

EACL 2012

**13th Conference of the European Chapter of the  
Association for Computational Linguistics**

**Proceedings of the Workshop on  
Semantic Analysis in Social Media**

April 23 2012  
Avignon, France

© 2012 The Association for Computational Linguistics

ISBN 978-1-937284-19-0

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

## Preface

Semantic analysis in social networks (SN) is important for applications such as understanding and enabling social networks, natural language interfaces and human behaviour on the web, e-learning environments, cyber communities and educational or online shared workspaces. These aspects are also important in security, privacy and identity, opinion mining, sentiment analysis, and in the larger area of affective computing.

This workshop provides a forum for discussion between leading names and researchers involved in text analysis and social networks in the context of natural language understanding, natural language generation, automatic categorization, topic detection, emotion analysis, and applications using computational approaches to process social networks.

Topics of interest include, but are not limited to:

- semantic analysis in sentences and web content from social networks
- classification of texts by emotion and mood from SN
- sociology of emotions and influence on inter-personal communications
- topic detection and clustering in SN
- SN analysis across different languages
- SN analysis from multimedia (text, speech, video)
- security and privacy issues in SNs
- automatic summarization from multiple sources and multiple languages
- analysis of sentiment and opinion in SN
- information extraction and indexing
- applications in which affective aspects are beneficial
- tools and resources for accessing, representing, and managing social network data in natural language processing frameworks (e.g., GATE, UIMA)
- other aspects of the computational treatment of SN and affect.

The workshop covers three main perspectives: government (e.g., security and criminology), industry (e.g., marketing), and academic (e.g., theoretical research related to SNs).

We would like to thank all the authors who submitted papers for the hard work that went behind their submissions. We express our deepest gratitude to the committee members for their thorough reviews. We also thank the EACL 2012 organizers for their help with administrative matters.

Diana Inkpen  
Atefeh Farzindar



**Organizers:**

Atefeh Farzindar (NLP Technologies Inc. and Université de Montréal, Canada)  
Diana Inkpen (University of Ottawa, Canada)

**Program Committee:**

Colin Cherry (National Research Council, Canada)  
Viola Ganter (HITS gGmbH, Heidelberg, Germany)  
Mitsuru Ishizuka (University of Tokyo, Japan)  
Fazel Keshtkar (University of Memphis)  
Wael Khreich (NLP Technologies Inc, Canada)  
Guy Lapalme (Université de Montréal, Canada)  
Eduarda Mendes Rodrigues (University of Porto, Portugal)  
Alena Neviarouskaya (Toyohashi University of Technology, Japan)  
Alexander Osherenko (Socioware Development, Germany)  
Victoria Rubin (University of Western Ontario, Canada)  
Carlo Strapparava (FBK-IRST, Povo-Trento, Italy)  
Mike Thelwall (University of Wolverhampton, UK)  
Alessandro Valitutti (University of Helsinki, Finland)  
Simone Teufel (Cambridge University, UK)  
René Witte (Concordia University, Canada)

## Keynote Speech

**Title:** Industrial Perspectives on Social Networks

**Keynote speaker:** Dr. Atefeh Farzindar, NLP Technologies Inc.

### **Abstract:**

Social media data is the collection of open source information that can be obtained publicly via the Web and social networks. Social information intelligence refers to an emerging data and semantic infrastructure that will enable organizations to create a new generation of business applications. This new class of applications will build on the rich set of assets already available within the organization. Social media has become a primary source of intelligence for Security Intelligence and Business Intelligence. Social data intelligence combines social media aspects and analytics to give important business insights, and is a convergence of several trends. Business intelligence from open intelligence incorporates knowledge management, social networking, plus social media monitoring and analytics, all combined into a new interface in the business intelligence environment.

In the context of analyzing social networks, finding powerful methods and algorithms to search for relevant data in large volumes, and various free formats from multiple sources and languages is a scientific challenge. Automatic processing of such data needs to evaluate the appropriate research methods for information extraction, automatic categorization and clustering, indexing data, generating automatic summaries, and statistical machine translation. With respect to machine learning approaches, we must consider developing innovative tools and integrating appropriate linguistic information in the fields of security and defence, and industry business intelligence.

There is great interest for social media data monitoring in the industry. Social media data can dramatically improve business intelligence and help both international and local markets. Businesses could achieve several goals by integrating social data into their corporate BI systems, such as branding and awareness, customer/prospect engagement and improving customer service.

### **Biography:**

Dr. Atefeh Farzindar is the founder of NLP Technologies Inc., a company specializing in natural language processing, automatic summarization, statistical machine translation and social media solutions. Dr. Farzindar received her Ph.D. in Computer Science from the Université de Montréal and Paris-Sorbonne University. She is an adjunct professor at the Department of Computer Science at the Université de Montréal. As president of NLP Technologies, she has managed multiple collaborative R&D projects with various industry and university partners. She is the chair of the language technologies sector of the Language Industry Association Canada (AILIA) and a board member of the Language Technologies Research Centre, co-chair of the Canadian Conference on Artificial Intelligence 2010 and industry chair for Canadian AI'2011 and AI'2012.

## Invited Talk

**Title:** Mining Online Discussions: An Applications to the Analysis of News Websites

**Invited Speaker:** Dr. Julien Velcin, Associate Professor in CS, ERIC Lab - University of Lyon

### **Abstract:**

News stories websites represent an important source of data for building a picture of what is going on in the world. Apart from providing the information itself, such websites allow Internet users to post their opinions and comments in the form of discussions.

These discussions bring valuable knowledge in order to understand a person's position regarding certain news. In the first part of the talk, I will briefly present this type of data and explain how they represent a challenge. After this introduction, I will discuss the content-oriented model as well as the recommended system we have developed for analyzing key messages posted in such online discussions. In addition, I will present the approach we have followed for extracting the implicit users' network. To this end, I will focus on a model involving three types of relations, two of which are based on the citations. The resulting user network is seen as an implicit social network and I will explain how it can be used to extract celebrities. Finally, I will introduce the problem of image extraction addressed in a new project, where an image is seen as a representation of various entities populating the Internet (e.g., politicians, companies, brands etc.). In particular, I will show how this kind of project interrelates traditional text/opinion mining with social network analysis.

### **Biography:**

After his MSc graduation in Artificial Intelligence and Pattern Recognition (2002), Julien Velcin defended a PhD in Computer Science at the University of Paris 6 in 2005. He is a researcher in the ERIC Lab, University of Lyon 2 since 2007. His research interests lie mainly in the extraction, evaluation and characterization of categories in unsupervised machine learning, with applications to text mining and web analysis. His work has been published in international journals and conferences such as WIAS, IJCAI, ICCS, ISMIS, ER. He is involved in international program committees such as ECML-PKDD, ASONAM, and EGC. He is a member of the editorial board of the International Journal of Data Analysis Techniques and Strategies (IJDATS). He is an active reviewer for the Association for Computing Machinery (ACM) since 2009. He is currently project manager of a national project on the study of images and opinion evolution through the Internet, funded by the French National Research Agency.





## Table of Contents

<i>Unsupervised Part-of-Speech Tagging in Noisy and Esoteric Domains With a Syntactic-Semantic Bayesian HMM</i>	
William M. Darling, Michael J. Paul and Fei Song .....	1
<i>The Role of Emotional Stability in Twitter Conversations</i>	
Fabio Celli and Luca Rossi .....	10
<i>Towards Scalable Speech Act Recognition in Twitter: Tackling Insufficient Training Data</i>	
Renxian Zhang, Dehong Gao and Wenjie Li .....	18
<i>Topic Classification of Blog Posts Using Distant Supervision</i>	
Stephanie Husby and Denilson Barbosa .....	28
<i>A User and NLP-Assisted Strategic Workflow for a Social Semantic OWL 2-Based Knowledge Platform</i>	
Jinan El-Hachem and Volker Haarslev .....	37
<i>A Hybrid Framework for Scalable Opinion Mining in Social Media: Detecting Polarities and Attitude Targets</i>	
Carlos Rodriguez-Penagos, Jens Grivolla and Joan Codina-Filba .....	46
<i>Predicting the 2011 Dutch Senate Election Results with Twitter</i>	
Erik Tjong Kim Sang and Johan Bos .....	53
<i>Opinion and Suggestion Analysis for Expert Recommendations</i>	
Anna Stavrianou and Caroline Brun .....	61



# Conference Program

## Monday, April 23, 2012

- 9:00–9:10 Opening Remarks by Diana Inkpen, University of Ottawa
- 9:10–9:30 Keynote speech: Industrial Perspectives on Social Networks by Atefeh Farzindar, NLP Technologies
- 9:30–10:00 *Unsupervised Part-of-Speech Tagging in Noisy and Esoteric Domains With a Syntactic-Semantic Bayesian HMM*  
William M. Darling, Michael J. Paul and Fei Song
- 10:00–10:30 Coffee break
- 10:30–11:00 *The Role of Emotional Stability in Twitter Conversations*  
Fabio Celli and Luca Rossi
- 11:00–11:30 *Towards Scalable Speech Act Recognition in Twitter: Tackling Insufficient Training Data*  
Renxian Zhang, Dehong Gao and Wenjie Li
- 11:30–12:00 *Topic Classification of Blog Posts Using Distant Supervision*  
Stephanie Husby and Denilson Barbosa
- 12:00–12:30 *A User and NLP-Assisted Strategic Workflow for a Social Semantic OWL 2-Based Knowledge Platform*  
Jinan El-Hachem and Volker Haarslev
- 12:30–14:00 Lunch break
- 14:00–15:00 Invited talk: Mining Online Discussions: an Application to the Analysis of News Websites by Julien Velcin, University Lyon 2
- 15:00–15:30 *A Hybrid Framework for Scalable Opinion Mining in Social Media: Detecting Polarities and Attitude Targets*  
Carlos Rodriguez-Penagos, Jens Grivolla and Joan Codina-Filba
- 15:30–16:00 Coffee break
- 16:00–16:30 *Predicting the 2011 Dutch Senate Election Results with Twitter*  
Erik Tjong Kim Sang and Johan Bos
- 16:30–17:00 *Opinion and Suggestion Analysis for Expert Recommendations*  
Anna Stavrianou and Caroline Brun

**Monday, April 23, 2012 (continued)**

17:00-17:30 Closing Remarks and Discussion