A Framework for Achieving Situational Awareness during Crisis based on Twitter Analysis

Andrea Zielinski (1), Laurissa Tokarchuk (2), Stuart Middleton (3), and Fernando Chaves (1)

(1) Fraunhofer Institute of Optronics, System Technologies and Image Exploitation - IOSB, Karlsruhe, Germany
(andrea.zielinski@iosb.fraunhofer.de), (2) Queen Mary and Westfield College, University of London, United Kingdom, (3)
University of Southampton - IT Innovation Centre, Southampton, United Kingdom

Decision Support Systems for Natural Crisis Management increasingly employ Web 2.0 and 3.0 technologies for future collaborative decision making, including the use of social networks like Twitter. However, human sensor data is not readily accessible and interpretable, since the texts are unstructured, noisy and available in various languages.

The present work focusses on the detection of crisis events in a multilingual setting as part of the FP7-funded EU project TRIDEC and is motivated by the goal to establish a Tsunami warning system for the Mediterranean.

It is integrated into a dynamic spatial-temporal decision making component with a command and control unit's graphical user interface that presents all relevant information to the human operator to support critical decision-support. To this end, a tool for the interactive visualization of geospatial data is implemented: All tweets with an exact timestamp or geo-location are monitored on the map in real-time so that the operator on duty can get an overall picture of the situation. Apart from the human sensor data, the seismic sensor data will appear also on the same screen.

Signs of abnormal activity from twitter usage in social networks as well as in sensor networks devices can then be used to trigger official warning alerts according to the CAP message standard. Whenever a certain threshold of relevant tweets in a HASC region (Hierarchical Administrative Subdivision Code) is exceeded, the twitter activity in this administrative region will be shown on a map.

We believe that the following functionalities are crucial for monitoring crisis, making use of text mining and network analysis techniques: *Focussed crawling*, *trustworthyness analysis geo-parsing*, and *multilingual tweet classification*.

In the first step, the Twitter Streaming API accesses the social data, using an adaptive keyword list (*focussed crawling)*. Then, tweets are filtered and aggregated to form counts for a certain time-span (e.g., an interval of 1-2 minutes). Particularly, we investigate the following novel techniques that help to fulfill this task: *trustworthyness analysis* (linkage analysis and user network analysis), *geo-parsing* (locating the event in space), and *multilingual tweet classification* (filtering out of noisy tweets for various Mediterranean languages). Lastly, an aberration algorithm looks for spikes in the temporal stream of twitter data.