

Preface

The 4th Workshop on *Making Sense of Microposts* (*#Microposts2014*) was held in Seoul, Korea, on the 7th of April 2014, during the 23rd International Conference on the World Wide Web (WWW'14). *#Microposts2014* sees a change in the workshop acronym from *#MSM*, to highlight our focus on *Microposts* – small chunks of information published online with minimal effort, via a variety of platforms and devices. The *#Microposts* journey started in 2011 at the 8th Extended Semantic Web Conference (ESWC 2011), then moved to WWW in 2012, where it has stayed, for the third year now.

The *#Microposts* series of workshops is unique in targeting researchers from a range of fields spanning both Computer Science and the Social Sciences. The aim is to harness the benefits different fields bring to research involving *Microposts*, and to maintain a focus on the end user and their interaction with other users and the physical and online worlds – the community who collectively publish this rich, varied information.

Microblogging platforms and other restricted size, text only and multi-media capable, instant communication tools are now so commonplace that applications are constantly being developed to enable their use not only on the desktop, but also on the go, from ordinary and smart phones, tablets and even public kiosks. While the well-known microblogging platforms – Twitter, Facebook, MySpace, Google+, Tumblr, Foursquare, Instagram and Pinterest, among others – cover a large portion of the online user base, other country and/or language-specific platforms such as Sina Weibo, and mobile-based messaging tools such as WhatsApp, are increasingly being used to share *Micropost*-type information. This medium of communication is no longer patronised predominantly by individual users sharing information informally within private networks and also with the wider public, but is used as mouth pieces by enterprise organisations and public bodies, to foster a feeling of more personal interaction with consumers and the wider, participating public. Disaster and emergency response and management, and political upheaval and crises, are two areas where social media has been shown to be particularly powerful for disseminating critical information to the individuals involved, and for broadcasting events as they occur to the outside world. Citizen reporters and scientists are now commonly accepted, or even anticipated, by organisations that traditionally relied mainly on trained experts. Microblogging and other social media platform usage is seen across all walks of life, from opinion mining and feedback solicitation for public consultations, to election campaigns and classroom participation.

With increasingly lower cost methods for publishing *Microposts* (often via mobile devices), and widespread use of informal and abbreviated language, the sheer scale and heterogeneity of *Micropost*

data presents challenges for analysis, knowledge extraction and aggregation, further dissemination and reuse in any of a range of applications. At the same time, today's end user, understandably, has very high expectations for intuitive, minimal effort applications for tailored search and information retrieval across myriad, interconnected devices, customised to their current context – situation, location and proximity of others within their social and other networks, and influenced by unknown users with similar interests.

The *#Microposts* workshop was created to bring together researchers exploring novel methods for analysing *Microposts*, and for reusing the resulting collective knowledge extracted from such posts, both online and in the physical world. With each year we have seen novel, leading edge approaches to exploring this now ubiquitous, but still very valued, means of communication and the knowledge it generates. We are able to report wide interest in the workshop, with a good number of submissions from a range of fields in and across disciplines, mainly from Computer Science and the Social Sciences. Along with reports of applications in different domains, our contributors and audience have re-confirmed each year the importance of *Microposts* to the ordinary end user and, increasingly, public organisations and industry. Many hearty thanks to all our contributors and participants. Submissions came from institutions all over the world – the main track saw authors from institutions across 11 different countries, and the challenge from 7. Interestingly, while challenges are often more popular with students, half the challenge submissions included authors from research institutions, including Microsoft Research, the Max Planck Institute, CNRS (France) and SAP Research. Our Programme Committee are even more varied, coming from universities and research institutions around the world, as well as from industry, more than half of whom have reviewed for each of our four workshops. A very special thanks goes to each of them; their valued feedback resulted in a rich collection of papers and posters, each of which adds to the state of the art in leading edge research. We are confident that the *#Microposts* series of workshops will continue to foster a vibrant community, as we continue to work with the rich body of knowledge generated by the many and varied end users whose social and working lives span the physical and online worlds.

Matthew Rowe University of Lancaster, UK
Milan Stankovic Sépage / Université Paris-Sorbonne, France
Aba-Sah Dadzie The University of Birmingham, UK
#Microposts2014 Organising Committee, April 2014

Introduction to the Proceedings

The main workshop track attracted 12 submissions, 6 of which, all long papers, were accepted, along with a poster. These covered topics from machine learning, on Micropost classification and extraction, to data mining and analysis, and sentic and sentiment analysis. Applications were seen in incident and emergency response and management, and topic and opinion mining. We provide a brief introduction to these below.

The proceedings include the abstract of the keynote, ‘*Computational Social Science and Microblogs – The Good, the Bad and the Ugly*’, presented by Markus Strohmaier of the Dept. of Computer Science, University of Koblenz-Landau, Germany.

Main Track Presentations

Micropost Mining and Analysis

Panisson *et al.*, in their paper *Mining Concurrent Topical Activity in Microblog Streams*, present a novel approach to topic mining from Twitter streams, in the context of recreating event timelines. Their evaluation, performed on a dataset sampled from the London 2012 Summer Olympics, shows a high degree of matching between the inferred timeline and the actual Olympics schedule.

Prapula G *et al.* introduce the notion of episode in the extraction of events from tweets, in *TEA: Episode Analytics on Short Messages*. Detection of *episodes* – significant moments when a particular entity gets traction on Twitter – constitute the basis for the application scenarios they present. Using data visualisation and social media monitoring, the approach is evaluated on selected famous personalities and entities, including sports and brands.

The paper *Sentic API: A Common-Sense Based API for Concept-Level Sentiment Analysis*, by Cambria *et al.* presents *Sentic API*, which makes use of a “bag-of-concepts” model, based on ontologies and semantic networks, and a “common-sense” knowledge base, with a combination of techniques: CF-IOF, the Affective-Space vector space and the “Hourglass of Emotions”, to improve on automatic extraction of semantics, sentics and sentiment from text. The authors conclude with a description of the application of Sentic API for opinion mining and sentiment analysis of patient opinion about the UK National Health Service, captured using a microblog-type feedback service.

The poster paper *Sentiment Analysis of Wimbledon Tweets*, by Sinha *et al.*, introduces novel ideas that may inspire future work on sentiment analysis on Twitter. The poster focuses on televised events where parallel annotation of video content and Twitter streams may give novel insight into the understanding of the emotional content of the events.

Micropost Classification and Extraction

Evaluating Multi-label Classification of Incident-related Tweets by A. Schulz *et al.* addresses the problem of assigning multiple labels to tweets, where such tweets are related to incidents that have occurred. The approach uses dependencies between labels to boost the performance of a multi-label classifier trained on specific label sequences. Schulz *et al.* demonstrate a good level of performance using this approach, tested on identification and classification of data concerning incidents and emergencies, with an exact-match percentage of 84.35%.

In *Combining Named Entity Recognition Methods for Concept Extraction in Microposts*, Dlugolinsky *et al.* present an approach for combining multiple named entity recognisers together. The authors demonstrate the improved performance that can be achieved, in particular in relation to recall, when using multiple, combined recognisers. We are pleased to report that Dlugolinsky *et al.* make use of the #MSM2013 Concept Extraction Challenge data¹, and reference their own and other contributions to the challenge.

Bellaachia & Al-Dhelaan, in *HG-RANK: A Hypergraph-based Keyphrase Extraction for Short Documents in Dynamic Genre*, propose an approach for extracting keyphrases from Microposts, by modeling the information as a hypergraph. The authors use a random walk approach to rank key phrases, and using the Opinosis dataset, containing Micropost-length product reviews, demonstrate the superiority of their approach with regard to state of the art baselines.

Named Entity Extraction & Linking (NEEL) Challenge

The workshop has over the years highlighted novel research directions while improving the analysis and reuse of Microposts using approaches in Information Extraction, Data Mining, Information Visualisation, Social Studies and other relevant areas. Each of these tackles these challenges from different perspectives, using a variety of state of the art and novel techniques. At the same time, the contributions to *Making Sense of Microposts* highlight the challenges still faced in research and applications using Micropost data. To respond to this challenge, #Microposts2014 hosted a *Named Entity Extraction & Linking (NEEL) Challenge*. While this and the first (*Context Extraction*) challenge, in #MSM2013, directly targeted only a sub-set of the Microposts and Social Web community, the dataset in each may be reused for other purposes, beyond information extraction and data mining. We aim to extend the challenge in the future to widen inclusion.

The #Microposts2014 NEEL challenge attracted good interest from the community, with 43 intents to submit, out of which 24 applied for a copy of the dataset, and 8 completed submission. Of these 4 were accepted, and a further 2 as posters. All challenge submissions also took part in the workshop’s poster session, whose aim is to exhibit practical application in the field, and foster further discussion about the ways in which knowledge content is extracted from Microposts and reused.

The NEEL challenge was chaired by A. Elizabeth Cano and Giuseppe Rizzo, with Andrea Varga as dataset chair. Many thanks to those who helped with the annotation of the training dataset – we name these contributors in the challenge summary paper.

¹The proceedings of the 2013 ‘Making Sense of Microposts’ (#MSM2013) Concept Extraction Challenge are available at: <http://ceur-ws.org/Vol-1019>

We provide a brief introduction to the challenge submissions here, and more detail about the evaluation process in the challenge summary paper included in the proceedings.

Chang *et al.*, who submitted the run with the highest F_1 score, in *E2E: An End-to-end Entity Linking System for Short and Noisy Text*, present a novel approach to the NEEL task. They jointly optimised the recognition and disambiguation tasks. Based on the local and global contexts of an entity mention they generated a set of surface candidates using normalised entity lexicons, and applied overlap resolution techniques to recognise and disambiguate entity mentions.

Habib *et al.*, in *Named Entity Extraction and Linking Challenge: University of Twente at #Microposts2014*, present a sequential approach to the NEEL task by first extracting entities and then disambiguating them into a DBpedia link. They make use of state of the art features for identifying named entity (NE) candidates, including the use of Tweet segments and regular expressions. Habib *et al.* followed an NE dictionary approach for matching for the named entity linking step.

In the submission *DataTXT at #Microposts2014 Challenge*, Scaiella *et al.* propose an approach which builds on their TAGME system. They train their disambiguation algorithm with the NEEL challenge dataset. The approach in Scaiella *et al.* relies strongly on Wikipedia features for extraction and disambiguation. Their final entity linking step integrates Wikipedia categories and DBpedia RDF types as features for deploying a C4.5 classifier. *DataTXT* assigns a confidence score to each entity annotation, and discards those that fall below a specified threshold.

Yosef *et al.*, in the submission *Adapting AIDA*, extend an existing tool for entity disambiguation, AIDA. AIDA extracts entity mentions from natural language text and maps these mentions to canonical entities appearing in YAGO. To cater for Micropost content they normalise abbreviations appearing as entity mentions and supporting entity mentions appearing as username and/or hashtags. Yosef *et al.* employ a graph-based approach for linking entities, which uses different similarity measures for weighting mention-entity edges.

Bansal *et al.*, in *Linking Entities in #Microposts*, present a sequential approach to NEEL (comprising NEE + NEL). They make use of existing off-the-shelf-tools for Named Entity Extraction (NEE), and also introduce novel features for entity linking. The latter rely on the recent popularity of an entity mention on Twitter. Along with other state-of-the-art features based on Wikipedia, they applied a LambdaMART approach for the final entity disambiguation and linking step.

Finally, in *Part-of-Speech is (almost) enough: SAP Research & Innovation at the #Microposts2014 NEEL Challenge*, Dahlmeier *et al.*, present a sequential approach which makes use of off-the-shelf-tools for both NEE and NEL. They extended these toolkits using gazetteers, and employed a series of heuristics for improving the disambiguation and linking steps.

Workshop Awards

Yorkshire Tea², manufactured by Taylor's of Harrogate, sponsored the best paper awards. Nominations were sought from the reviewers, and a final decision agreed by the Chairs, based on the nominations and review scores. The #Microposts2014 best paper award went to:

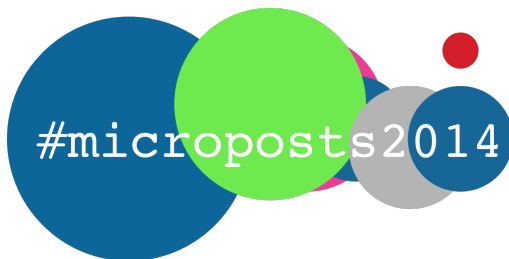
*André Panisson, Laetitia Gauvin, Marco Quaggiotto
& Ciro Cattuto*
for their submission entitled:

***Mining Concurrent Topical Activity in
Microblog Streams***

LinkedTV³ sponsored the NEEL Challenge award, an iPad, for the best submission. The challenge award was also determined by the results of the quantitative evaluation. The #Microposts NEEL Challenge award went to:

*Ming-Wei Chang, Bo-June Hsu, Hao Ma, Ricky Loynd
& Kuansan Wang*
for their submission entitled:

***E2E: An End-to-end Entity Linking
System for Short and Noisy Text***



Additional Material

The call for participation and all paper, poster and challenge abstracts are available on the #Microposts2014 website⁴. The full proceedings are also available on the CEUR-WS server, as Vol-1141⁵. The gold standard for the NEEL Challenge is available for download⁶.

The proceedings for the #MSM2013 main track are available as part of the WWW'13 Proceedings Companion⁷. The #MSM2013 Concept Extraction Challenge proceedings are published as a separate volume as CEUR Vol-1019⁸, and the gold standard is available for download⁹. The proceedings for #MSM2012 and #MSM2011 are available as CEUR Vol-838¹⁰, and CEUR Vol-718¹¹, respectively.



²<http://www.yorkshiretea.co.uk>

³<http://www.linkedtv.eu>

⁴<http://www.scc.lancs.ac.uk/microposts2014>

⁵#Microposts2014 Proc.:

<http://ceur-ws.org/Vol-1141>

⁶http://ceur-ws.org/Vol-1141/microposts2014-neel_challenge_gs.zip

⁷WWW'13 Companion:

<http://dl.acm.org/citation.cfm?id=2487788>

⁸#MSM2013 CE Challenge Proc.:

<http://ceur-ws.org/Vol-1019>

⁹http://ceur-ws.org/Vol-1019/msm2013-ce_challenge_gs.zip

¹⁰#MSM2012 Proc.: <http://ceur-ws.org/Vol-838>

¹¹#MSM2011 Proc.: <http://ceur-ws.org/Vol-718>

Main Track Programme Committee

Gholam R. Amin Sultan Qaboos University, Oman
Pierpaolo Basile University of Bari, Italy
Julie Birkholz Vrije University, The Netherlands
Uldis Bojars SIOC Project, Latvia
John Breslin NUIG, Ireland
A. Elizabeth Cano KMi, The Open University, UK
Marco A. Casanova Pontifícia Universidade Católica do Rio de Janeiro, Brazil
Óscar Corcho Universidad Politécnica de Madrid, Spain
Ali Emrouznejad Aston Business School, UK
Guillaume Erétéo Orange Labs, 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
Robert Jäschke University of Kassel, Germany
Jelena Jovanovic University of Belgrade, Serbia
Mathieu Lacage Alcméon, France
Vita Lanfranchi The University of Sheffield, UK
Philippe Laublet Université Paris-Sorbonne, France
João Magalhães Universidade Nova de Lisboa, Portugal
Diana Maynard The University of Sheffield, UK
José M. Morales del Castillo El Colegio de México, Mexico
Fabrizio Orlandi DERI, Ireland
Alexandre Passant seevl.net / MDG Web, Ireland
Bernardo Pereira Nunes Pontifícia Universidade Católica do Rio de Janeiro, Brazil
Danica Radovanovic University of Belgrade, Serbia
Yves Raimond BBC, UK
Giuseppe Rizzo Università di Torino, Italy
Harald Sack University of Potsdam, Germany
Bernhard Schandl Gnowsis.com, Austria
Sean W. M. Siqueira Universidade Federal do Estado do Rio de Janeiro, Brazil
Victoria Uren Aston Business School, UK
Andrea Varga The University of Sheffield, UK
Shenghui Wang Vrije University, The Netherlands
Katrin Weller GESIS Leibniz Institute for the Social Sciences, Germany
Ziqi Zhang The University of Sheffield, UK

Sub Reviewers

Pavan Kapanipathi Knoesis Center, Wright State University, USA
Flavio Martins Universidade Nova de Lisboa, Portugal
Filipa Peleja Universidade Nova de Lisboa, Portugal
Víctor Rodríguez Doncel Universidad Politécnica de Madrid, Spain
Nadine Steinmetz Hasso-Plattner Institute, Germany

NEEL Challenge Evaluation Committee

Ebrahim Bagheri Ryerson University, Canada
Pierpaolo Basile University of Bari, Italy
Óscar Corcho Universidad Politécnica de Madrid, Spain
Leon Derczynski The University of Sheffield, UK
Guillaume Erétéo Orange Labs, France
Miriam Fernandez KMi, The Open University, UK
Andrés Garcia-Silva Universidad Politécnica de Madrid, Spain
Anna Lisa Gentile The University of Sheffield, UK
Robert Jäschke University of Kassel, Germany
Diana Maynard The University of Sheffield, UK
José M. Morales del Castillo El Colegio de México, Mexico
Georgios Paltoglou University of Wolverhampton, UK
Bernardo Pereira Nunes Pontifícia Universidade Católica do Rio de Janeiro, Brazil
Daniel Preotjuc-Pietro The University of Sheffield, UK
Irina Temnikova Bulgarian Academy of Sciences, Bulgaria
Raphaël Troncy Eurecom, France
Victoria Uren Aston Business School, UK