Coling 2008

22nd International Conference on Computational Linguistics

Proceedings of the 3rd Textgraphs workshop on Graph-based Algorithms for Natural Language Processing

Workshop chairs:
Irina Matveeva, Chris Biemann, Monojit Choudhury, Mona Diab

24 August 2008
Manchester, UK

TextGraphs-3 Workshop at COLING 2008 was sponsored by Microsoft Research India

Order copies of this and other Coling 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 978-1-905593-57-6

Design by Chimney Design, Brighton, UK
Production and manufacture by One Digital, Brighton, UK

Introduction

Recent years have shown an increased interest in bringing the field of graph theory into Natural Language Processing. In many NLP applications entities can be naturally represented as nodes in a graph and relations between them can be represented as edges. Recent research has shown that graph-based representations of linguistic units as diverse as words, sentences and documents give rise to novel and efficient solutions in a variety of NLP tasks, ranging from part of speech tagging, word sense disambiguation and parsing to information extraction, semantic role assignment, summarization and sentiment analysis. The contribution of the graph representation, in addition to its intuitiveness, resides in the possibility to relate linguistic entities beyond pairwise comparison. This volume contains papers accepted for presentation at the TextGraphs-3 2008 Workshop on Graph- Based Algorithms for Natural Language Processing. This event took place on August 24, 2008, in Manchester, UK, immediately following COLING 2008, the 22nd International Conference on Computational Linguistics. It was the third workshop on this topic, building on the success of the first and second TextGraphs workshop at HLT-NAACL 2006 and 2007. The workshop aimed at bringing together researchers working on problems related to the use of graph-based algorithms for Natural Language Processing and on the theory of graph-based methods. It addressed a broad spectrum of research areas to foster exchange of ideas and to help identify principles of using the graph notions that go beyond an ad-hoc usage.

We issued calls for both regular and short, late-breaking papers. Six regular and three short papers were accepted for presentation, based on the careful reviews of our program committee. We are indebted to all program committee members for their thoughtful, high quality and elaborate reviews, especially considering our extremely tight time frame for reviewing. The papers appearing in this volume have surely benefited from their expert feedback. This year's workshop attracted papers employing graphs in a wide range of settings, so we are proud to present a very diverse program this year. N. Hathout acquires morphological structure from a lexicon employing the bipartite graph between headwords' formal semantic features. Mapping of text to a graph-based meaning representation is conducted by S. Muresan, using a recent grammar formalism. A. B. Massé et al. lay out a general theoretical framework for addressing the symbol grounding problem in digital dictionaries. A. Moschitti and F.M. Zanzotto use Kernel methods on tree pairs for recognizing textual entailment. Combining co-occurrence and phonological similarity, K. Ichioka and F. Fukumoto semantically cluster onomatopoetic words in Japanese. D. Rao et al. examine several random walk based approaches to measure word similarity. B. McGillivray et al. address cluster overlapping with correspondence analysis and apply their method to cluster English and Italian verbs and nouns. A domain-specific summarization method ranking nodes in a graph of concepts is introduced by L. Plaza Morales et al. The topology of associative concept dictionaries is modeled by H. Akama et al., who report interesting scale free properties of such networks.

Finally, having a prominent researcher as an invited speaker greatly contributes to the quality of the workshop. We thank Dragomir Radev for his talk and for the support he provided for this as well as all the previous Textgraphs workshops. We are also grateful to Microsoft Research India for sponsoring the travel and accomodation of the invited speaker.

Irina Matveeva, Chris Biemann, Monojit Choudhury and Mona Diab
August 2008

Organizers:

Irina Matveeva, Accenture Technology Labs
Chris Biemann, Powerset
Monojit Choudhury, Microsoft Research India
Mona Diab, Columbia University

Programme Committee:

Eneko Agirre, University of the Basque Country
Edo Airoldi, Princeton University
Regina Barzilay, MIT
Fernando Diaz, Yahoo! Montreal
Günes Erkan, Google
Michael Gamon, Microsoft Research Redmond
Andrew Goldberg, University of Wisconsin
Hany Hassan, IBM Egypt
Samer Hassan, University of North Texas
Gina Levow, University of Chicago
Rada Mihalcea, University of North Texas
Animesh Mukherjee, IIT Kharagpur
Dragomir Radev, University of Michigan
Uwe Quasthoff, University of Leipzig
Aitor Soroa, University of the Basque Country
Hans Friedrich Witschel, University of Leipzig
Fabio Massimo Zanzotto, University of Rome "Tor Vergata"
Thorsten Zesch, University of Darmstadt

Invited Speaker:

Dragomir Radev, University of Michigan, Ann Arbor

Table of Contents

Conference Programme

Sunday, August 24, 2008

9:20–9:25 Opening

9:30–10:30 Invited Talk by Dragomir Radev on "Lexical Affinity"

10:30–11:00 Break

Session I: Full Papers

11:00–11:30 *Acquistion of the Morphological Structure of the Lexicon Based on Lexical Similarity and Formal Analogy*
Nabil Hathout

11:30–12:00 *Learning to Map Text to Graph-Based Meaning Representations via Grammar Induction*
Smaranda Muresan

12:00–12:30 *How is Meaning Grounded in Dictionary Definitions?*
Alexandre Blondin Mass, Guillaume Chicoisne, Yassine Gargouri, Stevan Harnad, Odile Marcotte and Olivier Picard

12:30–14:00 Lunch

Session II: Full Papers

14:00–14:30 *Encoding Tree Pair-Based Graphs in Learning Algorithms: The Textual Entailment Recognition Case*
Alessandro Moschitti and Fabio Massimo Zanzotto

14:30–15:00 *Graph-Based Clustering for Semantic Classification of Onomatopoetic Words*
Kenichi Ichioka and Fumiyo Fukumoto

15:00–15:30 *Affinity Measures Based on the Graph Laplacian*
Delip Rao, David Yarowsky and Chris Callison-Burch

15:30–16:00 Break

Sunday, August 24, 2008 (continued)

Session III: Short Papers

16:00–16:20 *Semantic Structure from Correspondence Analysis*
Barbara McGillivray, Christer Johansson and Daniel Apollon

16:20–16:40 *Concept-Graph Based Biomedical Automatic Summarization Using Ontologies*
Laura Plaza, Alberto Daz and Pablo Gervs

16:40–17:00 *Random Graph Model Simulations of Semantic Networks for Associative Concept Dictionaries*
Hiroyuki Akama, Jaeyoung Jung, Terry Joyce and Maki Miyake