Extracting Entities of Emergent Events from Social Streams Based on a Data-Cluster Slicing Approach for Ontology Engineering

Chung-Hong Lee, Department of Electrical Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan
Chih-Hung Wu, Department of Electrical Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan

ABSTRACT

In this paper, we describe our work on developing a model and method for extracting key entities from the online social messages regarding emergent events for enhancing ontology engineering, enabling a sensible solution for prevention of similar disasters. Our work started with the development of an event modelling system using a data-cluster slicing approach, which combines analytics of social data and event lifecycle algorithms, allowing for large-scale emerging novel events to be quickly and accurately analyzed. Subsequently, our system computes the energy of each collected event data sets, and then encapsulates ranked temporal, spatial and topical keywords into a structured node for event-entity extraction, in order to update event ontologies for fast response of emergent events. The preliminary experimental results demonstrate that our developed system is workable, allowing for prediction of possible evolution and early warning of critical incidents with a support of dynamic entity extraction.

Keywords: Information Retrieval, Data Mining, Entity Extraction, Ontology, Ontology Engineering

1. INTRODUCTION

Entity extraction is a key process for ontology engineering. In most cases, the information of a specific emergent event involves several interdisciplinary domain ontologies. The corpus of the event related social-messages (e.g. tweets) contains fruitful information, which is valuable for establishing operational procedures that can contribute to solutions for disaster prevention. Integrating such diverse kinds of information presents a challenge to anyone seeking to

DOI: 10.4018/IJIRR.2015070101
establish a coherent picture of the developing situation and to enable sensible decision-making operations. For emergency management, one of the key challenges encountered in social data mining is event validation by consecutive social streams based on their temporal and spatial factors. The decision making process requires real-time processing of information need for situational awareness, and standardized operating procedure that people can follow for quick response to critical events. One valuable application might be an ontology entities extractor that discovers entities from the online social messages regarding specific events. Such extensible event ontology has a significant role to play in the task of improving the ability to process information. Motivated by this need, in this work we utilize social messages collected from Twitter for building an ontology model to cope with emerging disasters events. It is expected that the extended ontology and knowledge can be used to locate essential information to cope with novel disasters.

This study was developed using a novel event modelling method based on the data-cluster slicing method. By analysing the publicly available social streams, many large accidents or disasters could be immediately detected for investigation (Gong, Zeng, & Zhang, 2011; Lee et al., 2012). Although the applications related to event detection by using social datasets (e.g. tweets) have been largely developed in recent years (Aggarwal, Han, Wang, & Yu, 2003; Chi, Song, Zhou, Hino, & Tseng, 2007; Feng, Martin, Weinig, & Aoying, 2006; Hung-Leng, 2009; Zhu & Shasha, 2002), little work has been done to produce a consistent information source related to ontology engineering using real-time social data, with which an online extensible ontology can be learned and updated for emergency response. We first implement an early warning system that provides a faster way to detect events in their early stage with a new term weighting technique to deal with huge data of social media streams. Then, the detected-event datasets can be used to build an online event-ontology learning model that can directly identify and extract candidate terms for updating the event ontologies.

The goal of this work is to design a way to extract the key entities for updating extended event-ontologies, rather than starting to construct ontologies from scratch. We focus on the online community in Twitter and attributing the content associated with specific events for investigation, allowing for developing relevant applications for prediction of possible evolution and early warning of disasters.

2. RELATED WORK

In recent years, research on social media has been popular in fields of text mining and information retrieval, and such new types of user-generated content challenges researchers to mine valuable knowledge from massive message streams. One of the advantages of social messages is that it enables people to achieve a near real-time information awareness. The social messages associated with the events truly depict the actual disastrous reality and detailed description of the event development, which contains known and unknown knowledge provided by local people and experts. Thus, social messages offer real materials for learning the experiences to cope with unexpected emerging real-world events. As mentioned previously, the goal of this work is to design a way to extract the key entities for updating extended event-ontologies, rather than starting to construct ontologies from scratch. Through computational algorithms for analyzing the content of social messages, the extracted terms are useful for carrying out event-ontology engineering. Ontologies are shared models of a specific domain with the criteria including clarity, coherence, extendibility, minimal encoding bias, and minimal ontological commitment (Gruber, 1995). For the distribution and management of event information, many upper ontologies have defined the nature of events for information systems and are distributed on the Web. For instance, Masolo et al. (2002) built the DOLCE (Descriptive Ontology for Linguistic and Cogni-
The Impact of the Mode of Data Representation for the Result Quality of the Detection and Filtering of Spam

Named Entity Based Ranking with Term Proximity for XML Retrieval
[www.igi-global.com/article/named-entity-based-ranking-with-term-proximity-for-xml-retrieval/198965?camid=4v1a](www.igi-global.com/article/named-entity-based-ranking-with-term-proximity-for-xml-retrieval/198965?camid=4v1a)