GEOSPATIAL CROWDSOURCED DATA QUALIFICATION FOR SPATIAL DATA INFRASTRUCTURE BASED DISASTER MANAGEMENT

Koswatte S.[1], McDougall K.[1*] and Liu X.[2]

[1]Department of RSGIS, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka, Sri Lanka

[2]School of Civil Engineering and Surveying, Faculty of Health, Engineering and Science, University of Southern Queensland, Australia

[*]sam@geo.sab.ac.lk

The reporting of disasters has changed from official media reports to citizen reporters who are at the disaster scene. This kind of crowd based reporting, related to disasters or any other events, is often identified as 'Crowdsourced Data' (CSD). The quality of CSD is often problematic as it is often created by the citizens of varying skills and backgrounds. CSD is considered unstructured in general, and its quality remains poorly defined. This research designed a CSD quality assessment framework and tested the quality of the 2011 Australian floods' Ushahidi Crowdmap and Twitter data. Location availability of the Ushahidi Crowdmap and the Twitter data assessed the quality of available locations by comparing three different datasets i.e. Google Maps, OpenStreetMap (OSM) and Queensland Department of Natural Resources and Mines' (QDNRM) road data. Missing locations were semantically extracted using Natural Language Processing (NLP) and gazetteer lookup techniques. The Credibility of Ushahidi Crowdmap dataset was assessed using a naive Bayesian Network (BN) model commonly utilised in spam email detection. CSD relevance was assessed by adapting Geographic Information Retrieval (GIR) relevance assessment techniques which are also utilised in the IT sector. Thematic and geographic relevance were assessed using Term Frequency – Inverse Document Frequency Vector Space Model (TF-IDF VSM) and NLP based on semantic gazetteers.Results of the CSD location comparison showed that the combined use of non-authoritative and authoritative data improved location determination. The semantic location analysis results indicated some improvements of the location availability of the tweets and Crowdmap data. The results of the credibility and relevance analysis revealed that the spam email detection approaches and GIR techniques are feasible for CSD credibility and relevance detection.

Keywords: *Crowdsourced Data, SDI, Disaster Management, Credibility, Relevance*