Preface

The 1st Workshop on Making Sense of Microposts (#MSM2011) was held in Heraklion, Greece, on the 30th of May 2011, during the 8th Extended Semantic Web Conference (ESWC 2011).

Posting information about on-going events on Twitter and Facebook; checking into and contributing new information about Points of Interest on Foursquare; updating one's status on a variety of social networking sites, using a variety of devices, while stationary or on the move, are now so commonplace that such platforms are often the first point of call for searching for and sharing information covering a wide range of events, topics and personal or work-related interests.

As a result, enormous quantities of small user input are being piped into the data streams of the Web, leading to a rate of growth which has never before been witnessed. We refer to such small user inputs as Microposts. The #MSM2011 workshop discussed emerging to fairly advanced work on the research this has engendered. Micropost data, which spans disparate, heterogeneous topics, therefore requires new techniques for information extraction and the leveraging of semantics from Microposts, to glean the knowledge contained, and new tools that make optimal use of the semantics encoded in Microposts'. The discussions also looked at studies related to Microposts, both social and from a more technically oriented perspective, that should contribute to building appealing new systems based on this type of data.

#MSM2011 is unique in targeting both Semantic Web researchers and other fields, both within Computer Science, such as Human-Computer Interaction, and outside Computer Science, particularly the Social Sciences, in order to harness the benefits different fields bring to research involving Microposts. #MSM2011 recognises the importance of maintaining a focus on the end user of Microposts' – ranging from the mainstream mobile phone or computer user with little to no technical expertise, to the Semantic Web expert – in order to ensure that appealing and usable tools are developed, that harness the particular benefits of Semantic Web technology.

Many hearty thanks to all our contributors and participants, and also the Programme Committee whose 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 this is the start of a series of workshops that will continue to target the rich body of information generated by the many and varied authors in the online world.

Matthew Rowe KMi, The Open University, UK
Milan Stankovic Hypios / Université Paris-Sorbonne, France
Aba-Sah Dadzie The University of Sheffield, UK
Mariann Hardey The University of Durham, UK
#MSM2011 Organising Committee, May 2011

Introduction to the Proceedings

Out of a total of 19 paper submissions, 7 full and 2 short papers were accepted, around which two main discussions were held. 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 here.

Information Diffusion and Influence

Three submissions to the workshop addressed the topics of Information Diffusion and Influence within Microposts. Weller *et al.*'s paper, '*Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences*', analysed the flow of information through tweets at scientific conferences, by assessing the diffusion of URLs and retweets. '*Making Sense of Location Based Micro-posts Using Stream Reasoning*' by Celino *et al.* proposes a framework to identify so-called *mavens* – experts on a given POI – utilising stream reasoning to handle the deluge of Microposts and enable effective identification. A demo by Huron *et al.*, titled '*Polemical Video Annotation by Twitter*', models arguments and discussions between users on Twitter during video broadcasts. The application enables contentious points in videos to be identified, and leads to further information exchange and debate.

Entity Extraction and Semantics

Microposts often refer to entities within their content; identifying such entities enables effective tracking of mentions and consensus concerning opinion. However, the limited length of Microposts makes detecting and extracting such references challenging. '*Extracting Semantic Entities and Events from Sports Tweets*' by Choudhury and Breslin presents an approach to this problem in the form of named entity recognition from sports tweets, testing various features for the detection task – i.e. linguistic features, statistical analysis and domain knowledge. Entity extraction is utilised in '*Follow me: Capturing Entity-Based Semantics Emerging from Personal Awareness Streams*' by Cano *et al.*, to first detect entities in users' personal awareness streams – derived from status updates coupled with the context of the user – before using such entities to suggest concepts that correlate with the context of the user.

The interestingly titled '*Does Size Matter? When Small is Good Enough*' by Gentile *et al.* presents the novel experiment of truncating emails from Micropost size (i.e. 140 characters) up to the full size of each email in a given corpus, and then performing text classification over the truncated emails. Results are compared with the classification using the full emails, showing that truncated emails provide a sufficient summarisation for accurate classification. In '*Discovering the Dynamics of Terms' Semantic Relatedness through Twitter*', by Milikic *et al.*, the semantic relatedness of terms in Microposts is assessed. Their approach measures the dynamic aspect of semantic relatedness over time under the hypothesis that the relation between terms is incorrectly assumed to be static.

Politics and Sentiment

Microposts enable opinion to be expressed to a global audience with relative ease. As a result, platforms that provide functionality to publish Microposts are often central to emotive discussions, such as political uprisings. The workshop accepted three papers that explore work in this area. Skilters *et al.*'s paper, '*The Pragmatics of Political Messages in Twitter Communication*', assesses political messages on Twitter; their analysis reveals a link between retweet popularity and offline political consensus. '*Automatic Detection of Political Opinions in Tweets*' by Maynard and Funk presents an approach to political opinion detection and analysis in tweets. The authors conjecture that a middle ground is required between sophisticated NLP techniques that function over rich review corpora and more naïve, simplistic, weighted lexicon-based approaches. They present work that attempts to fills this gap and evaluate their work over a large corpus of political tweets from the 2010 UK General Election.

To gauge sentiment in tweets, Nielsen's paper, '*A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs*', proposes a new sentiment word list. Detection of sentiment in tweets normally utilises a weighted semantic lexicon, looking up individual terms and returning their valence. Existing lexicons are available for this task, therefore Nielsen presents his new word list and compares its performance in sentiment detection against, among others, the ANEW semantic lexicon. Results show comparable performance.

Workshop Awards

Two awards were made, sponsored by the EU FP7 *WeGov* project[1]. Best paper nominations were sought from the reviewers, and a final decision agreed by the Chairs, based on the nominations and review scores. The best poster/demo award was based on nominations by participants during the workshop.

Additional Material

The call for participation and all paper, poster and demo abstracts are available on the #MSM2011 website[2]. The full proceedings are also available on the CEUR-WS server, at: `http://ceur-ws.org/Vol-718`.

[1] `http://www.wegov-project.eu`
[2] `http://research.hypios.com/msm2011`

Programme Committee

Harith Alani KMi, The Open University, UK
Sofia Angeletou KMi, The Open University, UK
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
Guillaume Ereteo INRIA, France
Fabien Gandon INRIA, France
Andrés Garcia-Silva Universidad Politécnica de Madrid, Spain
Jon Hickman Birmingham City University, UK
Robert Jäschke University of Kassel, Germany
Jennifer Jones University of the West of Scotland, UK
Jelena Jovanovic University of Belgrade, Serbia
Philipe Laublet Université Paris-Sorbonne, France
Pablo Mendes Kno.e.sis, Wright State University, USA
Eric T. Meyer Oxford Internet Institute
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
Elena Simperl University of Innsbruck, Austria
Raphaël Troncy Eurecom, France
Mischa Tuffield Garlik, UK
Victoria Uren The University of Sheffield, UK
Claudia Wagner Joanneum Research, Austria
Shenghui Wang Vrije University, The Netherlands
Ziqi Zhang The University of Sheffield, UK

Multidisciplinary Steering Committee

Alexandre Monnin IRI, Centre Pompidou, Paris, France
Danica Radovanovic University of Belgrade, Serbia