



Multilingual Analysis of Twitter News in Support of Mass Emergency Events

A. Zielinski (1), U. Bügel (1), L. Middleton (2), S. E. Middleton (2), and L. Tokarchuk (3)

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

Social media are increasingly becoming an additional source of information for event-based early warning systems in the sense that they can help to detect natural crises and support crisis management during or after disasters.

Within the European FP7 TRIDEC project we study the problem of analyzing multilingual twitter feeds for emergency events. Specifically, we consider tsunami and earthquakes, as one possible originating cause of tsunami, and propose to analyze twitter messages for capturing testified information at affected points of interest in order to obtain a better picture of the actual situation. For tsunami, these could be the so called Forecast Points, i.e. agreed-upon points chosen by the Regional Tsunami Warning Centers (RTWC) and the potentially affected countries, which must be considered when calculating expected tsunami arrival times. Generally, local civil protection authorities and the population are likely to respond in their native languages. Therefore, the present work focuses on English as "lingua franca" and on under-resourced Mediterranean languages in endangered zones, particularly in Turkey, Greece, and Romania. We investigated ten earthquake events and defined four language-specific classifiers that can be used to detect natural crisis events by filtering out irrelevant messages that do not relate to the event.

Preliminary results indicate that such a filter has the potential to support earthquake detection and could be integrated into seismographic sensor networks. One hindrance in our study is the lack of geo-located data for asserting the geographical origin of the tweets and thus to be able to observe correlations of events across languages. One way to overcome this deficit consists in identifying geographic names contained in tweets that correspond to or which are located in the vicinity of specific points-of-interest such as the forecast points of the tsunami scenario. We also intend to use twitter analysis for situation picture assessment, e.g. for planning relief actions.

At present, a multilingual corpus of Twitter messages related to crises is being assembled, and domain-specific language resources such as multilingual terminology lists and language-specific Natural Language Processing (NLP) tools are being built up to help cross the language barrier. The final goal is to extend this work to the main languages spoken around the Mediterranean and to classify and extract relevant information from tweets, translating the main keywords into English.