

Preface

The 2nd Workshop on *Making Sense of Microposts* (#MSM2012) was held in Lyon, France, on the 16th of April 2012, during the 21st International Conference on the World Wide Web (WWW'12). #MSM2012 follows on from a successful 1st workshop, #MSM2011, at the 8th Extended Semantic Web Conference (ESWC 2011), which, with approximately 50 participants, was the most popular workshop at ESWC 2011.

The #MSM series of workshops is unique in targeting both Semantic Web researchers and other fields, both within Computer Science, such as Human-Computer Interaction and Visualisation, and in other areas, particularly the Social Sciences. The aim is to harness the benefits different fields bring to research involving microposts. Moving the 2nd workshop to WWW allowed us to reach a wider and more varied audience.

Posting information about on-going events, exchanging information with one's social and working circles, or simply publishing one's train of thought on online social media platforms such as Twitter and Facebook, or contributing information about points of interest on Foursquare, is increasingly the norm in the online world. Support for the use of online social media on or via ubiquitous, small, mobile devices and near-permanent connectivity have further lowered the barrier to interaction with the online world. This has resulted in an explosion of small chunks of information being published with minimal effort (e.g. a 'tweet' or a 'check-in' on foursquare) – we refer to such user input as *microposts*. The reality of the trend of microposts' domination in online, end user-generated content can be seen in the appearance of new services that focus primarily on low-effort user input, such as Google+, whose aim is to bootstrap microposts in order to more effectively tailor search results to a user's social graph and profile.

The sheer scale of micropost data, generated using a variety of devices and on multiple platforms, by myriad users in as many different situations, requires new techniques to glean knowledge and provide useful services and applications sitting atop the amalgamation of this heterogeneous, distributed data. Further, the brevity of user expression in microposts imposes additional challenges for

analysis. The #MSM workshop series was born to bring together researchers exploring novel methods for analysing microposts, and for reusing the resulting collective knowledge extracted from such posts, both on the Web and in the physical world. The #MSM2012 workshop discussed emerging to fairly advanced work on the research these challenges have engendered.

#MSM2012 continues to highlight the importance of maintaining a focus on the end user – ranging from the mainstream user of what is now ubiquitous technology, such as the mobile phone, tablet or desktop computer, with little to no technical expertise, to the Semantic Web expert – to ensure that appealing, useful and usable tools are designed and built, which harness the particular benefits of Semantic Web technology.

Many hearty thanks to all our contributors and participants, and also the Programme Committee whose valued feedback resulted in a rich collection of papers, posters and demos, each of which adds to the state of the art in leading edge research. We are confident that the #MSM series of workshops will continue to foster a vibrant community, and target the rich body of information generated by the many and varied authors whose social and working lives span the physical and online worlds.

Matthew Rowe KMi, The Open University, UK
Milan Stankovic Hypios / Université Paris-Sorbonne, France
Aba-Sah Dadzie The University of Sheffield, UK
#MSM2012 Organising Committee, April 2012

Introduction to the Proceedings

Out of a total of 19 paper submissions, 6 full and 3 short papers were accepted. This was in addition to a poster and demo session, to exhibit practical application in the field, and foster further discussion of the ways in which data extracted from Microposts is being reused. The accepted submissions cover an array of topics; we highlight these below.

The proceedings include also the abstract of the keynote, *'Information Theoretic Tools for Social Media'*, presented by Greg Ver Steeg, of the Information Sciences Institute at the University of Southern California.

Sentiment and Semantics

Platforms, such as Twitter and Facebook, that support micropost publication allow users to vent their frustrations and express their opinions in a centralised and public space. Passive networks formed on such platforms are comprised of users *listening* to the signals produced by other users and consuming their published information. As a consequence, sentiment analysis of microposts has become a useful means for companies and organisations to gauge the collective sentiment and opinion regarding different entities and topics. In Saif et al.'s paper, *'Alleviating Data Sparsity for Twitter Sentiment Analysis'*, the authors describe an approach to alleviate the data sparsity problem that affects sentiment analysis on Twitter through the use of semantic and sentiment-topic features. The authors demonstrate the efficacy of these additional features by outperforming a baseline model which neglects such additional information.

Semantics within microposts forms the basis for discussion in *'Small talk in the Digital Age: Making Sense of Phatic Posts'* by Radovanovic & Ragnedda. In this paper the authors present a theoretical discussion and analysis of the importance of microposts in providing diverse information across the Web. Following on from this theme of diversity and information utility is Zangerle et al.'s paper titled *'Exploiting Twitter's Collective Knowledge for Music Recommendations'*. In this work the authors demonstrate the utility of microposts in providing music recommendations through collective knowledge. In both works microposts are described as a useful source for diverse information that can in turn be used in differing applications and contexts.

Information Extraction

The masses of microposts published every day cover a wide range of subjects and topics. Extracting information from microposts related to the same entity or topic can provide a diverse perspective with regard to public perception and/or opinion. Prior to performing opinion analysis, entities must be recognised within the microposts and the information extracted accordingly. One issue of the diversity of microposts, however, is the prevalence of term ambiguity – where the same term can have multiple meanings. Context provides one mechanism for disambiguation; however, given the limited information size of a single micropost, obtaining such contextual information is challenging. This issue is addressed in the paper by Castro Reis et al. titled *'Extracting Unambiguous Keywords from Microposts Using Web and Query Logs Data'*, by automatically detecting, and hence, enabling the extraction of terms in microposts which are not ambiguous. The authors present a hand-crafted classifier that uses background knowledge of term features to yield high levels of precision, outperforming a Support Vector Machine in the same setting.

The second paper to address the topic of information extraction is *'Knowledge Discovery in distributed Social Web sharing activities'* by Scerri et al. In this work the authors address the challenges in managing and making optimal use of personal information, by arguing that social activity streams, both of a given user and members of his/her social network, provide useful means for the enrichment of personal information spaces. To this end the authors propose a framework for the extraction of information from disparate activity streams and the integration of the extracted information into existing personal information spaces, through the LivePost ontology presented.

Visualisation, Search and Networks

The scale and volume of microposts makes interpretation and analysis of such data limited to end users. One solution is to visualise microposts in a coherent and readable form, thereby facilitating sense-making and data exploration. The paper by Hubmann-Haidvogel et al. titled *'Visualizing Contextual and Dynamic Features of Microposts'* presents work that enables the visualisation of large volumes of microposts. The approach supports multi-faceted views to enable a range of information-seeking tasks. For instance, by presenting geographical information alongside topic-volume statistics, the end user is presented with an overview of microposts at a higher level of abstraction.

Search over microposts has recently become a topic of great interest, with the creation of the first 'Microblog' track¹ at the Text REtrieval Conference 2011. The diversity and ambiguity of terms found within microposts limits current retrieval paradigms and therefore requires the exploration of new methods for retrieval. In the paper by Tao et al. titled *'What makes a tweet relevant for a topic?'* the authors explore the effects of various features on retrieval performance over microposts. The features investigate the topic-sensitive and topic-independent effects on retrieval performance and find that by taking the former information into account performance is improved.

The final two papers investigate the dynamics and effects of networks associated with microposts. The first, by Wagner et al. titled *'When social bots attack: Modeling susceptibility of users in online social networks'*, assesses which users are likely to fall foul of socialbot attacks and be influenced by the content the bots produce. The authors explore three different feature sets to describe users and find that users who engage a lot in conversational behaviour with their social network are more susceptible to attacks. The second paper in the area of networks is *'Understanding co-evolution of social and content networks on Twitter'* by Singer et al. In this work the authors explore how networks change over time through time-series analysis of social network measures. Their findings indicate that social networks have an influence on content networks.

¹<https://sites.google.com/site/microblogtrack/home>

Workshop Awards

The Parisian Open Innovation startup, Hypios², sponsored an award for the submission that contributed best to making innovation happen on the Web. Best paper nominations were sought from the reviewers, and a final decision agreed by the Chairs, based on the nominations and review scores.

Additional Material

The call for participation and all paper, poster and demo abstracts are available on the #MSM2012 website³. The full proceedings are also available on the CEUR-WS server, as Vol-838⁴. The proceedings for the 1st workshop are available as CEUR Vol-718⁵.



Programme Committee

Fabian Abel Leibniz University Hannover, Germany
Gholam R. Amin Sultan Qaboos University, Oman
Sofia Angeletou KMi, The Open University, UK
Pierpaolo Basile University of Bari, Italy
Uldis Bojars University of Latvia, Latvia
David Beer University of York
John Breslin NUIG, Ireland
A. Elizabeth Cano The University of Sheffield, UK
Óscar Corcho Universidad Politécnica de Madrid, Spain
Danica Damljanovic The University of Sheffield, UK
Ali Emrouznejad Aston Business School, UK
Guillaume Ereteo INRIA, France
Miriam Fernandez KMi, The Open University, UK
Fabien Gandon INRIA, Sophia-Antipolis, France
Andrés Garcia-Silva Universidad Politécnica de Madrid, Spain
Anna Lisa Gentile The University of Sheffield, UK
Jon Hickman Birmingham City University, UK
Seth van Hooland Free University of Brussels, Belgium
Jennifer Jones University of the West of Scotland, UK
Jelena Jovanovic University of Belgrade, Serbia
Vita Lanfranchi, The University of Sheffield, UK
Philippe Laublet Université Paris-Sorbonne, France
Pablo Mendes Kno.e.sis, Wright State University, USA
João Magalhães Universidade Nova de Lisboa, Portugal
Julie Letierce DERI, Galway, Ireland
Diana Maynard The University of Sheffield, UK
Pablo Mendes Freie Universität of Berlin, Germany
José M. Morales del Castillo Universidad de Granada, Spain
Alexandre Passant DERI, Galway, Ireland
Danica Radovanovic University of Belgrade, Serbia
Yves Raimond BBC, UK
Harald Sack University of Potsdam, Germany
Bernhard Schandl University of Vienna, Austria
Andreas Sonnenbichler KIT, Germany
Raphaël Troncy Eurecom, France
Victoria Uren Aston Business School, UK
Claudia Wagner Joanneum Research, Austria
Shenghui Wang Vrije University, The Netherlands
Katrin Weller University of Düsseldorf, Germany
Ziqi Zhang The University of Sheffield, UK

²<http://hypios.com>

³<http://socsem.open.ac.uk/msm2012>

⁴<http://ceur-ws.org/Vol-838>

⁵<http://ceur-ws.org/Vol-718>