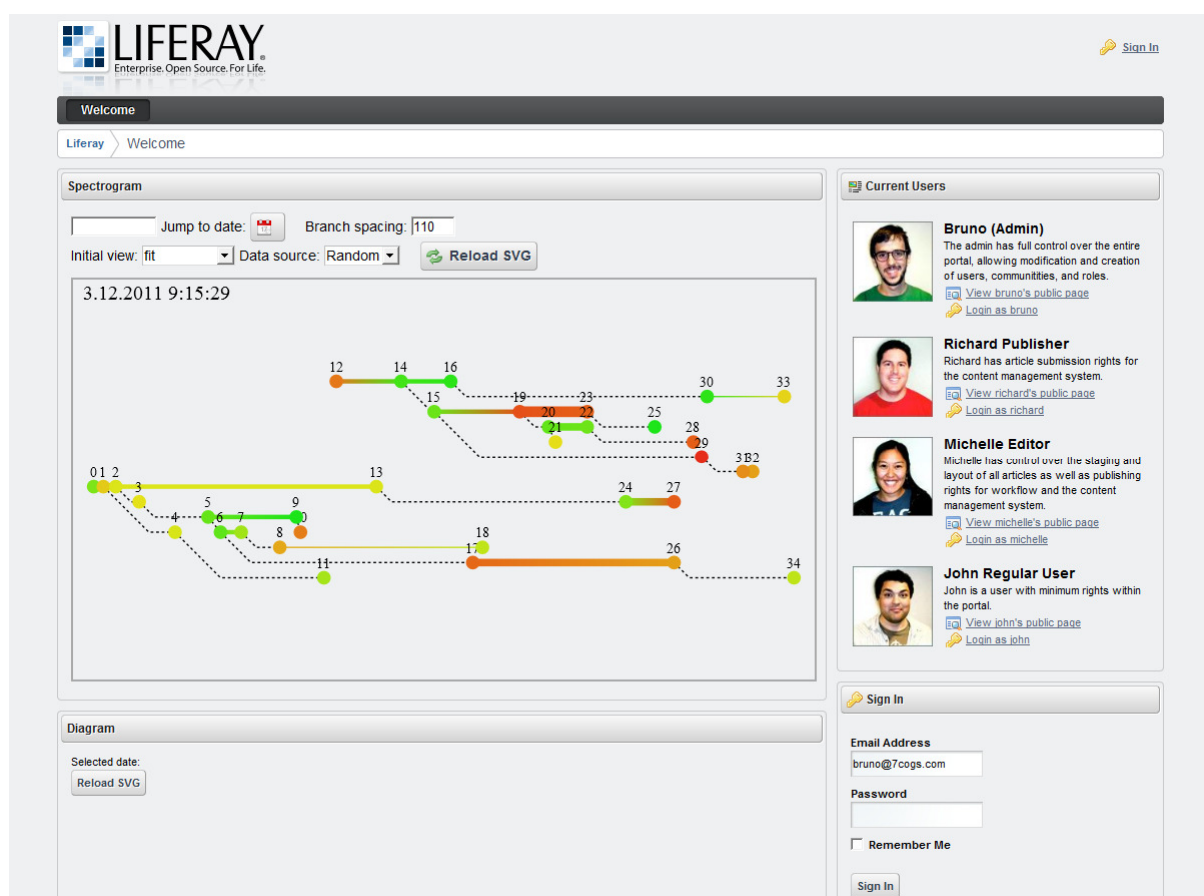

Figure 7: Liferay portlet with spectrograph.

7 Conclusion

Liferay platform proves to be suitable for implementing the above described analysis and visualization portlets. The system is in the final phase and portlets for attendance, topic and sentiment analysis are currently deployed in production environment.

Analytic functions were developed especially over the data of expert discussions of a technical nature (e.g. discussions about mobile technologies) and the data from the field of finance, in Czech, Slovak and English. In the future, there are plans to expand to other European languages including Cyrillic texts.

LITERATURE

[1]   PREECE, Jenny. Online communities: designing usability, supporting sociability. New York: John Wiley, 2000.

[2]   BOYD, Danah M.; ELLISON, Nicole B. Social Network Sites: Definition, History, and Scholarship . In: Journal of computer-mediated communication JCMC. Indiana University, 2007.

[3]   O'REILLY, Tim. Various Things I've Written . Tim O'Reilly.com. 2011.

[4]   EVANS, Dave. Social media marketing: an hour a day. Indianapolis, Ind.: Wiley, Serious skills, 2008.

[5]   JOHNSON, Luke. Social media help bosses tell their story. The financial times. London, 2011.

[6]   MINISTR, Jan; RÁČEK, Jaroslav; FIALA, Josef. Business intelligence jako nástroj hodnocení sentimentu nestrukturovaných textů sociálních sítí. IT pro praxi. VŠB-TUO, Ostrava, 2011.