ACL-IJCNLP 2009

People's Web 2009

2009 Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources

Proceedings of the Workshop

7 August 2009
Suntec, Singapore

Production and Manufacturing by
World Scientific Publishing Co Pte Ltd
5 Toh Tuck Link
Singapore 596224

Deutsche
Forschungsgemeinschaft
DFG

VolkswagenStiftung

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-932432-55-8 / 1-932432-55-8

Preface

Welcome to the proceedings of the ACL Workshop "The People's Web Meets NLP: Collaboratively Constructed Semantic Resources". The workshop attracted 21 submissions, of which 9 are included in these proceedings. We are gratified by this level of interest.

This workshop was motivated by the observation that the NLP community is currently considerably influenced by online resources, which are collaboratively constructed by ordinary users on the Web. In many works, such resources have been used as semantic resources overcoming the knowledge acquisition bottleneck and coverage problems pertinent to conventional lexical semantic resources. The resource that has gained the greatest popularity in this respect so far is Wikipedia. However, the scope of the workshop deliberately exceeded Wikipedia. We are happy that the proceedings include papers on resources such as Wiktionary, Mechanical Turk, or creating semantic resources through online games. This encourages us in our belief that collaboratively constructed semantic resources are of growing interest for the natural language processing community.

We should also add that we hoped to bring together researchers from both worlds: those using collaboratively created resources in NLP applications and those using NLP applications for improving the resources or extracting different types of semantic information from them. This is also reflected in the proceedings, although the stronger interest was taken in using semantic resources for NLP applications.

We thank the Volkswagen Foundation and the German Research Foundation for supporting the workshop.

Iryna Gurevych and Torsten Zesch

Organizers:

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

Program Committee:

Delphine Bernhard, Technische Universität Darmstadt
Paul Buitelaar, DERI, National University of Ireland, Galway
Razvan Bunescu, University of Texas at Austin
Pablo Castells, Universidad Autnonoma de Madrid
Philipp Cimiano, Delft University of Technology
Irene Cramer, Dortmund University of Technology
Andras Csomai, Google Inc.
Ernesto De Luca, University of Magdeburg
Roxana Girju, University of Illinois at Urbana-Champaign
Andreas Hotho, University of Kassel
Graeme Hirst, University of Toronto
Ed Hovy, University of Southern California
Jussi Karlgren, Swedish Institute of Computer Science
Boris Katz, Massachusetts Institute of Technology
Adam Kilgarriff, Brighton, Lexical Computing Ltd, UK
Chin-Yew Lin, Microsoft Research
James Martin, University of Colorado Boulder
Olena Medelyan, University of Waikato
David Milne, University of Waikato
Saif Mohammad, University of Maryland
Dan Moldovan, University of Texas at Dallas
Kotaro Nakayama, University of Tokyo
Ani Nenkova, University of Pennsylvania
Guenter Neumann, DFKI Saarbrücken
Maarten de Rijke, University of Amsterdam
Magnus Sahlgren, Swedish Institute of Computer Science
Manfred Stede, Potsdam University
Benno Stein, Bauhaus University Weimar
Tonio Wandmacher, University of Osnabrück
Rene Witte, Concordia University Montral
Hans-Peter Zorn, European Media Lab, Heidelberg

Invited Speaker:

Rada Mihalcea, University of North Texas

Sponsors:

Volkswagen Foundation as part of the Lichtenberg-Professorship Program under grant No. I/82806 and German Research Foundation (DFG) under grant 798/1-3.

Table of Contents

Conference Program

Friday, August 7, 2009

8:45–9:00 Opening Remarks

09:00–09:30 *A Novel Approach to Automatic Gazetteer Generation using Wikipedia*
Ziqi Zhang and Jose Iria

09:30–10:00 *Named Entity Recognition in Wikipedia*
Dominic Balasuriya, Nicky Ringland, Joel Nothman, Tara Murphy and James R. Curran

10:00–10:30 Coffee Break

10:30–11:00 *Wiktionary for Natural Language Processing: Methodology and Limitations*
Emmanuel Navarro, Franck Sajous, Bruno Gaume, Laurent Prévot, ShuKai Hsieh, Ivy Kuo, Pierre Magistry and Chu-Ren Huang

11:00–11:20 *Using the Wiktionary Graph Structure for Synonym Detection*
Timothy Weale, Chris Brew and Eric Fosler-Lussier

11:20–11:40 *Automatic Content-Based Categorization of Wikipedia Articles*
Zeno Gantner and Lars Schmidt-Thieme

11:40–12:00 *Evaluating a Statistical CCG Parser on Wikipedia*
Matthew Honnibal, Joel Nothman and James R. Curran

12:10–13:50 Lunch Break

13:50–14:50 Invited Talk by Rada Mihalcea

15:00–15:30 *Construction of Disambiguated Folksonomy Ontologies Using Wikipedia*
Noriko Tomuro and Andriy Shepitsen

15:30–16:00 Coffee Break

16:00–16:20 *Acquiring High Quality Non-Expert Knowledge from On-Demand Workforce*
Donghui Feng, Sveva Besana and Remi Zajac

Friday, August 7, 2009 (continued)

16:20–16:40 *Constructing an Anaphorically Annotated Corpus with Non-Experts: Assessing the Quality of Collaborative Annotations*
Jon Chamberlain, Udo Kruschwitz and Massimo Poesio

16:40–17:00 Discussion