

Semantic Web and Information Extraction

SWAIE 2012

Workshop in conjunction with the  
18th International Conference on  
Knowledge Engineering and Knowledge Management  
Galway City, Ireland, 9 October 2012

Edited by:  
Diana Maynard  
Marieke van Erp  
Brian Davis



# Preface

There is a vast wealth of information available in textual format that the Semantic Web cannot yet tap into: 80% of data on the Web and on internal corporate intranets is unstructured, hence analysing and structuring the data - social analytics and next generation analytics - is a large and growing endeavour. The goal of the 1st workshop on Semantic Web and Information Extraction was to bring researchers from the fields of Information Extraction and the Semantic Web together to foster inter-domain collaboration. To make sense of the large amounts of textual data now available, we need help from both the Information Extraction and Semantic Web communities. The Information Extraction community specialises in mining the nuggets of information from text: such techniques could, however, be enhanced by annotated data or domain-specific resources. The Semantic Web community has already taken great strides in making these resources available through the Linked Open Data cloud, which are now ready for uptake by the Information Extraction community. The workshop invited contributions around three particular topics: 1) Semantic Web-driven Information Extraction, 2) Information Extraction for the Semantic Web, and 3) applications and architectures on the intersection of Semantic Web and Information Extraction.

SWAIE 2012 had a number of high-quality submissions. From these, the 6 best papers were chosen for the two paper sessions of the programme: 4 long paper presentations and 2 short ones. Additionally, we held a lightning talks session where attendees could present brief 3-minute talks about late-breaking work or demos around the workshop themes, and a panel session involving some general discussion about these themes. To initiate the workshop, a keynote talk was provided by D.J. McCloskey, NLP Architect in IBM's new Watson Solutions division. The keynote presented the post-Watson role of Information Extraction and its intersection with the Semantic Web.

We would like to thank the many people who helped make SWAIE 2012 such a success: the Programme Committee, the paper contributors, the invited speaker and panellists, and all the participants present at the workshop who engaged in lively debate.

**Diana Maynard**, University of Sheffield  
**Marieke van Erp**, VU University Amsterdam  
**Brian Davis**, DERI Galway

## Programme Committee

- Georgeta Bordea, DERI Galway, Ireland
- Matje van de Camp, Tilburg University, The Netherlands
- Christian Chiarcos, Information Sciences Institute, USA
- Hamish Cunningham, University of Sheffield, UK
- Thierry DeClerck, DFKI, Germany
- Robert Engels, Western Norwegian Research Institute, Norway
- Phil Gooch, City University London, UK
- Seth Grimes, Alta Plana Corporation
- Siegfried Handschuh, DERI, Ireland
- Dirk Hovy, Information Sciences Institute, USA
- Laurette Pretorius, University of South Africa, South Africa
- Birgit Proell, Johannes Kepler University of Linz, Austria
- Giuseppe Rizzo, EURECOM, France
- Piek Vossen, VU University, The Netherlands
- Marie Wallace, IBM Dublin, Ireland
- René Witte, Concordia University, Montreal, Canada

# Contents

Unsupervised Improvement of Named Entity Extraction in Short Informal Context Using Disambiguation Clues <i>Mena B. Habib and Maurice van Keulen . . . . .</i>	1
LODIE: Linked Open Data for Web-scale Information Extraction <i>Fabio Ciravegna, Anna Lisa Gentile and Ziqi Zhang . . .</i>	11
Ontologies as a Source for the Automatic Generation of Grammars for Information Extraction Systems <i>Thierry Declerck and Paul Buitelaar . . . . .</i>	23
Identifying Consumers' Arguments in Text <i>Jodi Schneider and Adam Wyner . . . . .</i>	31
Identifying and Extracting Quantitative Data in Annotated Text <i>Don J. M. Willems, Hajo Rijgersberg and Jan L. Top . .</i>	43
Scenario-Driven Selection and Exploitation of Semantic Data for Optimal Named Entity Disambiguation <i>Panos Alexopoulos, Carlos Ruiz and José-Manuel Gómez-Pérez . . . . .</i>	55