Semantic Web 5 (2014) 337–339
DOI 10.3233/SW-140141
IOS Press

Editorial

# The Semantics of Microposts

Aba-Sah Dadzie [a], Matthew Rowe [b] and Milan Stankovic [c]
[a] *The University of Birmingham, UK*
*E-mail: a.dadzie@cs.bham.ac.uk*
[b] *Lancaster University, UK*
*E-mail: m.rowe@lancaster.ac.uk*
[c] *Université Paris-Sorbonne/Sépage, Paris, France*
*E-mail: milstan@gmail.com*

## 1. Introduction

The call for this special issue on the *Semantics of Microposts* resulted from discussions initiated by the first *Making Sense of Microposts* workshop. The workshop was born out of the recognition of the invaluable knowledge contained in the vast, heterogeneous, distributed, continuously updating data store comprised of microposts – very small chunks of typically instantaneously published data, posted via social media and other communication channels that restrict publication unit size.

Microposts continue to grow in popularity as a means of communication. This is despite, or maybe because of, the restrictions in post size (typically 140/160 characters – Twitter/SMS, and/or associated, small multi-media elements, e.g., short YouTube videos, Instagram photos, Pinterest pins, Foursquare check-ins, g+ chats, Facebook likes). Near permanent connectivity and the advent of inexpensive smartphones, and specialised apps and short codes for microblogging via even feature phones, has resulted in an explosion in the generation of microposts, as they make possible near real-time dissemination of news and events, both public and personal. Twitter, for instance, reported as at 2014, 78% of its active users tweeting from a phone.[1] Facebook reported for Mar 2014, 609 million users active via a mobile device, out of a total of 802 million active daily.[2]

The availability of technology across all walks of life increases the ease with which microposts are published, from personal and shared devices, and even public kiosks and digital displays. As a result, microposts are often used to document both day-to-day and unusual events in the lives of individuals, support intercommunication between “friends” and with the general public, and enable interaction with other personas. (Twitter and Facebook accounts for pets are not unusual; personas may also be created to disseminate information from or about a non-human entity, e.g., the Mars rover Curiosity, wearable devices). Microposts are used to ask questions, publish opinions about people and places, post news on breaking events, make announcements within a community, or share a fleeting thought with the general public. Spontaneously formed and purpose-driven hashtags, a particular feature of microposts and used across different communication channels, provide a sense of (transient) community and a filter that may be used in subsequent data aggregation, exploration and analysis.

The ‘Making Sense of Microposts’ workshop has the sub-title *Big things come in small packages*. While each individual post is very small, as a collection microposts capture a broad view on the physical and online lives of the connected public. Foursquare, for example, reported over 6 billion check-ins as at May 2014, with millions more each day.[3] Twitter, arguably one of the most popular microblogging tools today, re-

[1] https://about.twitter.com/company

[2] newsroom.fb.com/company-info

[3] https://foursquare.com/about

ported 500 million tweets per day as at 2014.[1] The large volume of microposts published on a daily basis, from a variety of devices and platforms, comprises collective intelligence that may be mined for reuse in different scenarios, e.g., to gauge the public mood at large scale entertainment events, during periods of political upheaval and in political or marketing campaigns; to locate people and resources during natural and manmade emergencies and disasters; to provide context- and location-based guides to the traveller; to encourage discussion and collaboration in education.

The tasks involved in making sense of microposts, which requires extraction of its semantic content, are therefore addressed from a number of perspectives. Within the Semantic Web community, Natural Language Processing and Computational Linguistics, approaches used typically involve automated and semi-supervised methods for large-scale data mining, information extraction and content analysis. In Social and Web Sciences, focused, qualitative analysis is carried out on smaller datasets, to allow deeper insight to be derived from investigation of, among others, the social element of this phenomenon. Findings from such analysis are typically fed into augmenting large-scale content analysis and knowledge acquisition. Visualisation is also often employed, to present the results of data mining and analysis, and as research progresses, to support deeper exploration and complex analysis of micropost data. Information visualisation is also popular in data browsing and search applications targeted at the lay user that reuse the content of microposts, e.g., tools that guide tourists through a city or visitors to an exhibition.

A key challenge in the analysis of microposts is that the brevity that eases publication and dissemination introduces complexity in analysis, resulting as it does in the use of non-standard abbreviations and terse language. The need for novel approaches for effective, efficient analysis and reuse of this data continues to engage the research community and, increasingly, industry and commerce, government, state and other public bodies. This special call seeks to showcase the outcomes of work toward this aim and to point toward new avenues of research and applications.

We received 5 complete submissions, some of which were extended versions of papers submitted to the first 'Making Sense of Microposts' (#MSM2011) workshop. The special issue features 2 out of these, one of which follows on from work presented at #MSM2011, and the other an independent submission to the open call.

In *Reality Mining on Micropost Streams*, Balduini et al. use the augmented reality Android app *BOTTARI* to demonstrate the application of information retrieval with opinion mining and stream reasoning to provide personalised, location-based recommendations. BOTTARI is built on the LarKC Semantic Web framework,[4] and combines push with pull to provide timely, context-based information to its users. The system relies on the BOTTARI ontology, which extends the SIOC ontology[5] and the WGS84 vocabulary,[6] to allow encoding of opinion and location respectively. BOTTARI has been trialled in a restaurant district in Seoul, Korea, with 319 restaurants in a 2 km$^2$ area. The authors use, with a manually curated knowledge base containing information about these restaurants, a corpus of 109,390 tweets generated by approximately 30,000 users over 3 years, who rated 245 of the restaurants. Variation in the collection methods however resulted in 85% of the tweets falling within the last 6 months of the collection period. In an evaluation of the quality and the efficacy of its recommendations the authors found that data skew, with a large number of tweets posted by a small number of users, resulted in at best just over 50% of users receiving suitable or interesting recommendations. An interesting result discovered by the analysis was seasonal variation in preference for meal types and restaurants. The evaluation also reported very good scalability of the approach employed. As at the time of writing the paper, plans were underway to extend BOTTARI across Korea, and to integrate information from other social media platforms along with Linked Data.

*Social Influence Analysis in Microblogging Platforms – A Topic-Sensitive based Approach*, by Cano et al., assesses the influence exerted by posters about a topic, by examining how others engage with the ideas and opinions they express on micro-blogging platforms such as Twitter. The authors enrich tweet content using Zemanta[7] and OpenCalais,[8] following which they derive semantic profiles by translating the enriched posts into triples, using the SIOC,[5] OPO[9]

[4] See http://www.larkc.eu.

[5] Semantically-Interlinked Online Communities ontology: http://sioc-project.org/ontology.

[6] Basic Geo (WGS84 lat/long) Vocabulary: http://www.w3.org/2003/01/geo.

[7] http://www.zemanta.com

[8] http://www.opencalais.com

[9] Online Presence Ontology: http://online-presence.net.

and AO[10] ontologies. These are then used to measure the importance attached to the original tweet or poster, with relevance to the topic of interest, and also as an endorsement of quality (of the original tweet). The paper describes the derivation of a new algorithm, the *topic-entity PageRank (TPR)*, based on the Topic-Sensitive PageRank algorithm. The authors evaluate their approach to ranking influence using a corpus containing 100,000 retweets written in English, out of over 2.2 million tweets collected over 5 days. The retweet network comprised 32,626 unique users, of which 1,417 were retweeters, with 50739 edges (retweet relationships). The evaluation compares TPR with three state of the art algorithms: *Hyperlink-Induced Topic Search (HITS)*, *In-degree (ID)* and *Topic-Sensitive PageRank (TSPR)*, using two tasks, the first to rank influence in 18 pre-selected topics, and a recommendation task. The authors found performance of TPR and correlation between the different approaches to be influenced by topic. Further, with the exception of 1 topic, TPR outperformed the other three algorithms. The paper concludes with a description of a prototype developed for visual exploration of the topic-based influence measure derived by TPR.

Putting together this special issue was a team effort. To all who submitted papers, whether accepted or not, interest and continued research in the topic is vital in advancing the field. Each piece of work contributes to our efforts to improve the extraction of semantics and the knowledge content in micropost data, and therefore optimise use of this rich knowledge.

Our reviewers, listed below, provided detailed feedback to the authors and suggestions for improving the original submissions. We thank each one who dedicated their time and resources toward making this issue possible.

**Sofia Angeletou** British Broadcasting Corporation, UK
**Uldis Bojars** SIOC Project, Latvia
**Guillaume Erétéo** Orange Labs, France
**Annalisa Gentile** The University of Sheffield, UK
**Jelena Jovanovic** University of Belgrade, Serbia
**Philippe Laublet** Université Paris-Sorbonne, France
**Diana Maynard** The University of Sheffield, UK
**Pablo Mendes** IBM Research Almaden, USA
**Alexandre Monnin** INRIA Sophia Antipolis, France
**José M. Morales del Castillo** El Colegio de México, Mexico
**Harald Sack** University of Potsdam, Germany
**Bernhard Schandl** mySugr GmbH, Vienna, Austria
**Mischa Tuffield** PeerIndex, UK
**Victoria Uren** Aston Business School, UK
**Claudia Wagner** GESIS – the Leibniz Institute for the Social Sciences, Germany

[10] Annotation Ontology: https://code.google.com/p/annotation-ontology.