

TextGraphs-8

# **Graph-Based Methods for Natural Language Processing**

**Proceedings of the Workshop**

18 October 2013  
Grand Hyatt Seattle  
Seattle, Washington, USA

©2013 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)

ISBN 978-1-937284-97-8

## Introduction to TextGraphs-8

For the past 7 years, the series of TextGraphs workshops have exposed and encouraged the synergy between the field of Graph Theory (GT) and Natural Language Processing (NLP). The mix between the two started small, with graph theoretical framework providing efficient and elegant solutions for NLP applications that focused on single documents for part-of-speech tagging, word sense disambiguation and semantic role labeling. It then got progressively larger with ontology learning and information extraction from large text collections, and have reached web scale through the new fields of research that focus on information propagation in social networks, rumor proliferation, e-reputation, multiple entity detection, language dynamics learning and future events prediction to name but a few.

The 8th edition of the TextGraphs workshop aimed to be a new step in the series, focused on issues and solutions for large-scale graphs, such as those derived for web-scale knowledge acquisition or social networks. We encouraged the description of novel NLP problems or applications that have emerged in recent years which can be addressed with graph-based solutions, as well as novel graph-based solutions to known NLP tasks. Continuing to bring together researchers interested in Graph Theory applied to Natural Language Processing, provides an environment for further integration of graph-based solutions into NLP tasks. A deeper understanding of new theories of graph-based algorithms is likely to help create new approaches and widen the usage of graphs for NLP applications.

This edition of the TextGraphs workshop took place on October 18<sup>th</sup>, 2013, in Seattle, WA, immediately preceding the Conference on Empirical Methods in Natural Language Processing – EMNLP 2013.

This volume contains papers accepted for presentation at the workshop. We issued calls for regular papers, short late-breaking papers, and demos. After careful review by the program committee of the 15 submissions received – 12 regular papers, 2 short papers and 1 demo – 8 regular papers, 2 short papers and 1 demo were accepted for presentation. The accepted papers address varied problems – from theoretical and general considerations, to NLP and also "real-world" applications - through interesting variations in known and also novel graph-based methods.

We are thankful to the members of the program committee, who have provided high quality reviews in a timely fashion despite the holiday season, and all submissions have benefited from this expert feedback.

We were lucky to have two excellent speakers for this year's event. We thank Oren Etzioni and Pedro Domingos for their enthusiastic acceptance and presentations.

Zornitsa Kozareva, Irina Matveeva, Gabor Melli, Vivi Nastase

October, 2013



**Organizers:**

Zornitsa Kozareva, ISI, University of South California (USA)  
Irina Matveeva, NexLP (USA)  
Gabor Melli, VigLink (USA)  
Vivi Nastase, FBK (Italy)

**Program Committee:**

Patricio Barco, University of Alicante (Spain)  
Chris Biemann, Darmstadt University of Technology (Germany)  
Razvan Bunescu, Ohio University (USA)  
Guillaume Cleuziou, University of Orléans (France)  
Bruno Crémilleux, University of Caen (France)  
Mona Diab, Columbia University (USA)  
Aram Galystan, ISI (University of Southern California (USA)  
Lise Getoor, University of Maryland (USA)  
Filip Ginter, University of Turku (Finland)  
Amit Goyal, Yahoo! Labs (USA)  
Udo Hahn, University of Jena (Germany)  
Liang Huang, City University of New York (USA)  
Kristina Lerman, ISI, University of South California (USA)  
Mausam, University of Washington (USA)  
Sofus Macskassy, ISI, University of Southern California (USA)  
Rada Mihalcea, University of North Texas (USA)  
Gerard de Melo, University of Berkeley (USA)  
Andres Montoyo, University of Alicante (Spain)  
Alessandro Moschitti, University of Trento (Italy)  
Animesh Mukherjee, Indian Institute of Technology (India)  
Philippe Muller, Paul Sabatier University (France)  
Rafael Munoz, University of Alicante (Spain)  
Preslav Nakov, Qatar Foundation (Qatar)  
Roberto Navigli, Sapienza, Università di Roma (Italy)  
Guenther Neumann, DFKI (Germany)  
Zaiqing Nie, Microsoft (China)  
Tae-Gil Noh, University of Heidelberg (Germany)  
Arzucan Özgür, Bogazici University (Turkey)  
Manuel Palomar, University of Alicante (Spain)  
Simone Paolo Ponzetto, Sapienza Università di Roma (Italy)  
Octavian Popescu, FBK (Italy)  
Alan Ritter, University of Washington (USA)  
Paolo Rosso, Technical University of Valencia (Spain)  
Stephen Soderland, University of Washington (USA)

Thamar Solorio, University of Alabama (USA)  
Veselin Stoyanov, Johns Hopkins University (USA)  
Mihai Surdeanu, University of Arizona (USA)  
Paul Tarau, University of North Texas (USA)  
Konstantin Voevodski, Google (USA)  
Rui Wang, DFKI GmbH (Germany)  
Fabio Massimo Zanzotto, University of Rome (Italy)

**Invited Speakers:**

Oren Etzioni, Allen Institute for Artificial Intelligence (USA)  
Pedro Domingos, University of Washington (USA)

## Table of Contents

<i>Event-Centered Information Retrieval Using Kernels on Event Graphs</i> Goran Glavaš and Jan Šnajder .....	1
<i>JoBimText Visualizer: A Graph-based Approach to Contextualizing Distributional Similarity</i> Chris Biemann, Bonaventura Coppola, Michael R. Glass, Alfio Gliozzo, Matthew Hatem and Martin Riedl .....	6
<i>Merging Word Senses</i> Sumit Bhagwani, Shrutiranjana Satapathy and Harish Karnick .....	11
<i>Reconstructing Big Semantic Similarity Networks</i> Ai He, Shefali Sharma and Chun-Nan Hsu .....	20
<i>Graph-Based Unsupervised Learning of Word Similarities Using Heterogeneous Feature Types</i> Avneesh Saluja and Jiri Navratil .....	29
<i>From Global to Local Similarities: A Graph-Based Contextualization Method using Distributional Thesauri</i> Martin Riedl and Chris Biemann .....	39
<i>Understanding seed selection in bootstrapping</i> Yo Ehara, Issei Sato, Hidekazu Oiwa and Hiroshi Nakagawa .....	44
<i>Graph-Structures Matching for Review Relevance Identification</i> Lakshmi Ramachandran and Edward Gehring .....	53
<i>Automatic Extraction of Reasoning Chains from Textual Reports</i> Gleb Sizov and Pinar Öztürk .....	61
<i>Graph-based Approaches for Organization Entity Resolution in MapReduce</i> Hakan Kardes, Deepak Konidena, Siddharth Agrawal, Micah Huff and Ang Sun .....	70
<i>A Graph-Based Approach to Skill Extraction from Text</i> Ilkka Kivimäki, Alexander Panchenko, Adrien Dessy, Dries Verdegem, Pascal Francq, Hugues Bersini and Marco Saerens .....	79





# Workshop Program

**Friday, October 18, 2013**

7:30-8:50 Breakfast and Registration

8:50-9:00 Opening Remarks

**(9:00-10:30) Session One**

9:00-10:05 Invited talk: Opportunities for Graph-based Machine Reading by Oren Etzioni

10:05–10:20 *Event-Centered Information Retrieval Using Kernels on Event Graphs*  
Goran Glavaš and Jan Šnajder

10:20–10:30 *JoBimText Visualizer: A Graph-based Approach to Contextualizing Distributional Similarity*  
Chris Biemann, Bonaventura Coppola, Michael R. Glass, Alfio Gliozzo, Matthew Hatem and Martin Riedl

10:30-11:00 Coffee Break and Demo

**(11:00-12:30) Session Two**

11:00–11:25 *Merging Word Senses*  
Sumit Bhagwani, Shrutiranjana Satapathy and Harish Karnick

11:25–11:50 *Reconstructing Big Semantic Similarity Networks*  
Ai He, Shefali Sharma and Chun-Nan Hsu

11:50–12:15 *Graph-Based Unsupervised Learning of Word Similarities Using Heterogeneous Feature Types*  
Avneesh Saluja and Jiri Navratil

12:15–12:30 *From Global to Local Similarities: A Graph-Based Contextualization Method using Distributional Thesauri*  
Martin Riedl and Chris Biemann

12:30-2:00 Lunch Break

**Friday, October 18, 2013 (continued)**

**(2:00-3:30) Session Three**

2:00-3:05 Invited talk: Extracting Tractable Probabilistic Knowledge Graphs from Text by Pedro Domingos

3:05–3:30 *Understanding seed selection in bootstrapping*  
Yo Ehara, Issei Sato, Hidekazu Oiwa and Hiroshi Nakagawa

3:30-4:00 Coffee Break

**(4:00-5:40) Session Four**

4:00–4:25 *Graph-Structures Matching for Review Relevance Identification*  
Lakshmi Ramachandran and Edward Gehringer

4:25–4:50 *Automatic Extraction of Reasoning Chains from Textual Reports*  
Gleb Sizov and Pinar Öztürk

4:50–5:15 *Graph-based Approaches for Organization Entity Resolution in MapReduce*  
Hakan Kardes, Deepak Konidena, Siddharth Agrawal, Micah Huff and Ang Sun

5:15–5:40 *A Graph-Based Approach to Skill Extraction from Text*  
Ilkka Kivimäki, Alexander Panchenko, Adrien Dessy, Dries Verdegem, Pascal Francq, Hugues Bersini and Marco Saerens

5:40-6:00 Closing Remarks