

# **Joint Workshop on Automatic Knowledge Base Construction and Web-scale 2012**

**(AKBC-WEKEX 2012)**

**Held at the 2012 Conference of the North American Chapter of the Association for  
Computational Linguistics: Human Language Technologies (NAACL-HNT 2012)**

**Montreal, Quebec, Canada  
3 – 8 June 2012**

**ISBN: 978-1-62276-591-1**

**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© (2012) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2012)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**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>Towards Distributed MCMC Inference in Probabilistic Knowledge Bases</i> Mathias Niepert, Christian Meilicke and Heiner Stuckenschmidt .....	1
<i>Collectively Representing Semi-Structured Data from the Web</i> Bhavana Dalvi, William Cohen and Jamie Callan .....	7
<i>Unsupervised Content Discovery from Concise Summaries</i> Horacio Saggion .....	13
<i>Automatic Evaluation of Relation Extraction Systems on Large-scale</i> Mirko Bronzi, Zhaochen Guo, Filipe Mesquita, Denilson Barbosa and Paolo Merialdo .....	19
<i>Relabeling Distantly Supervised Training Data for Temporal Knowledge Base Population</i> Suzanne Tamang and Heng Ji .....	25
<i>Web Based Collection and Comparison of Cognitive Properties in English and Chinese</i> Bin Li, Jiajun Chen and Yingjie Zhang .....	31
<i>Population of a Knowledge Base for News Metadata from Unstructured Text and Web Data</i> Rosa Stern and Benoît Sagot .....	35
<i>Real-time Population of Knowledge Bases: Opportunities and Challenges</i> Ndapandula Nakashole and Gerhard Weikum .....	41
<i>Adding Distributional Semantics to Knowledge Base Entities through Web-scale Entity Linking</i> Matt Gardner .....	46
<i>KrakeN: N-ary Facts in Open Information Extraction</i> Alan Akbik and Alexander Löser .....	52
<i>Structural Linguistics and Unsupervised Information Extraction</i> Ralph Grishman .....	57
<i>A Context-Aware Approach to Entity Linking</i> Veselin Stoyanov, James Mayfield, Tan Xu, Douglas Oard, Dawn Lawrie, Tim Oates and Tim Finin .....	62
<i>Evaluating the Quality of a Knowledge Base Populated from Text</i> James Mayfield and Tim Finin .....	68
<i>Constructing a Textual KB from a Biology TextBook</i> Peter Clark, Phil Harrison, Niranjana Balasubramanian and Oren Etzioni .....	74
<i>Knowledge Extraction and Joint Inference Using Tractable Markov Logic</i> Chloé Kiddon and Pedro Domingos .....	79

<i>Entity Linking at Web Scale</i>	
Thomas Lin, and Oren Etzioni .....	84
<i>Human-Machine Cooperation: Supporting User Corrections to Automatically Constructed KBs</i>	
Michael Wick, Karl Schultz and Andrew McCallum .....	89
<i>Annotated Gigaword</i>	
Courtney Napoles, Matthew Gormley and Benjamin Van Durme .....	95
<i>Rel-grams: A Probabilistic Model of Relations in Text</i>	
Niranjana Balasubramanian, Stephen Soderland, and Oren Etzioni .....	101
<i>Automatic Knowledge Base Construction using Probabilistic Extraction, Deductive Reasoning, and Human Feedback</i>	
Daisy Zhe Wang, Yang Chen, Sean Goldberg, Christian Grant and Kun Li .....	106
<i>Monte Carlo MCMC: Efficient Inference by Sampling Factors</i>	
Sameer Singh, Michael Wick and Andrew McCallum .....	111
<i>Probabilistic Databases of Universal Schema</i>	
Limin Yao, Sebastian Riedel and Andrew McCallum .....	116
<i>Using Textual Patterns to Learn Expected Event Frequencies</i>	
Jonathan Gordon and Lenhart Schubert .....	122