

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

**At the 23rd International Conference on Computational Linguistics 2010**

**Beijing, China  
28 August 2010**

**ISBN: 978-1-61782-348-0**

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2010) by the Chinese Information Processing Society of China  
All rights reserved.

Printed by Curran Associates, Inc. (2011)

For permission requests, please contact the Chinese Information Processing Society of China  
at the address below.

Chinese Information Processing Society of China  
No. 4 Zhongguancun South 4th St.  
Haidian District Beijing  
China 100190

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2634  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Constructing Large-Scale Person Ontology from Wikipedia</i> Yumi Shibaki, Masaaki Nagata and Kazuhide Yamamoto .....	1
<i>Using the Wikipedia Link Structure to Correct the Wikipedia Link Structure</i> Benjamin Mark Pateman and Colin Johnson .....	10
<i>Extending English ACE 2005 Corpus Annotation with Ground-truth Links to Wikipedia</i> Luisa Bentivogli, Pamela Forner, Claudio Giuliano, Alessandro Marchetti, Emanuele Pianta and Kateryna Tymoshenko .....	19
<i>Expanding textual entailment corpora from Wikipedia using co-training</i> Fabio Massimo Zanzotto and Marco Pennacchiotti .....	28
<i>Pruning Non-Informative Text Through Non-Expert Annotations to Improve Aspect-Level Sentiment Classification</i> Ji Fang, Bob Price and Lotti Price .....	37
<i>Measuring Conceptual Similarity by Spreading Activation over Wikipedia's Hyperlink Structure</i> Stephan Gouws, G-J van Rooyen and Herman A. Engelbrecht .....	46
<i>Identifying and Ranking Topic Clusters in the Blogosphere</i> M. Atif Qureshi, Arjumand Younus, Muhammad Saeed, Nasir Touheed, Emanuele Pianta and Kateryna Tymoshenko .....	55
<i>Helping Volunteer Translators, Fostering Language Resources</i> Masao Utiyama, Takeshi Abekawa, Eiichiro Sumita and Kyo Kageura .....	63

## 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 from Wikipedia 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