Investigating the relationship between earthquakes and online news

Stephen Camilleri (1), Joel Azzopardi (1), and Matthew Agius (2)

(1) University of Malta, Faculty of Information and Communication Technology, Malta, (2) University of Malta, Faculty of Science, Malta

News agencies have the leading role of deciding which events are broadcast and the coverage allotted especially when a major disaster such as an earthquake strikes. Very few studies discuss how news agencies react when it comes to the reporting of such events, particularly the relationships between earthquake magnitude, causalities, damage and distance away from the earthquake. As more news agencies are turning to online platforms, analysing the numerous online news reports becomes a challenge in its own right.

We present results from a software application that harvests and processes automatically multilingual online news reports published out on the fly. This study exploits the use of text mining tools to automatically map in near real-time earthquake events with newspaper articles published by news agencies. The application downloads the multilingual articles published by 23 leading news agencies in real-time from six continents around the world. Articles discussing quakes are then identified, clustered, and processed for feature extraction. Earthquake-related features that are mined include the magnitude, location and timestamp from the corresponding news reports, as well as the number of casualties, injured and structural damage caused. Cluster feature extraction is also carried out for cross-referencing purposes. Each cluster is then mapped against earthquake events, aggregated from seismic bulletin stored by the United States Geological Survey (USGS). The results are complemented with a series of tests evaluating the accuracy of the extraction and clustering, followed by a modest impact analysis and recommendations for future work in this domain area. This tool paves the way for a more in-depth analysis on the possible temporal correlation (how long the disaster and its effects remain discussed in the news) and spatial correlation (geographical distance between news sources reporting the event and the location of the disaster itself) that may exist between newspaper articles and earthquakes.