RANLP 2015

Second Workshop on Natural Language Processing and Linked Open Data

(NLP&LOD2)

Proceedings of the Workshop

September 11, 2015
Hissar, Bulgaria

Second Workshop on Natural Language Processing and Linked Open Data
Associated with the International Conference
Recent Advances in
Natural Language Processing'2015

PROCEEDINGS

Hissar, Bulgaria
11 September 2015

ISBN 978-954-452-035-9

Designed and Printed by INCOMA Ltd.
Shoumen, BULGARIA

Introduction

NLP started to use extensively LOD in various scenarios, such as: exploring knowledge datasets (DBPedia, FreeBase, GeoNames, etc.) for annotation and information extraction; publishing language resources as LOD (WordNet, FrameNet, etc.); aggregating of the available data for various tasks (BabelNet, Global WordNet Grid); creation of standards for LOD (LEMON); building ontologies for different domains.

At the same time, the NLP processing pipelines have been developed towards the recognition and extraction of entities and events from raw stream data. Handling of events, however, requires also the inclusion of high quality modules like NER, NED, Semantic Role Labeling (SRL), sense and valency annotation. These modules rely not only on canonical resources, but also on the LOD datasets for extracting information about people, facts, and organizations. Additionally NLP techniques are used for creation of LOD datasets on the basis of new textual information.

Since there is some experience gained now in the interaction between NLP and LOD as well as between LOD and NLP, some problems have been identified, too. These are: general failure of NLP technology to meet completely the requirements of LOD; incompleteness of LOD datasets; sparseness of LOD datasets through various languages and domains; lack of robust reasoning mechanisms in NLP and LOD; still inefficient handling of natural language non-literal phenomena, such as metonymy, polysemy, figurative expressions; usability and re-usability of NLP and LOD applications.

Thus, a number of issues are related to the interaction between NLP and LOD. These are: reasons for low precision and inconsistencies; enhancing NLP applications with LOD; information extraction from LOD using NLP techniques; manipulating LOD (cleaning, adding information, deleting information, reconstructing facts) with NLP techniques; LOD as a corpus; mapping LOD to common sense ontologies and language data; storing LOD in RDF bases; methodological and theoretical approaches to LOD; handling polysemy and metonymy of entities in LOD; incompleteness of LOD data; LOD as unbalanced data through countries, cultures and topics of interest; insufficient reasoning in NLP and LOD; dynamics of LOD and NLP: versioning, replication, provenance, etc.

There are 6 papers accepted at the workshop. They cover the following topics: resources for Linked Open Data; representation of language phenomena in Linked Open Data; unified access to Linked Open Data and an ontology-based POS tagger using unifications of different tagsets.

We wish you a pleasant reading!

The Organizers

Organizers:

Piek Vossen, VU University Amsterdam
German Rigau, University of the Basque Country
Petya Osenova, Sofia University, Bulgaria
Kiril Simov, Bulgarian Academy of Sciences, Bulgaria

Program Committee:

Eneko Agirre, University of the Basque Country, Spain
Kalina Boncheva, Sheffield University, UK
António Branco, University of Lisbon, Portugal
Aljoscha Burchardt, DFKI, Germany
Nicoletta Calzolari, Istituto di Linguistica Computazionale, Italy
Philipp Cimiano, University of Bielefeld, Germany
Christian Chiarcos, Goethe-University Frankfurt am Main, Germany
Thierry Declerck, DFKI, Germany
Markus Egg, Humboldt-University Berlin, Germany
Jan Hajič, Charles University in Prague, Czech Republic
John McCrae, University of Bielefeld, Germany
Petya Osenova, Sofia University and IICT-BAS, Bulgaria
Maciej Piasecki, Wroclaw University of Technology, Poland
German Rigau, University of the Basque Country, Spain
Felix Sasaki, DFKI, Germany
Kiril Simov, IICT-BAS, Bulgaria
Gertjan van Noord, University of Groningen, The Netherlands
Piek Vossen, Vrije Universiteit Amsterdam, The Netherlands

Invited Speaker:

German Rigau, University of the Basque Country

Table of Contents

Workshop Program

09:00–09:30 *Welcome*

Invited lecture

09:30–10:30 *Cross-lingual Event Detection in Discourse*
German Rigau

Session 1:

10:30–11:00 *Generating Lexicalization Patterns for Linked Open Data*
Rivindu Perera and Parma Nand

11:00–11:30 **Coffee break**

Session 2:

11:30–12:00 *Small in Size, Big in Precision: A Case for Using Language-Specific Lexical Resources for Word Sense Disambiguation*
Steven Neale, João Silva and António Branco

12:00–12:30 *Towards the Representation of Hashtags in Linguistic Linked Open Data Format*
Thierry Declerck and Piroska Lendvai

12:30–14:00 **Lunch break**

Session 3:

14:00–14:30 *An Ontology-based Approach To Automatic Part-of-Speech Tagging Using Heterogeneously Annotated Corpora*
Maria Sukhareva and Christian Chiarcos

14:30–15:00 *Accessing Linked Open Data via A Common Ontology*
Kiril Simov and Atanas Kiryakov

15:00–15:30 *The GuanXi network: a new multilingual LLOD for Language Learning applications*
Ismail El Maarouf, Hatem Mousselly Sergieh, Eugene Alferov, Haofen Wang, Zhijia Fang and Doug Cooper

15:30–16:00 **Discussion and Closing**