Preface

The 3rd Workshop on *Making Sense of Microposts (#MSM2013)* was held in Rio de Janeiro, Brazil, on the 13th of May 2013, as part of the 22nd International Conference on the World Wide Web (WWW'13). #MSM2013 is the third in a series of successful workshops. The #MSM workshop was first held at the 8th Extended Semantic Web Conference (ESWC 2011), and with approximately 50 participants, was the most popular workshop at ESWC 2011. The second workshop was held at the 21st International Conference on the World Wide Web (WWW'12), and had approximately 60 participants, as did this year's workshop.

The #MSM series of workshops is unique in targeting both Semantic Web researchers and other fields, 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. The workshop also encourages the demonstration of the generation and use of Microposts through different physical and online media, as well as application of research, and re-use of Micropost data in real-world scenarios. Continuing to hold the workshop at WWW allows us to reach a wider and more varied audience and target research and applications at the leading edge of technology. The 2013 edition was an occasion to expand our community, and with the conference in Rio de Janeiro, to connect with local researchers from Brazil and South America, opening the way for new synergy and interesting discussions within the local cultural context.

In a world where more and more data is becoming available to machines, questions related to the use of this data for increasing machine intelligence naturally arise. Big Data treatment efforts exploit masses of data using statistical approaches in order to conceive anticipatory systems able to predict future human behaviour and adapt to it. Semantic analysis of Web content, including Microposts, is another complementary perspective to the goal of making machines more intelligent and more capable of supporting daily human activity, decision making and communication. We are seeing a very large increase in systems relying on Semantic Web technologies being deployed: Intelligent Assistants, such as Siri[1], rely on Semantic Data Graphs to provide users with factual responses to their questions. Facebook Graph Search[2] allows users to formulate complex queries over a socio-semantic graph constructed from people's *likes* and structural knowledge about things being *liked*. While static knowledge bases are largely employed in such systems, exploiting the dynamic, evolving knowledge that resides in the growing masses of Microposts, invaluable as they are acknowledged to be, remains a major challenge.

Each year we make a little step toward resolving this challenge, due largely to what makes publishing via Microposts so popular – their brevity, and as a

---

[1] http://www.apple.com/ios/siri
[2] https://www.facebook.com/about/graphsearch

---

result, the use of non-standard abbreviations, informal language and grammar. With each workshop we have found that our research community continues to open exciting new possibilities for constructing increasingly intelligent and useful services.

New to this year's workshop is the **Concept Extraction Challenge**, sponsored by eBay. Existing concept extraction tools are intended for use over news corpora and similar document-based corpora with relatively long length. The aim of the challenge was to foster research into novel, more accurate concept extraction for (much shorter) Micropost data. The keen interest in concept extraction that is shared by our community motivated this challenge, focused for this first time on a rather general task. The interest shown in the challenge by both academia and industry has confirmed its relevance. We aim to pursue the challenge in the future editions of #MSM, and are investigating new challenge tasks and the use of different collections of data, prompted by the challenge results and further research it continues to foster.

This first run of the challenge has been a learning curve, with contributions from participants, not just in their formal submissions, but also to corrections in the training data that fed into the cycle of updates that resulted in the final gold standard. The #MSM2013 Concept Extraction Challenge received 22 complete submissions, out of which 6 were accepted for presentation at the workshop, and a further 7 for presentation as posters. Submissions came from institutions across 12 countries, with 13% of submitting authors from Brazilian institutions.

Many hearty thanks to all our contributors and participants, and also the Programme Committees whose valued feedback resulted in a rich collection of work, each of which adds to the state of the art in leading edge research in the challenging task of information extraction from Microposts. Especial thanks to Andrea Varga, who was largely responsible for generating the challenge dataset, and Pablo Mendes who gave us very useful suggestions on collaborative annotation of the data. 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.

Amparo E. Cano   KMi, The Open University, UK
Matthew Rowe   Lancaster University, UK
Milan Stankovic   Université Paris-Sorbonne, France
Aba-Sah Dadzie   The University of Sheffield, UK
#MSM2013 Concept Extraction Challenge Organising Committee, May '13

Summary of Other Contributions to #MSM2013

Published with ACM as a companion volume to the WWW'13 proceedings, the main track[3] received 13 paper submissions, out of which 4 full and 2 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 about the ways in which data extracted from Microposts is being reused. The accepted submissions cover an array of topics, including a variety of approaches to concept extraction – again reinforcing its importance with respect to research on Microposts, among these, rule-based, machine learning and hybrid methods. Other topics covered range from research from a social science perspective, on the use of Microposts to publicise and discuss trending events and topics, and the extraction of intent, meaning and sentiment. Submissions came from 9 countries, with 29% of all authors from institutions in Brazil. Thanks to our local chair, Bernardo Pereira Nunes, who helped, among other things, to promote the workshop and challenge to local institutions.

The main track proceedings include also the keynote abstract, '*Urban*: Crowdsourcing for the good of London*'[4], presented by Daniele Quercia, of the Cambridge Networks Network at the University of Cambridge, England, UK.

The #MSM2013 award for the best paper, based on nominations by the reviewers and confirmed by the workshop chairs, was awarded to:

> *Lisa Posch, Claudia Wagner, Philipp Singer & Markus Strohmaier*
> for the paper:
> > **Meaning as Collective Use: Predicting Hashtag Semantics on Twitter**[5]

Introduction to the #MSM2013 Challenge Proceedings

This volume includes first a challenge report, with a summary of the state of the art and a comparison of the performance of the approaches taken for the 13 submissions accepted. This provides an overview of the capability of the state of the art in Concept Extraction approaches to date. This introductory paper details the challenge objectives and task, and the dataset construction and validation processes. We also provide a comprehensive description of the

---

[3] #MSM2013 welcome: `http://dl.acm.org/citation.cfm?id=2490000.2487998`
[4] #MSM2013 keynote: `http://dl.acm.org/citation.cfm?id=2487788.2488000`
[5] Best paper, main track: `http://dl.acm.org/citation.cfm?id=2487788.2488008`

quantitative evaluation methodology followed and the performance and ranking metrics used.

Participants' descriptions of the systems implemented complete the proceedings. Each submission was peer reviewed, to provide the authors with feedback on their approach and to identify interesting and promising work to present at the workshop. The quantitative evaluation described in the report was also carried out to rank submission runs – this was the final criterion, with a cut-off for acceptance, and the key measure for the challenge award.

Concept Extraction Challenge Award

eBay[6] sponsored the challenge award: US$ 1,500, for the best submission. Nominations were sought from the reviewers, and a final decision agreed by the challenge chairs, based on their nominations, review scores and the results of the quantitative evaluation. The #MSM2013 Concept Extraction Challenge Award went to:

Mena Habib, Maurice Van Keulen & Zhemin Zhu
for their submission entitled:
University of Twente at #MSM2013

---

[6] `http://www.ebayinc.com`

Challenge Evaluation Committee

Naren Chittar eBay, USA
Óscar Corcho Universidad Politécnica de Madrid, Spain
Danica Damljanovic Kuato Studios, London, UK
Anna Lisa Gentile The University of Sheffield, UK
Diana Maynard The University of Sheffield, UK
Peter Mika Yahoo! Research, Spain
Enrico Motta KMi, The Open University, UK
Daniel Preotiuc The University of Sheffield, UK
Alan Ritter University of Washington, USA
Guiseppe Rizzo Eurecom, France
Raphaël Troncy Eurecom, France
Victoria Uren Aston Business School, UK
Andrea Varga The University of Sheffield, UK

Additional Material

The call for participation and all challenge abstracts, in addition to those for the main workshop track, are available on the #MSM2013 website[7]. The full challenge proceedings are also available on the CEUR-WS server, as Vol-1019[8]. The proceedings for the main track are available as part of the WWW'13 Proceedings Companion[9]. The proceedings for the 1st and 2nd workshops are available as CEUR Vol-718[10] and Vol-838[11] respectively.

---

[7] Challenge web pages: `http://oak.dcs.shef.ac.uk/msm2013/ie_challenge.html`
[8] #MSM2013 Challenge proceedings: `http://ceur-ws.org/Vol-1019`
[9] WWW'13 Proceedings Companion: `http://dl.acm.org/citation.cfm?id=2487788`
[10] #MSM2011 proceedings: `http://ceur-ws.org/Vol-718`
[11] #MSM2012 proceedings: `http://ceur-ws.org/Vol-838`