EACL 2014

14th Conference of the European Chapter of the
Association for Computational Linguistics

Proceedings of the 5th Workshop on Language Analysis for
Social Media (LASM)

April 26-30, 2014
Gothenburg, Sweden

Introduction

These proceedings contain the papers presented at the 5th Workshop on Language Analysis in Social Media (LASM 2014). The workshop is held in Gothenburg, Sweden, on April 26–30, 2014, and hosted in conjunction with the 14th Conference of the European Chapter of the Association for Computational Linguistics.

Over the past few years, online social networking sites (Facebook, Twitter, Youtube, Flickr, MySpace, LinkedIn, Metacafe, Vimeo, etc.) have revolutionized the way we communicate with individuals, groups and communities, and altered everyday practices. The unprecedented volume and variety of user-generated content as well as the user interaction network constitute new opportunities for understanding social behavior and building socially-intelligent systems.

This 5th workshop attracted several submissions from around the world. Each paper was assigned to four reviewers. For the final workshop program, and for inclusion in these proceedings, nine regular papers were selected. The workshop program features two keynote presentations: one by Kalina Bontcheva, Senior Researcher in the Natural Language Processing Group, Department of Computer Science, University of Sheffield, and one on Industrial perspectives presented by NLP Technologies, Montreal Canada, on social media monitoring and innovative tools.

One of the goals of LASM 2014 was to reflect a wide range of different research efforts and results of language analysis with implications for fields such as natural language processing, computational linguistics, sociolinguistics and psycholinguistics. We invited original and unpublished research papers on all topics related to the analysis of language on social media, including the following topics:

- What are people talking about on social media?

- How are they expressing themselves?

- Why do they scribe?

- Natural language processing techniques for social media analysis

- How do language and social network properties interact?

- Semantic Web / Ontologies / Domain models to aid in social data understanding

- Characterizing Participants via Linguistic Analysis

- Language, Social Media and Human Behavior

This workshop would not have been possible without the hard work of many people. We would like to thank all Program Committee members and external reviewers for their effort in providing high-quality reviews in a timely manner. We thank all the authors who submitted their papers, as well as the authors whose papers were selected, for their help with preparing the final copy. We are in debt to the EACL 2014 Workshop co-Chairs. We would also like to thank our industry partners for their support and for making LASM 2014 a successful workshop; NLP Technologies, Microsoft Research and IBM Almaden.

March 2014

Atefeh Farzindar, Diana Inkpen, Michael Gamon, and Meena Nagarajan

Organizing Committee:

Atefeh Farzindar (NLP Technologies Inc. and Universite de Montreal Canada)

Diana Inkpen (University of Ottawa, Canada)

Michael Gamon (Microsoft Research, USA)

Meena Nagarajan (IBM Research, USA)

Program Committee:

Colin Cherry (NRC Canada)

Cindy Chung (University of Texas)

Munmun De Choudhury (Microsoft Research)

Jacob Eisenstein (Georgia Institute of Technology)

Jennifer Foster (Dublin City University)

Kevin Hass (Microsoft)

Guy Lapalme (Universite de Montreal)

Saif Mohammad (NRC Canada)

Smaranda Muresan (Rutgers University)

Alexander Osherenko (Humboldt-Universität zu Berlin)

Patrick Pantel (Microsoft Research)

Alan Ritter (University of Washington)

Mathieu Roche (Universite de Montpellier)

Victoria Rubin (University of Western Ontario)

Hassan Sayyadi (University of Maryland)

Valerie Shalin (Wright State)

Mike Thelwall (University of Wolverhampton)

Alessandro Valitutti (University of Helsinki)

Julien Velcin (Universite de Lyon)

Wei Xu (University of Washington)

Table of Contents

Conference Program

(9.00 am) Introductions

(9.05 am) Industrial Key Note:
Atefeh Farzindar, NLP Technologies, Montreal Canada
Industrial perspectives on social media monitoring and innovative tools

(9.15 am) Invited Key Note:
Kalina Bontcheva, Department of Computer Science, University of Sheffield
Natural Language Processing for Social Media: Are We There Yet?
(see *workshop web page* for abstract)

(10.30 am) Coffee Break

(11.00 am)

Mining Lexical Variants from Microblogs: An Unsupervised Multilingual Approach
Alejandro Mosquera and Paloma Moreda Pozo

(11.30 am)

Estimating Time to Event from Tweets Using Temporal Expressions
Ali Hürriyetoğlu, Nelleke Oostdijk and Antal van den Bosch

(12.00 pm)

Accurate Language Identification of Twitter Messages
Marco Lui and Timothy Baldwin

(12.30 pm) Lunch Break

(2.00 pm)

The (Un)Predictability of Emotional Hashtags in Twitter
Florian Kunneman, Christine Liebrecht and Antal van den Bosch

(continued)

(2.30 pm)

Finding Arguing Expressions of Divergent Viewpoints in Online Debates
Amine Trabelsi and Osmar R. Zaiane

(3.00 pm)

Aspect Term Extraction for Sentiment Analysis: New Datasets, New Evaluation Measures and an Improved Unsupervised Method
John Pavlopoulos and Ion Androutsopoulos

(3.30 pm) Coffee Break

(4.00 pm)

Vowel and Diacritic Restoration for Social Media Texts
Kübra Adali and Gülşen Eryiğit

(4.30 pm)

A Cascaded Approach for Social Media Text Normalization of Turkish
Dilara Torunoğlu and Gülşen Eryiğit

(5.00 pm)

Experiments to Improve Named Entity Recognition on Turkish Tweets
Dilek Kucuk and Ralf Steinberger

(5.30 pm) Closing Remarks