Coling 2010

23rd International Conference on Computational Linguistics

Proceedings of the

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

Produced by
Chinese Information Processing Society of China
All rights reserved for Coling 2010 CD production.

To order the CD of Coling 2010 and its Workshop Proceedings, please contact:

Chinese Information Processing Society of China
No.4, Southern Fourth Street
Haidian District, Beijing, 100190
China
Tel: +86-010-62562916
Fax: +86-010-62562916
`cips@iscas.ac.cn`

Introduction

This volume contains papers accepted for presentation at the 2nd Workshop on Collaboratively Constructed Semantic Resources that took place on August 28, 2010, as part of the Coling 2010 conference in Beijing. Being the second workshop on this topic, we were able to build on the success of the previous workshop on this topic held as part of ACL-IJCNLP 2009.

In many works, collaboratively constructed semantic resources 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. 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 such resources to enhance Natural Language Processing 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 collaboratively constructed semantic resources is a fundamental issue, as they lack editorial control and entries are often incomplete. Thus, techniques for link prediction or information extraction have been proposed to guide the ”crowds” while constructing resources of better quality.

We issued calls for both long and short papers. Seven long papers and one short paper were accepted for presentation, based on the careful reviews of our program committee. We would like to thank the program committee members for their thoughtful, high quality, and elaborate reviews, especially considering the tight schedule for reviewing. The call for papers attracted submissions on a wide range of topics showing that collaboratively constructed semantic resources are of growing interest in different fields of Natural Language Processing.

The workshop aimed at bringing together researchers from different worlds, for example those using collaboratively constructed resources as sources of lexical semantic information for Natural Language Processing purposes such as information retrieval, named entity recognition, or keyword extraction, and those using Natural Language Processing techniques to improve the resources or extract and analyze different types of lexical semantic information from them. Looking at the final proceedings, we can safely say that this goal has been achieved.

Iryna Gurevych and Torsten Zesch

Organizers:

Iryna Gurevych, UKP Lab, Technische Universität Darmstadt
Torsten Zesch, UKP Lab, Technische Universität Darmstadt

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 Universitt 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
Gy¨rgy Szarvas, Technische Universität Darmstadt
Hans-Peter Zorn, European Media Lab, Heidelberg
Jos Iria, University of Sheffield
Laurent Romary, LORIA, Nancy
Magnus Sahlgren, Swedish Institute of Computer Science
Manfred Stede, Potsdam University
Omar Alonso, Microsoft
Pablo Castells, Universidad Autnonoma de Madrid
Paul Buitelaar, DERI, Galway
Philipp Cimiano, Delft University of Technology
Razvan Bunescu, University of Texas at Austin
Rene Witte, Concordia University Montral
Roxana Girju, University of Illinois at Urbana-Champaign
Saif Mohammad, University of Maryland
Samer Hassan, University of North Texas
Sren Auer, Leipzig University
Tonio Wandmacher, CEA, Paris

Invited Speaker:

Tat-Seng Chua, National University of Singapore

Title: Extracting Knowledge from Community Question-Answering Sites

Abstract: Community question-answering (QA) services, like Yahoo! Answers, contain a huge amount of information in the form of QA pairs accumulated over many years. The information covers a wide variety of topics on questions of great interests to and frequently asked by the users. To make this huge amount of information accessible by general users, research has been carried out to help users find similar questions with readily available answers. However, a better approach is to organize all relevant QA pairs around a given topic into a knowledge structure to help users better understand the overall topic. To accomplish this, our research leverages on appropriate topic prototype hierarchy automatically acquired from the Web or Wikipedia to guide the organization of the un-structured user-generated-contents in community QA sites. More specifically, we propose a prototype-hierarchy based clustering algorithm that utilizes the category structure information, article contents of Wikipedia, as well as distribution of relevant QA pairs around the topic based on a multi-criterion optimization function. This talk discusses our research to transform unstructured community QA resources into knowledge structure.

Short Bio: Chua Tat-Seng the KITHC Chair Professor at the School of Computing, National University of Singapore (NUS). He was the Acting and Founding Dean of the School of Computing during 1998-2000. He joined NUS in 1983, and spent three years as a research staff member at the Institute of Systems Science (now I2R) in the late 1980s. Dr Chua's main research interest is in multimedia information retrieval, in particular, on the analysis, retrieval and question-answering (QA) of text and image/video information. He is currently working on several multi-million-dollar projects: interactive media search, local contextual search, and real-time live media search. His group participates regularly in TREC-QA and TRECVID video retrieval evaluations. Dr Chua has organized and served as program committee member of numerous international conferences in the areas of computer graphics, multimedia and text processing. He is the conference co-chair of ACM Multimedia 2005, CIVR (Conference on Image and Video Retrieval) 2005, and ACM SIGIR 2008. He serves in the editorial boards of:ACM Transactions of Information Systems (ACM), Foundation and Trends in Information Retrieval (NOW), The Visual Computer (Springer Verlag), and Multimedia Tools and Applications (Kluwer). He is the member of steering committee of CIVR, Computer Graphics International, and Multimedia Modeling conference series; and as member of International Review Panels of two large-scale research projects in Europe.

Table of Contents

Conference Program

Saturday, August 28, 2010

9:15–9:30	Opening Remarks
9:30–10:00	*Constructing Large-Scale Person Ontology from Wikipedia* Yumi Shibaki, Masaaki Nagata and Kazuhide Yamamoto
10:00–10:30	*Using the Wikipedia Link Structure to Correct the Wikipedia Link Structure* Benjamin Mark Pateman and Colin Johnson
10:30–11:00	Coffee Break
11:00–11:30	*Extending English ACE 2005 Corpus Annotation with Ground-truth Links to Wikipedia* Luisa Bentivogli, Pamela Forner, Claudio Giuliano, Alessandro Marchetti, Emanuele Pianta and Kateryna Tymoshenko
11:30–12:00	*Expanding textual entailment corpora fromWikipedia using co-training* Fabio Massimo Zanzotto and Marco Pennacchiotti
12:00–12:30	*Pruning Non-Informative Text Through Non-Expert Annotations to Improve Aspect-Level Sentiment Classification* Ji Fang, Bob Price and Lotti Price
12:30–14:00	Lunch Break
14:00–15:00	Invited Talk by Tat-Seng Chua, National University of Singapore
15:00–15:30	*Measuring Conceptual Similarity by Spreading Activation over Wikipedia's Hyperlink Structure* Stephan Gouws, G-J van Rooyen and Herman A. Engelbrecht
15:30–16:00	Coffee Break
16:00–16:30	*Identifying and Ranking Topic Clusters in the Blogosphere* M. Atif Qureshi, Arjumand Younus, Muhammad Saeed, Nasir Touheed, Emanuele Pianta and Kateryna Tymoshenko
16:30–16:50	*Helping Volunteer Translators, Fostering Language Resources* Masao Utiyama, Takeshi Abekawa, Eiichiro Sumita and Kyo Kageura
16:50–17:30	Discussion