2nd Workshop on "The People's Web meets NLP: Collaboratively Constructed Semantic Resources", Beijing, China

Submitted by witte on Wed, 2010-05-12 06:38

- 2010
- beijing
- ccsr
- china
- coling
- workshop

Start: 2010-08-28
Timezone: Asia/Hong Kong

COLING 2010

2nd Workshop on
"The People's Web meets NLP:
Collaboratively Constructed Semantic Resources"

Beijing
August 28th, 2010
http://www.ukp.tu-darmstadt.de/scientific-community/coling-2010-workshop/

Keywords:
Wikipedia, Wiktionary, Mechanical Turk, Games with a purpose,
Folksonomies, Twitter, Social Networks

INTRODUCTION

The workshop builds upon the success of the first ACL "The People's Web meets NLP" Workshop in 2009 that attracted 21 submissions. Accepted submissions included papers on Wikipedia [1], Wiktionary [2], Mechanical Turk [3], and game-based construction of semantic resources [4]. This clearly demonstrates a substantial and growing interest of the NLP community in collaboratively constructed semantic resources (CSRs), also evidenced by the increasing number of publications in this area and the EMNLP 2009 Web 2.0 track. In many works, CSRs have been used to overcome the knowledge acquisition bottleneck and coverage problems pertinent to conventional lexical semantic resources. The greatest popularity in this respect can so far certainly be attributed to Wikipedia [1]. However, other resources, such as folksonomies or the multilingual collaboratively constructed dictionary Wiktionary, have also shown great potential. Thus, the scope of the workshop deliberately includes any collaboratively constructed resource, not only Wikipedia.

Effective deployment of CSRs to enhance NLP introduces a pressing need to address a set of fundamental challenges, e.g. the interoperability with existing resources, or the quality of the extracted lexical semantic knowledge. Interoperability between resources is crucial as no single resource provides perfect coverage. The quality of CSRs is

a fundamental issue, as they lack editorial control and entries are often incomplete. Thus, techniques for link prediction [5] or information extraction [6] have been proposed to guide the "crowds" while constructing resources of better quality.

[1] Olena Medelyan, David Milne, Catherine Legg and Ian H. Witten.
Mining meaning from Wikipedia.
In: International Journal of Human-Computer Studies. 67(9), 2009.
[2] Torsten Zesch, Christof Mueller and Iryna Gurevych
Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary
Proceedings of the Conference on Language Resources and Evaluation (LREC), 2008.
http://www.ukp.tu-darmstadt.de/software/jwpl/
http://www.ukp.tu-darmstadt.de/software/jwktl/
[3] Rion Snow, Brendan O'Connor, Daniel Jurafsky and Andrew Y. Ng.
Cheap and Fast---But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks.
Proceedings of EMNLP. 2008.
[4] Luis von Ahn and Laura Dabbish.
General Techniques for Designing Games with a Purpose.
Communications of the ACM, 2008.
[5] Rada Mihalcea and Andras Csomai
Wikify!: Linking Documents to Encyclopedic Knowledge.
Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, CIKM 2007.
[6] Daniel S. Weld et al.
Intelligence in Wikipedia.
Twenty-Third Conference on Artificial Intelligence (AAAI), 2008.

TOPICS

The workshop will bring together researchers from different worlds, for example those using collaboratively constructed resources as sources of lexical semantic information for NLP purposes such as information retrieval, named entity recognition, or keyword extraction, and those using NLP techniques to improve the resources or extract and analyze different types of lexical semantic information from them. We will especially welcome contributions of interdisciplinary nature, e.g. those applying discourse analysis techniques from computational linguistics to the content of CSRs to better understand their properties.

Specific topics include but are not limited to:
* Analysis of collaboratively constructed resources, such as wiki-based platforms, folksonomies, Twitter, or social networks;
* Using collaboratively constructed resources for NLP purposes such as information retrieval, text categorization, information extraction, etc.;
* Using special features of collaboratively constructed resources to create novel resource types, for example revision-based corpora, simplified versions of resources, etc.;
* Analyzing the structure of collaboratively constructed resources related to their use in NLP;
* Interoperability of collaboratively constructed resources with conventional lexical semantic resources and between themselves;
* Mining social and collaborative content for constructing structured

semantic resources and the corresponding tools;
* Mining multilingual information from collaboratively constructed resources;
* Quality and reliability of collaboratively constructed semantic resources.

We especially encourage short papers describing publicly available tools for accessing or analyzing collaboratively constructed resources that can serve as a multiplier in the NLP community.

The workshop is intended to be highly interdisciplinary. Thus, we encourage the participation of researchers working on computational linguistics aspects (e.g. parsing or discourse analysis) or NLP applications (e.g. information retrieval, information extraction, question answering, and knowledge representation) as well as researchers from other areas who might benefit from collaboratively constructed semantic resources.

Substantially extended versions of the best papers from the workshop can be submitted to a planned Special Issue in one of the major computational linguistics journals. The revised papers will have to undergo a separate reviewing process required for journal publications.

IMPORTANT DATES

Paper submission deadline (full and short): May 30, 2010
Notification of acceptance of papers: June 30, 2010
Camera-ready copy of papers due: July 10, 2010
COLING 2010 Workshop: Aug 28, 2010

ORGANIZERS

Iryna Gurevych
Torsten Zesch

Ubiquitous Knowledge Processing Lab
Technische Universität Darmstadt, Germany

PROGRAM COMMITTEE

Andras Csomai Google Inc.
Anette Frank Heidelberg University
Benno Stein Bauhaus University Weimar
Bernardo Magnini ITC-irst Trento
Christiane Fellbaum Princeton University
Dan Moldovan University of Texas at Dallas
Delphine Bernhard LIMSI-CNRS, Orsay
Diana McCarthy Lexical Computing Ltd
Elke Teich Technische Universität Darmstadt
Emily Pitler University of Pennsylvania
Eneko Agirre University of the Basque Country
Erhard Hinrichs Eberhard Karls Universität Tübingen
Ernesto De Luca Technische Universität Berlin
Florian Laws University of Stuttgart
Gerard de Melo MPI Saarbrücken
German Rigau University of the Basque Country

Graeme Hirst University of Toronto
Günter Neumman DFKI Saarbrücken
György Szarvas Technische Universität Darmstadt
Hans-Peter Zorn European Media Lab, Heidelberg
José Iria University of Sheffield
Laurent Raumary LORIA, Nancy
Magnus Sahlgren Swedish Institute of Computer Science
Manfred Stede Potsdam University
Omar Alonso A9.com, Inc.
Pablo Castells Universidad Autónonoma de Madrid
Paul Buitelaar DERI, National University of Ireland, Galway
Philipp Cimiano Delft University of Technology
Razvan Bunescu University of Texas at Austin
Rene Witte Concordia University Montréal
Roxana Girju University of Illinois at Urbana-Champaign
Saif Mohammad University of Maryland
Samer Hassan University of North Texas
Sören Auer Leipzig University
Tonio Wandmacher CEA, Paris