

**NAACL-HLT 2012**

**Proceedings of the Joint Workshop on  
Automatic Knowledge Base Construction  
and  
Web-scale Knowledge Extraction  
(AKBC-WEKEX 2012)**

**June 7-8, 2012  
Montréal, Canada**

Production and Manufacturing by  
*Omnipress, Inc.*  
2600 Anderson Street  
Madison, WI 53707  
USA

©2012 The Association for Computational Linguistics

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](mailto:acl@aclweb.org)

ISBN13: 978-1-937284-20-6, ISBN10: 1-937284-20-4

## Introduction

Recently, there has been a significant amount of interest in automatically creating large-scale knowledge bases (KBs) from unstructured text. The Web-scale knowledge extraction task presents a unique set of opportunities and challenges. The resulting knowledge bases can have the advantage of scale and coverage. They have been enriched by linking to the Semantic Web, in particular the growing linked open dataset (LOD). These semantic knowledge bases have been used for a wide variety of Natural Language Processing, Knowledge Representation, and Reasoning applications such as semantic search, question answering, entity resolution, ontology mapping etc. The automatic construction of these KBs has been enabled by research in areas including natural language processing, information extraction, information integration, databases, search and machine learning. There are substantial scientific and engineering challenges in advancing and integrating such relevant methodologies.

With this year's workshop, we would like to resume the positive experiences from two previous workshops: AKBC-2010 and WEKEX-2011. The joint AKBC-WEKEX workshop will serve as a forum for researchers working in the area of automated knowledge harvesting from text. By having invited talks by leading researchers from industry, academia, and the government, and by focusing particularly on vision papers, we aim to provide a vivid forum of discussion about the field of automated knowledge base construction. For more details on the workshop, please visit: <http://akbcwekex2012.wordpress.com/>

James Fan, Raphael Hoffman, Aditya Kalyanpur, Sebastian Riedel, Fabian Suchanek, Partha Talukdar



**Organizers:**

James Fan (IBM Research)  
Raphael Hoffman (University of Washington)  
Aditya Kalyanpur (IBM Research)  
Sebastian Riedel (University of Massachusetts, Amherst)  
Fabian Suchanek (Max-Planck Institute for Informatics)  
Partha Pratim Talukdar (Carnegie Mellon University)

**Steering Committee:**

Oren Etzioni (University of Washington)  
Andrew McCallum (University of Massachusetts, Amherst)  
Fernando Pereira (Google Research)  
Gerhard Weikum (Max-Planck Institute for Informatics)

**Invited Speaker:**

Nilesh Dalvi (Yahoo Research)  
Bonnie Dorr (DARPA)  
Oren Etzioni (University of Washington)  
James Fan / Aditya Kalyanpur (IBM Research)  
Ed Hovy (University of Southern California/ISI)  
Andrew McCallum (University of Massachusetts, Amherst)  
Fernando Pereira (Google Research)  
Tom Mitchell (Carnegie Mellon University)  
Patrick Pantel (Microsoft Research)  
Chris Re (University of Wisconsin, Madison)  
Steffen Staab (University of Koblenz)

**Program Committee:**

Soren Auer (University of Leipzig)  
Ken Barker (IBM Research)  
Peter Clark (Vulcan Inc.)  
Amol Deshpande (University of Maryland)  
Anhai Doan (University of Wisconsin, Madison)  
Oren Etzioni (University of Washington)  
Tony Fader (University of Washington)  
Alfio Gliozzo (IBM Research)  
Alon Halevy (Google Research)  
Ed Hovy (University of Southern California/ISI)  
Zack Ives (University of Pennsylvania)  
Vladimir Kolovski (Novartis)  
Xiao Ling (University of Washington)

Andrew McCallum (University of Massachusetts, Amherst)  
Goran Nenadic (University of Manchester)  
Patrick Pantel (Microsoft Research)  
Marius Pasca (Google Research)  
Chris Re (University of Wisconsin, Madison)  
Alan Ritter (University of Washington)  
Sunita Sarawagi (IIT Bombay)  
Sameer Singh (University of Massachusetts, Amherst)  
Martin Theobald (Max-Planck Institute for Informatics)  
Gerhard Weikum (Max-Planck Institute for Informatics)  
Limin Yao (University of Massachusetts, Amherst)

## 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, Niranjan 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>	
Niranjan 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, Christan 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



# Workshop Program

## Thursday, June 7, 2012

- 8:30–9:00      Opening Remarks
- 9:00–9:45      Invited Talk: Oren Etzioni
- 9:45–10:30     Invited Talk: Patrick Pantel
- 10:30–11:00    Break
- 11:00–11:45    Invited Talk: Andrew McCallum
- 11:45–12:30    Invited Talk: James Fan / Aditya Kalyanpur
- 12:30–14:00    Lunch
- 14:00–14:45    Poster Quick Presentations
- 14:45–15:30    Top 3 Talks
- 15:30–16:00    Break
- 16:00–18:00    Poster Session

## Friday, June 8, 2012

- 9:00–9:45      Invited Talk: Fernando Pereira
- 9:45–10:30     Invited Talk: Tom Mitchell
- 10:30–11:00    Break
- 11:00–11:45    Invited Talk: Chris Re
- 11:45–12:30    Invited Talk: Steffen Staab

**Friday, June 8, 2012 (continued)**

12:30–14:00 Lunch

14:00–14:45 Invited Talk: Ed Hovy

14:45–15:30 Invited Talk: Nilesh Davi

15:30–16:00 Break

16:00–16:30 Invited Talk: Bonnie Dorr

16:30–18:00 Unconference