ScatterScopes: Understanding Events in Real-time through Spatiotemporal Indication and Hierarchical Drilldown

VAST 2014 Mini Challenge 3 Recognition: “Honorable Mention for Good Support for Situation Awareness”

Dennis Thom, Michael Wörner, and Steffen Koch, Member, IEEE

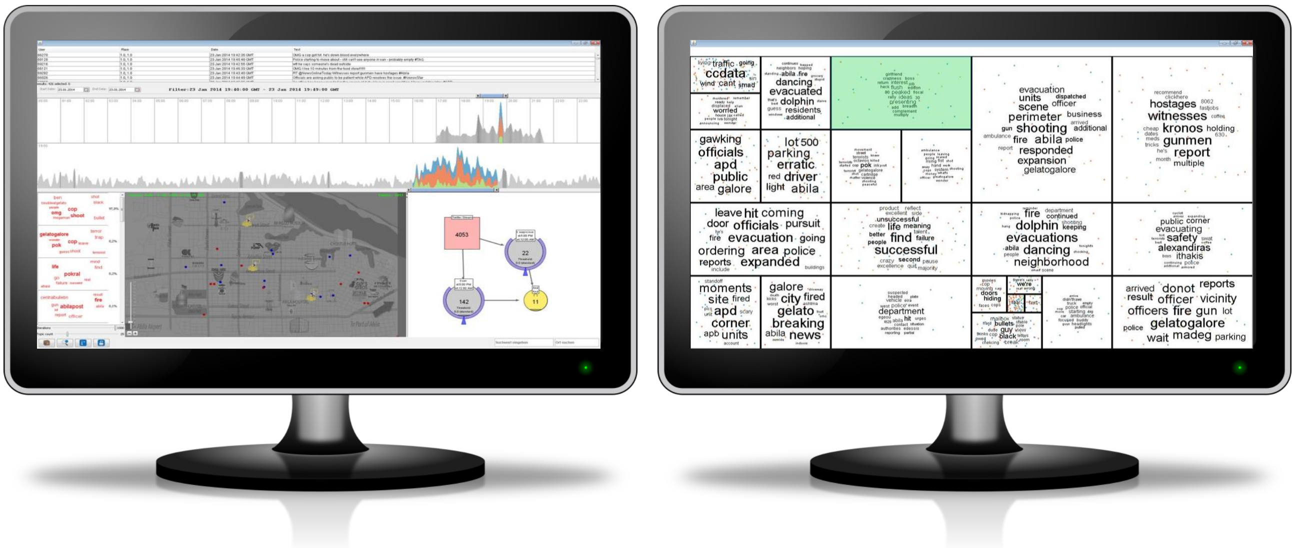

Fig. 1. *ScatterScopes* is designed to be employed in a two-display setup. The left display is used to detect relevant messages in the data stream using content search, spatiotemporal overview, sentiment analysis and selection management. The right display shows an interactive, explorable treemap of selected messages based on agglomerative cluster analysis.

Index Terms—User Interfaces, Information Search and Retrieval, Social Media

1 Introduction

The VAST Challenge 2014 MC3 featured a dataset of roughly 4000 microblog messages and 200 emergency callcenter reports produced in the fictitious city of Abila. The task was to identify relevant events and outliers in the data and highlight observations that are connected to an earlier kidnapping of employees from a company called GASTECH. In contrast to earlier challenges, the focus was on real-time data processing. The microblog messages were hosted on a web-server that simulated real-time streaming of the data during a 4.5 hour period and the challenge participants had to monitor and analyze them as they were transmitted. To tackle the challenge we developed *ScatterScopes*, a real-time enabled visual analytics system that fosters understanding of ongoing events by means of spatiotemporal overview and hierarchical drilldown. Trends and anomalies in space, time and content can be quickly identified with the system using interactive maps, sentiment timelines and textual search. Once interesting or suspicious subsets of elements are selected, their inherent topic structure can be further dissected based on a highly interactive treemap of message clusters. Subset selection and recombination is furthermore supported by a filter-and-flow mechanism that can also be used to formulate and test hypotheses based on Boolean logic. *ScatterScopes* has been used by our team to successfully identify and describe all events hidden in the MC3 data streams.

2 System Overview

In 2013 our team introduced *Scatterblogs2* [1], a visual monitoring and analysis system for Twitter that can already handle up to 20 million daily messages in real-time and thus seemed perfectly suited for the challenge. Scatterblogs established situational awareness based on geolocated tweets and employed SVM classifiers trained on large volumes of historic data to identify relevant information. The MC3 challenge data, however, featured only a relatively small portion of geo-tweets, and historic data for Abila is not available for supervised training. Both are conditions that resemble situations often encountered in real-world scenarios, rendering the challenge very relevant. To solve it, we re-combined several modules from *Scatterblogs2* for scalable stream data processing and spatiotemporal exploration and enhanced them with means to handle non-geolocated messages as well as to identify and examine anomalies based on unsupervised indication.

The resulting system, *ScatterScopes*, is designed to be employed in a two-monitor setup. The first display, which can be seen in Figure 2,

• *All authors are with the Institute for Visualization and Interactive Systems, University of Stuttgart. E-mail: firstname.lastname@vis.uni-stuttgart.de.*

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014; date of publication xx xxx 2014; date of current version xx xxx 2014. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

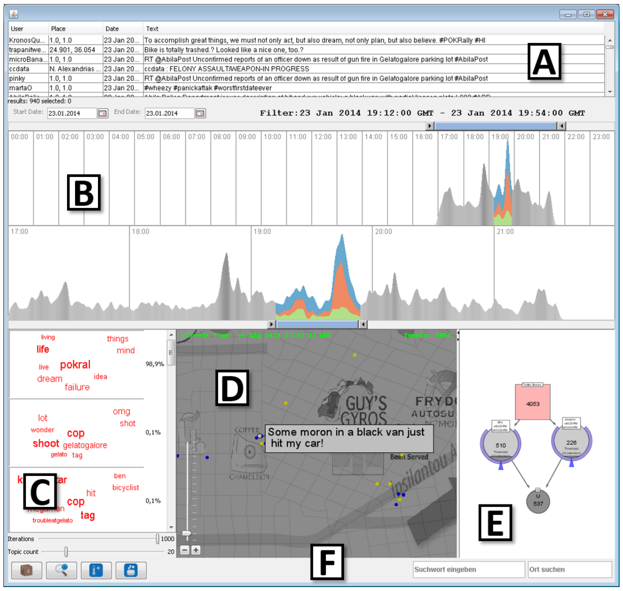

Fig. 2. First display: Space-time overview and selection management.

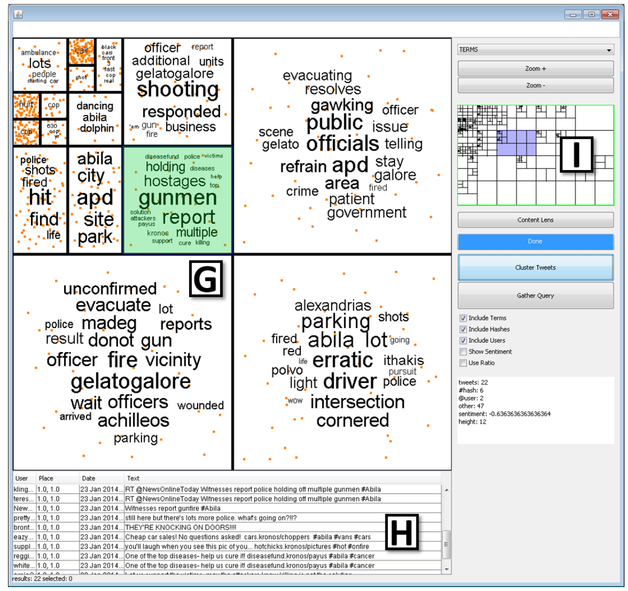

Fig. 3. Second display: Hierarchical treemap for topic drilldown.

provides an instant overview of streamed data and is composed of several linked components: A table (A) shows details of user-selected messages such as author id, timestamp, place and text. A temporal overview (B), generated with kernel-density estimation, reflects the overall volume of messages over time and indicates trends. For this purpose, the system performs a real-time sentiment analysis and uses colors to show the respective volumes of positive (=green), negative (=red) and neutral (=blue) messages. The temporal overview can furthermore be used for three-step temporal filtering. First, a broader time-frame is selected by defining start and end date using the two calendar widgets. Second, the broader frame can be interactively explored using the slider above the upper plot, which will immediately show the range in more detail in the lower plot and the corresponding messages on the map. Third, the slider below the lower plot can be used to select individual peaks and trends. Once a range of messages is selected by the user, an LDA model is used to produce a quick overview of their contents by showing a list of little tag clouds of the topic words (C). Prevalent topics appear higher in the list. The available geo-located messages and emergency reports are represented on the map (D), with red dots showing regular messages and blue dots showing emergency reports. If geo-location is not explicitly given, we try to find location-names in the text and meta-data and estimate possible locations by assigning them to crossroads. A filter-flow graph (E) can furthermore be used to store and combine filter configurations using Boolean operations. Finally, a control bar (F) provides means for textual filtering and allows to activate several map overlays including a movable term lens. If real-time data is currently received within the selected broader time frame, immediate feedback is given to the user as the plot height in the temporal overview increases on the right and new geo-enabled messages are highlighted through animations on the map.

The second display is shown in Figure 3. It provides a drilldown view of data based on an explorable topic tree generated on user request. Once the user has selected a subset of messages with temporal filters, textual filters or combinations of both, he or she can send them to the second display by clicking a button in the control panel. Agglomerative clustering is then applied to extract the underlying topic hierarchy of the subset. In order to properly group messages according to similar content and focus, we developed a novel distance metric that reflects the specifics of microblog data based on the qualities hashtags, user-mentions, and non-stopword terms. From each of these qualities a feature vector is created for each message and the distance between messages is computed using weighted cosine-similarity of these vectors. The result of the clustering is a binary tree of agglomerated subsets, which is then visualized in a treemap display of representative tag clouds (G). The user can explore this hierarchical overview using the mouse wheel, which changes the display layer of the tree hierarchy, and by selecting nodes, which shows the individual messages contained in a sub-tree of the hierarchy in a table (H). Selected nodes can be further zoomed in by clicking the zoom buttons in the control panel. In this case, a minimap (I) shows the current viewport within the complete tree structure.

3 Analysis Process: Usage Example

The challenge data comprised the time-frame from 17:00 to 21:30 on January 23, 2014. One of the more important events was a large fire that affected the Dancing Dolphin apartment building on the east side of Abila around 18:40. The event can be detected immediately as one of the larger peaks in the temporal overview that also includes several negative (red) messages. As the events unfold, the user can create spatial and temporal selections to investigate them in the hierarchy view, which indicates that people have been rescued, that paramedics are treating them outside the building and that an evacuation perimeter has been established. A second large peak dominated by negative sentiments appears at around 19:30 (see Fig. 2). Investigating it with the hierarchical view quickly highlights a shooting that takes place near the Gelatogalore ice cream shop between the police and drivers of a black van. It is also indicated that two hostages are held captive in the van (see Fig. 3). By highlighting further messages about the van using a Boolean combination of keyword searches, the analyst can trace its previous movements on the map. The messages show that the van hit another car and a bicyclist earlier and that it started its hasty run in the east part of the city. Based on these observations the analyst can suspect that the police presence and evacuation measures around the Dancing Dolphin fire might have caused the escape run and conclude that the kidnappers might have their hideout in that area.

Acknowledgments

We would like to thank the IEEE VAST Challenge 2014 team and all reviewers.

References

[1] H. Bosch, D. Thom, F. Heimerl, E. Puttmann, S. Koch, R. Kruger, M. Worner, and T. Ertl. Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2022–2031, 2013.