

Geocoding Tweets Approach Based on Conceptual Representations in the Context of the Knowledge Society

Imelda Escamilla, CIC, Instituto Politécnico Nacional, Mexico City, Mexico

Miguel Torres-Ruiz, CIC, Instituto Politécnico Nacional, Mexico City, Mexico

Marco Moreno-Ibarra, CIC, Instituto Politécnico Nacional, Mexico City, Mexico

Rolando Quintero, CIC, Instituto Politécnico Nacional, Mexico City, Mexico

Giovanni Guzmán, CIC, Instituto Politécnico Nacional, Mexico City, Mexico

Vladimir Luna-Soto, CIC, Instituto Politécnico Nacional, Mexico City, Mexico

ABSTRACT

In this paper, an approach to geocode tweets published in Spanish is proposed. The tweets are related to traffic events within an urban context of the Mexico City. They are generated by a particular phenomenon for knowing the behavior of the involved geographic entities. In order to disambiguate and verify the consistency of information, an application ontology was defined. Thus, the core goal is to identify location as well as spatial relationships between entities presented in the events, using semantic and spatial analysis of the collected dataset. In consequence, a visualization method for presenting the results was also proposed. The paper describes the methodology for enabling the discovery of spatial patterns within traffic tweets and provides useful information to make timely decisions and contribute in the context of Knowledge Society.

KEYWORDS

Application Ontology, Conceptual Representation, Geographic Object, Knowledge Society, Named Entity Recognition, Semantic Geocoding, Spatial Relationship, Tweets

1. INTRODUCTION

The current society has knowledge as one of its most important values and indeed this is often called Knowledge Society. The application of advanced software technologies in the context of the Knowledge Society is a bold contribution of the software engineering scientific community and a joint vision for applied humanistic computing.

DOI: 10.4018/IJSWIS.2016010103

Copyright © 2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

In the last years, Twitter has become a popular micro blogging platform, with over 400 million tweets posted daily according to tweespeed.com¹, making it a tool that can help significantly in the knowledge society due to its agile reading (no more than 140 characters), dynamic (information available in real time), accessible (for almost any device connected to Internet), functional (allows you to embed pictures, videos and links to other content), organized (with hashtags that represent subjects and ordered by date of publication), interactive (can view posts from other people, follow them, respond, share your posts by retweet or save them to mark them as favorites), non-invasive (no chat Instant Messaging) and with the possibility of anonymity (using nicknames or impersonal nicknames) (Pérez et al., 2012; Kassens, 2012; Welch & Bonnan, 2012).

This has led many research efforts on various topics to exploit this information such as event detection (Agarwal et al., 2012; Atefeh & Khreich, 2015), health monitoring (Nielsen et al., 2015), emergency detection (Seol et al., 2013), and among others. Many of these applications can benefit from information about the location, where the events occur, but unfortunately, this information is very poor, because only 1% of tweets contain geo-tags (Takhteyev et al., 2012).

The extraction of information from tweets presents some challenges, i.e., information is completely unstructured and its limited to 140 characters, tweets can contain grammatical errors, and abbreviations and each user has its own writing style, so information can be incomplete, false or not credible (Ritter, 2012).

However, Gutierrez et al. (2015) and Oussalah et al. (2013) established that the use of information content in tweets, provides geographic information, because the texts commonly refers to further locations. The tweet analysis allows us to know and evaluate social and natural events. Nevertheless, geocoding methods are used to translate geographic locations represented in the text (*e.g.* detection and location of events in a geographic area). They have focused on point feature type (Quin et al., 2013; Hart & Zandbergen, 2013; Krumm & Horvitz, 2015) and there are not approaches oriented towards polygon representation.

Thus, in this paper, a methodology focused on geocoding events appearing in tweets about traffic events of the Mexico City is proposed. The work consists of identifying events, geographic features and their spatial relationships, supported by conceptual representations, Natural Language Processing techniques and classification algorithms. Unlike other works, the detection of trending topics is not supported in this paper.

This paper is organized as follows: Section 2 describes the related work in the fields of geocoding based on short texts. The proposed methodology is presented in Section 3. In Section 4 the evaluation and comparison of the proposed method with other approaches are presented. Section 5 outlines the conclusions and future work.

2. RELATED WORK

According to Blanco et al. (2015) the automatically generating links to a knowledge base is one way of providing semantics to digital items. Entity linking has become an ubiquitous approach to add semantics to different media types, most notably in text (Zhao et al., 2015; Moro et al., 2014) and across different genres, such as news, archives, or tweets (Park & Lee, 2011; Bron et al., 2011).

Blanco et al., (2015) described that the main challenge entity linking systems have to face is ambiguity, *i.e.*, how to distinguish among different entities with the same label and also a possible lack of specificity, in which a case more generic and possibly less meaningful entities are identified.

These issues are generally addressed by identifying segments, key phrases or weighting terms based on text documents associated with a particular geographic context. There are several approaches

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/article/geocoding-tweets-approach-based-on-conceptual-representations-in-the-context-of-the-knowledge-society/149202?camid=4v1

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Computer Science, Security, and Information Technology, InfoSci-Select, InfoSci-Computer Systems and Software Engineering eJournal Collection, InfoSci-Networking, Mobile Applications, and Web Technologies eJournal Collection, InfoSci-Journal Disciplines Engineering, Natural, and Physical Science, InfoSci-Select. Recommend this product to your librarian:

www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

Hidden Markov Models for Context-Aware Tag Query Prediction in Folksonomies

Chiraz Trabelsi, Bilel Moulahi and Sadok Ben Yahia (2012). *Collaboration and the Semantic Web: Social Networks, Knowledge Networks, and Knowledge Resources* (pp. 168-190).

www.igi-global.com/chapter/hidden-markov-models-context-aware/65693?camid=4v1a

RIKEN MetaDatabase: A Database Platform for Health Care and Life Sciences as a Microcosm of Linked Open Data Cloud

Norio Kobayashi, Satoshi Kume, Kai Lenz and Hiroshi Masuya (2018). *International Journal on Semantic Web and Information Systems* (pp. 140-164).

www.igi-global.com/article/riken-metadatabase/193932?camid=4v1a

A Comparative Analysis of Online Social Networking Sites and Their Business Models

T. Andrew Yang and Dan J. Kim (2010). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications* (pp. 662-672).

www.igi-global.com/chapter/comparative-analysis-online-social-networking/39197?camid=4v1a

Information Retrieval by Semantic Similarity

Angelos Hliaoutakis, Giannis Varelas, Epimenidis Voutsakis, Euripides G.M. Petrakis and Evangelos Milios (2006). *International Journal on Semantic Web and Information Systems* (pp. 55-73).

www.igi-global.com/article/information-retrieval-semantic-similarity/2824?camid=4v1a