SiCi Explorer: Situation Monitoring of Cities in Social Media Streaming Data

Andreas Weiler, Michael Grossniklaus, and Marc H. Scholl
Database and Information Systems Group, University of Konstanz
Box D188, 78457 Konstanz, Germany
firstname.lastname@uni-konstanz.de

ABSTRACT

The continuous growth of social networks and the active use of social media services result in massive amounts of user-generated data. More and more people worldwide report and distribute up-to-date information about almost any topic. Therefore, we argue that this kind of data is a good basis to observe ongoing situations in cities as well as related situations from outside about these cities in real-time.

This paper presents a visualization for monitoring the situation (current topics and emotions) in cities and about cities, which is reflected in the live message stream of the social microblogging service *Twitter*, by using continuously updating and with sentiment colored TagClouds.

1. INTRODUCTION AND MOTIVATION

The high volume and distribution speed of tweets makes it difficult for users to follow the evolution of topics within the continuous data flow. It is a big challenge to discriminate between normal behavior of the social sensor or unusual and abnormal behavior, which usually is an indicator for an interesting event in the area. However, the amount of useful information in the generated data increases as well. Another advantage of user-generated data is the automatic enrichment of the textual information by geographical information of the user's mobile device. Hereby, it is possible to classify the incoming information as local report or report about a city from outside.

In this paper, we present a visualization for monitoring situations in cities and surrounding areas, which is filtered by geographical coordinates, and the situation, which is reported from outside by filtering for keywords, in the live and continuous streaming data of Twitter. Our work presents a compact visualization for time series event data, which supports users to identify interesting data points inside the cities and about these cities from outside. This supports users in following the evolution of topics and emotions of locals in defined geographical areas and also to visualize the evolution of topics and emotions of people about the area. It also supports the users in differentiating between the situation inside the area and the situation from outside.

2. DESIGN

To visualize the evolution of topics and the emotion in and about a city over time, we use different layers of granularity displayed on a map. Each layer is split into two parts. The first part (inside the oval area, filtered by geocoordinates) reflects the situation inside the area of the city and the second part (inside the rectangle area without the oval area, filtered by keywords) reflects the situation from outside about the city. The TagClouds of both parts are adjusted to the shape of the area, by using a processing library[1], and so no overlapping terms between the two areas exists. The upper layers are divided into slices, which reflect the sentiment from the underlying layers. Hereby, it is possible to recognize interesting points in time in underlying layers even if the user just observes the top layer.

Top-Layer: The lowest granularity layer reflects the topics and emotions in hourly frames. To reflect the sentiment of the middle-layer both areas are split into six slices. These six slices are colored in the corresponding sentiment of the middle-layer.

Middle-Layer: The middle granularity layer reflects the situation in ten minute frames. Both areas are split into ten slices, which reflect the corresponding sentiment from 1 minute frames from the bottom-layer.

Bottom-Layer: The highest granularity layer reflects the topics and emotions in one minute frames in real-time.

The emotion of a time frame is visualized by using the color dimension. The fill color of the shape signifies the average sentiment (red = negative, green = positive) of the tweets in the data window. The value of the sentiment for a tweet is obtained by using an external library[2], which analyzes the text of the message and returns sentiment values between -5 (extremely negative) and 5 (extremely positive).

3. USE CASE

The use case describes the observation of the city *Boston* (filtered for 5 miles around the city center and the keyword "boston") on the day of the 15th April 2013. Figure 1 shows the overview of the whole use case for the hours from 5PM to 9PM. The most important term "marathon" in the first

[1] `http://wordcram.org/`
[2] `http://sentistrength.wlv.ac.uk/`

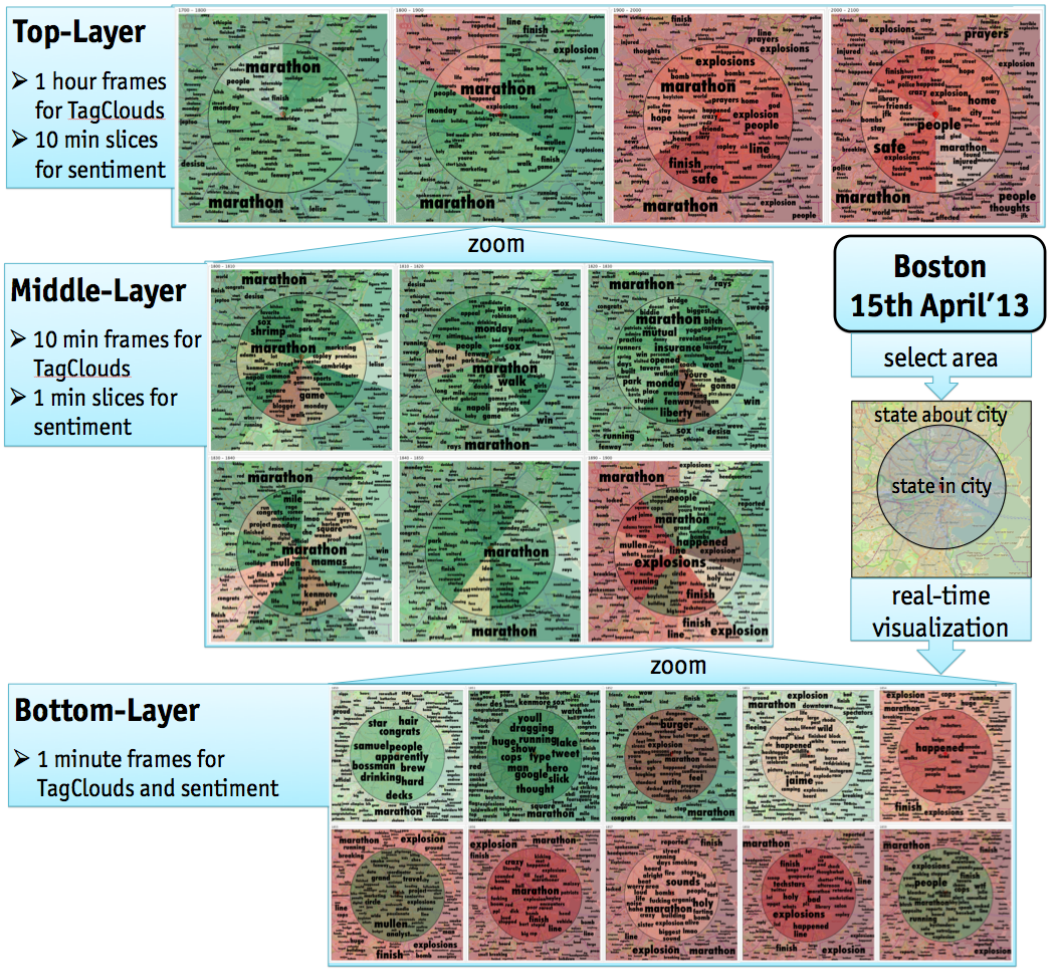

Figure 1: Process of the use case for the city Boston on 15th April 2013, from 5PM to 9PM (UTC).

hour signifies the ongoing sports event in the city and also reflects a very positive emotion for inside the city and from outside. Further indicators for the Marathon are terms like "running", "finish", "desisa" (name of the male winner), and "jeptoo" (name of the female winner). The second hour also shows very positive emotion for the first five segments of the hour. However, in the sixth segment, the last ten minutes of the hour, the emotion shifts to negative for inside the area and from the outside. Also, the term "explosion" can be seen in the outside and inside part. By looking at the two following hours, we notice that the negative emotion increases sharply in both areas. Further terms like "bombs", "prayers", and "thoughts" indicate that a bad event happened in the city of Boston. An interesting observation is that the terms "prayers" and "thoughts" are more frequent on the outside. Since the second hour shows this abnormal behavior and we are further interested in the evolution of the situation, we zoom into the second hour of the visualization. In this more detailed view, we can see that the term "marathon" is always very frequent, however in the last ten minutes of the hour the terms "explosion", "finish", and "line" on the outside and the terms "explosions", "happened", and "finish" on the inside reflect the just happened *Boston Marathon Bombings*. After analyzing the middle layer of the visualization, we are further interested in the situations in the last ten minutes of the second hour. Therefore we zoom into the last ten minutes of the middle-layer, where we can identify that the term "explosion" appears in both areas in the second minute. In contrast, the emotion is still positive at that time, because the event not yet widely spread. We can summarize that by using the visualization the explosion can be identified by terms only two minutes after the event took place. However, the emotion changes only slightly in the first minutes, but then the negative emotion increases and gives a good indication about the extent of the tragedy.

4. CONCLUSIONS

In this paper, we demonstrated a visualization for monitoring the situation (current topics and emotions) within and about cities, which is reflected in the live message stream of the social microblogging service Twitter. Our use case shows that it is possible to visualize the current and past topics and emotions for cities. We can also conclude that the live observation can support local people and news reporters in getting up-to-date information about the state of emotion and topics in and about a city.