

#Microposts2015 – 5th Workshop on ‘Making Sense of Microposts’

Big things come in small packages

Matthew Rowe
Sch. of Computing & Comms.
University of Lancaster
Lancaster, UK
m.rowe@lancaster.ac.uk

Milan Stankovic
Sépage / Université
Paris-Sorbonne
Paris, France
milstan@sepage.fr

Aba-Sah Dadzie
Sch. of Computer Science
The University of Birmingham
Birmingham, UK
a.dadzie@cs.bham.ac.uk

Categories and Subject Descriptors

H.3.5 [Online Information Services]: Web-based services

Keywords

Microposts, Microblogs, Social Media, User & Data Analysis

1. INTRODUCTION

#Microposts2015, the 5th workshop on *Making Sense of Microposts*, is summarised by the sub-theme: *big things come in small packages*. The workshop was borne out of research we were each carrying out as microblogging platforms became increasingly popular, and their value as a publishing platform and the data generated as a result began to be recognised. This phenomenon continues to grow, as microblogs provide a low-effort means of publishing information within private, but moreso public, fora, giving a voice to all in all arenas. As Clark & Aufderheide observed in their 2009 whitepaper: “*the people formerly known as the audience now are at the center of media*” [3]. We coined the term *Microposts* to symbolise these posts - very small in terms of individual size and effort to publish, but as a collective source of constantly updating, typically context-sensitive information and knowledge, very large scale. Microposts use predominantly text (e.g., tweets, Facebook shares, Google +1s and Foursquare check-ins). However, we see, with an increase in posting in the moment, often from camera phones and tablets, microposts incorporating or comprising solely of (tagged) images and video, e.g., as hosted on Instagram and Pinterest, and embedded into tweets and check-ins.

Importantly, the #Microposts workshop does not focus only on the data generated by end users, but encourages studies about the users themselves, their local context, which platforms they use, the networks they build, and the devices from which they send Microposts. With each workshop we learn more about the subject, and the practical challenges researchers face in the study of Microposts and

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
<http://dx.doi.org/10.1145/2740908.2743860>.

their publishers. With this in mind, the #Microposts2015 workshop continues a tradition of encouraging participation from a wide range of fields both within and outwith Computer Science as well as cross-disciplinary work, and also actively targets Social Science research by including a dedicated track. To respond to challenges in information extraction for this particular type of data – key being brevity and the use of non-standard language, we are running, for the third time, an information extraction challenge: the Named Entity rEcognition and Linking (NEEL) Challenge.

2. WORKSHOP SCOPE

This year’s topics are grouped under three main areas:

UNDERSTAND – keen to highlight the importance of a focus on the human in the generation, analysis and reuse of Microposts, we encourage work that improves understanding of how situation and context drive individual and collective publication of Microposts, and the effect of targeting these at specific entities. Such studies address, among others, collective awareness and citizen empowerment, civil action, politics, ethics, media, profiling, cultural and sociological issues, and as technology advances, the use of wearable and other personal devices to generate and consume microposts.

DISCOVER – the extraction and analysis of information content from Micropost data contribute to pattern and trend discovery. This is key to knowledge discovery and application, using approaches including emergent semantics, data mining, sentiment analysis, prediction, network analysis, community and influence detection. As research on information extraction from Microposts advances, Microposts are being linked to the Web of Linked Data, to enrich knowledge content and allow more effective reuse of this collective intelligence.

APPLY – applications papers and case studies demonstrate re-use of Micropost data in a variety of areas, including citizen participation, emergency response, crowd tracking, collective intelligence, business and marketing intelligence, personalisation, recommendation, event and topic detection, and opinion mining.

3. WORKSHOP STRUCTURE

The workshop will be opened with a keynote talk that sets the theme for the year. We received twelve submissions from institutions in Europe, Asia, the Middle East and Africa

for the main and social sciences tracks; from these were selected five for presentation. The workshop features also a poster session, which gives the audience a chance to discuss emerging topics and provide feedback to participants on particularly interesting submissions that do not make it into the main track, work in progress and challenge submissions. The workshop will close with brief presentations from selected NEEL submissions and the award ceremony.

To allow all tracks to be published in a single volume, the proceedings will be published using the online, open-access CEUR scientific workshop proceedings publication service.

3.1 Main Track

Two full and one short paper were selected for presentation: Wang *et al.* use inherent features in tweets to investigate the effectiveness of each of five algorithms with four sets of features, to develop an approach to identify social spam in large Twitter datasets. They aim ultimately to support spam detection in real time. De Boom *et al.* present a novel perspective on event-detection in tweets. They demonstrate that combining machine learning with explicit semantics associated with tweets and hashtags can yield considerable improvement over state of the art event clustering approaches. Karatay & Karagoz explore, in depth, user profiling using named entity detection in tweets. Their aim is to filter and personalise information, to manage information load, and therefore, improve user engagement and experience.

Two posters were accepted: Agarwal & Sureka employ keyword and tag usage to identify individuals and communities taking advantage of the ability to post anonymously on Tumblr to promote hate and extremism. Edwards & Guy distinguish spam from non-spam on Twitter, as well as differences in spam accounts and behaviour exhibited following other spam accounts. They observe that genuine Twitter users tend to connect with specific types of spammers.

3.2 Social Sciences Track

New to 2015, a dedicated Social Sciences track is being chaired by Danica Radovanović (University of Novi Sad) and Katrin Weller (GESIS). In addition to topics that fall mainly under obtaining an understanding of Micropost generation and consumption, this track aims to address also two emerging topics: inequality in access and use of digital, social media, and the social and communication dynamics that result from the use of Micropost-based services.

Two papers were accepted for this track. A major event that spiked engagement on social media in 2015 was the Charlie Hebdo shooting in France on 7th Jan. Giglietto & Lee, in one of the first studies of reactions on Twitter to this event, study the use of the hashtag *#JeNeSuisPasCharlie*, in contrast to *#JeSuisCharlie*. Using data analysis approaches including activity patterns and word frequency, the authors demonstrate how tweets including *#JeNeSuisPasCharlie* resemble crisis communication patterns and support expressions of self-identity such as grief and resistance. Coelho *et al.* present a model for discursive analysis of posts generated on net activist groups, to identify elements that promote social empowerment in the political vitality present in digital culture. They create and articulate methods, instruments and resources for the collection and treatment of what constitutes big data, and for qualitative analysis of content.

The track is sponsored by *GESIS*, the Leibniz Institute for the Social Sciences [1], the largest service and infrastructure

institution for the Social Sciences in Germany. The award for the best submission is to highlight the contribution of the Social Sciences to research on Microposts and the users and phenomena that drive this specific type of data publication.

3.3 NEEL Challenge

With each year the (NEEL) challenge increases in complexity, to tackle an issue not addressed in previous challenges. The challenge continues to be an exciting and in itself challenging part of the workshop, run as a multi-step process over the months leading up to the workshop. NEEL is chaired by A. Elizabeth Cano (KM_i, OU), and Giuseppe Rizzo (EURECOM), assisted by dataset chairs Andrea Varga (Swiss Re) and Bianca Pereira (INSIGHT).

The challenge continues to receive expressions of interest from both industry and academia, and as in previous years, returning participants tackling the new challenge from new perspectives. The 2015 NEEL task has been extended to require the recognition and typing of entities, followed by linking to corresponding DBpedia entities. Additionally, the runtime computation efficiency of systems submitted will be measured. As at the time of writing the training dataset for the challenge has been released, and testing of the APIs is in progress. The challenge received 28 expressions of interest, out of which 21 registered for the final evaluation.

Both to encourage competition and support attendance at the workshop, as in previous years the winning submission will receive an award, sponsored in 2015 by *SpazioDati* [2], an Italian startup that focuses on text analytics and big data. We are pleased to note that *SpazioDati* were among the participants in the 2014 challenge.

4. ACKNOWLEDGMENTS

Putting *#Microposts2015* together was a team effort, and we acknowledge first the contribution of the chairs of our two extra tracks. A good number of the programme committee have reviewed for each of the five workshops, and we have welcomed again new members. Especial thanks to returning PC, whose support contributed to acceptance of our proposal for a 5th workshop. Thanks too, to all PC members, who provided valuable feedback to both us, as chairs, in selecting papers for presentation, and to the authors, resulting in a collection of papers that contribute to research on Microposts specifically and the wider Semantic Web. We thank all authors who submitted their work; each submission helps to advance the state of the art and contributes to the discussions on the workshop day itself.

Finally, we thank our sponsors, *GESIS* and *SpazioDati*, whose support has helped to encourage submission and reinforces the value of the research and practical work discussed at the *#Microposts* workshops.

5. REFERENCES

- [1] *GESIS*. <http://www.gesis.org>.
- [2] *SpazioDati*. <http://spaziodati.eu>.
- [3] J. Clark and P. Aufderheide. Public media 2.0: Dynamic, engaged publics. Available at: <http://cmsimpact.org/sites/default/files/documents/pages/publicmedia2.0.pdf>, 2009.
- [4] *#Microposts2015* website. <http://www.scc.lancs.ac.uk/microposts2015>.