SWAIE 2014

The Third Workshop on Semantic Web and Information Extraction

Proceedings of the Workshop

August 24, 2014
Dublin, Ireland

© 2014 The Authors
The papers in this volume are licensed by the authors under a
Creative Commons Attribution 4.0 International License.

ISBN 978-1-873769-48-5
Proceedings of the Third Workshop on Semantic Web and Information Extraction
(SWAIE 2014)
Diana Maynard, Marieke van Erp and Brian Davis (eds.)

Introduction

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. Here, the Information Extraction community could help as they specialise in mining the nuggets of information from text. Information Extraction techniques could be enhanced by annotated data or domain-specific resources. The Semantic Web community has 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. Following the previous two SWAIE workshops at EKAW 2012 and RANLP 2013 respectively, we have focused our attention on fostering awareness of how Semantic Web technologies can benefit the traditional IE and NLP communities.

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 2014 had a number of high-quality submissions. From these, 6 high quality papers were selected.

Two keynote speakers were invited to the workshop. The first talk was provided by D.J. McCloskey, NLP Architect in IBM's Watson Solutions division. The keynote presented the post-Watson role of Information Extraction and its intersection with the Multilingual Semantic Web. The second talk was provided by Lorraine Goeuriot, CNGL, DCU on Ontology Based IE for the medical domain. We would like to thank the many people who helped make SWAIE 2014 such a success: the Programme Committee, the paper contributors, the invited speakers 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, INSIGHT@NUI Galway

Organizers:

Diana Maynard, University of Sheffield
Marieke van Erp, VU University Amsterdam
Brian Davis, INSIGHT@NUI Galway

Program Committee:

Eneko Agirre, University of the Basque Country, Spain
Paul Buitelaar, INSIGHT@ NUIGalway, Ireland
Elena Cabrio, INRIA Sophia Antipolis Méditerannée, France
Philipp Cimiano,CITEC University of Bielefeld, Germany
Hamish Cunningham, University of Sheffield, UK
Mariana Damova, Mozajka, Bulgaria
Dana Dannélls, University of Gothenburg, Sweden
Thierry DeClerck, DFKI, Germany
Antske Fokkens, VU University Amsterdam, the Netherlands
Jens Grivolla, Universitat Pompeu Fabra, Spain
Dirk Hovy, University of Copenhagen, Denmark
Phil Gooch, City University London, UK
Siegfried Handschuh, INSIGHT@ NUIGalway, Ireland
John Judge, CNGL Dublin City University, Ireland
David Lewis, CNGL Trinity College Dublin, Ireland
Alex O'Conor, CNGL Trinity College Dublin, Ireland
Laurette Pretorius, University of South Africa, South Africa
Birgit Proell, Johannes Kepler University Linz, Austria
German Rigau, University of the Basque Country, Spain
Marco Rospocher, Fondazione Bruno Kessler, Italy
Sara Tonelli,Fondazione Bruno Kessler, Italy

Invited Speaker:

D.J. McCloskey, IBM Watson Solutions, Dublin
Lorraine Goeuriot, CNGL, Dublin

Table of Contents

Corpus-based Translation of Ontologies for Improved Multilingual Semantic Annotation
Claudia Bretschneider, Heiner Oberkampf, Sonja Zillner, Bernhard Bauer and Matthias Hammon 1

Information Extraction for Social Media
Mena Badieh Habib Morgan and Maurice van Keulen 9

Seed Selection for Distantly Supervised Web-Based Relation Extraction
Isabelle Augenstein 17

Ontology-based Extraction of Structured Information from Publications on Preclinical Experiments for Spinal Cord Injury Treatments
Benjamin Paassen, Andreas Stöckel, Raphael Dickfelder, Jan Philip Göpfert, Nicole Brazda, Tarek Kirchhoffer, Hans Werner Müller, Roman Klinger, Matthias Hartung and Philipp Cimiano 25

Semi-supervised Sequence Labeling for Named Entity Extraction based on Tri-Training: Case Study on Chinese Person Name Extraction
Chien-Lung Chou, Chia-Hui Chang and Shin-Yi Wu 33

Towards a robust framework for the semantic representation of temporal expressions in cultural legacy data
Daniel Isemann, Gerard Lynch and Raffaella Lanino 41

Workshop Program

Corpus-based Translation of Ontologies for Improved Multilingual Semantic Annotation
Claudia Bretschneider, Heiner Oberkampf, Sonja Zillner, Bernhard Bauer and Matthias Hammon

Information Extraction for Social Media
Mena Badieh Habib Morgan and Maurice van Keulen

Seed Selection for Distantly Supervised Web-Based Relation Extraction
Isabelle Augenstein

Ontology-based Extraction of Structured Information from Publications on Preclinical Experiments for Spinal Cord Injury Treatments
Benjamin Paassen, Andreas Stöckel, Raphael Dickfelder, Jan Philip Göpfert, Nicole Brazda, Tarek Kirchhoffer, Hans Werner Müller, Roman Klinger, Matthias Hartung and Philipp Cimiano

Semi-supervised Sequence Labeling for Named Entity Extraction based on Tri-Training: Case Study on Chinese Person Name Extraction
Chien-Lung Chou, Chia-Hui Chang and Shin-Yi Wu

Towards a robust framework for the semantic representation of temporal expressions in cultural legacy data
Daniel Isemann, Gerard Lynch and Raffaella Lanino