Proceedings of the

EACL 2009 Workshop on

GEMS: GEometrical Models of Natural Language Semantics

Endorsed by

the Association for Computational Linguistics
SIGLEX and SIGSEM, two Special Interest Groups of ACL

Edited by

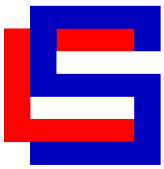
Roberto Basili
and
Marco Pennacchiotti

31 March 2009
Megaron Athens International Conference Centre
Athens, Greece

Production and Manufacturing by
TEHNOGRAFIA DIGITAL PRESS
7 Ektoros Street
152 35 Vrilissia
Athens, Greece

Endorsed by:

ACL SIGLEX



ACL SIGSEM

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

Introduction

The geometry of distributional models of lexical semantics represent a core topic in contemporary computational linguistics for its impact on several advanced Natural Language Processing tasks and some related knowledge fields (as social science and humanities).

The goal of the EACL 2009 GEMS Workshop on "*GEometrical Models of natural Language Semantics*" was to stimulate research on semantic spaces and distributional methods in NLP, by adopting an interdisciplinary view. This aimed to enforce the proper exchange of ideas, results and resources among often independent communities. The workshop provided a common ground for a fruitful discussion among experts of distributional approaches, collocational corpus analysis and machine learning, researchers interested in the use of quantitative models in NLP applications (like question answering, summarization or textual entailment), experts in formal computational semantics and in other fields of science as well.

The workshop successfully gathered a relevant number of high quality contributions to problems of meaning representation, acquisition and use, based on distributional and vector space models. We received 21 submissions, including short and long papers. Long papers were peer-reviewed by three members of the program committee, short papers by two. As an outcome of the review process, the program committee selected 11 papers for a full presentation, and 4 for short ones. All selected paper have been included in these proceedings. The paper are representative of the current state of the art in the subject, including:

- cutting edge researches on geometric methods and machine learning, such as tensor factorization, kernel methods and Dirichlet process mixture models;
- applications of semantic space models to NLP tasks, such as Textual Entailment Recognition, Ontology Learning, Induction of Selectional Preferences, Verb Classification and Machine Translation
- novel uses of distributional methods for advanced linguistic studies, such as lexical variation and evolution as well as for educational purposes;
- reference comparative studies among different types of semantic spaces.

The papers included in this volume shed some light on the state of the art and the potential applications of semantic spaces in NLP and in related linguistic fields.

We would like to thank all the authors for their hard work dedicated to the submissions. Our deepest gratitude goes to the members of the program committee for their precious reviewing. Most of the impact of this volume is entirely due to their careful analysis and meaningful suggestions to the authors. A special thank goes to Patrick Pantel for his stimulating and visionary invited talk, supported by his own institution. Finally, we acknowledge the EACL 2009 workshop chairs, Miriam Butt, Stephan Clark as well as Kemal Oflazer and David Schlangen, for their constant support across all the preparatory work.

Roberto Basili, University of Roma, *Tor Vergata*, Italy
Marco Pennacchiotti, Yahoo! Inc, Santa Clara, US.

March, 2009

Organizers:

Roberto Basili, University of Roma *Tor Vergata* (Italy)
Marco Pennacchiotti, Yahoo! Inc., Santa Clara, CA (US)

Program Committee:

Marco Baroni, University of Trento (Italy)
Michael W. Berry, University of Tenneesee (US)
Gemma Boleda, Pompeu Fabra University of Barcelona (Spain)
Johan Bos, University of Roma "*La Sapienza*" (Italy)
Paul Buitelaar, DFKI (Germany)
John A. Bullinaria, University of Birmingham (UK)
Rodolfo Delmonte, University *Ca' Foscari* Venice (Italy)
Katrin Erk, University of Texas (US)
Stefan Evert, University of Osnabruck (Germany)
Alfio Massimiliano Gliozzo, STLab - ISTC-CNR (Italy)
Jerry Hobbs, University of Southern California (US)
Alessandro Lenci, University of Pisa (Italy)
Jussi Karlgren, Swedish Institute of Computer Science (Sweden)
Will Lowe, University of Nottingham (UK)
Diana McCarthy, University of Sussex (UK)
Alessandro Moschitti, University of Trento (Italy)
Saif Mohammad, University of Maryland (US)
Sebastian Padó, Stanford University (US)
Patrick Pantel, Yahoo! Inc. (US)
Massimo Poesio, University of Trento (Italy)
Magnus Sahlgren, Swedish Institute of Computer Science (Sweden)
Sabine Schulte im Walde, University of Stuttgart (Germany)
Hinrich Schütze, University of Stuttgart (Germany)
Fabrizio Sebastiani, CNR (Italy)
Suzanne Stevenson, University of Toronto (Canada)
Peter D. Turney, National Research Council (Canada)
Dominic Widdows, Google Research (US)
Yorick Wilks, University of Sheffield (UK)
Fabio Massimo Zanzotto, University of Roma "*Tor Vergata*" (Italy)

Table of Contents

Conference Program

Tuesday, March 31, 2009

8:45–9:00 Opening Remarks

9:00–10:00 Invited Talk by Patrick Pantel

Session 1

9:50–10:15 *One Distributional Memory, Many Semantic Spaces*
Marco Baroni and Alessandro Lenci

10:15–10:40 *Word Space Models of Lexical Variation*
Yves Peirsman and Dirk Speelman

10:40–11:00 Coffee break

Session 2

11:00–11:25 *Unsupervised Classification with Dependency Based Word Spaces*
Klaus Rothenhäusler and Hinrich Schütze

11:25–11:50 *A Study of Convolution Tree Kernel with Local Alignment*
Lidan Zhang and Kwok-Ping Chan

11:50–12:15 *BagPack: A General Framework to Represent Semantic Relations*
Amaç Herdağdelen and Marco Baroni

Tuesday, March 31, 2009 (continued)

Short Presentations

12:15–12:30 *Positioning for Conceptual Development using Latent Semantic Analysis*
Fridolin Wild, Bernhard Hoisl and Gaston Burek

12:30–12:45 *Semantic Similarity of Distractors in Multiple-Choice Tests: Extrinsic Evaluation*
Ruslan Mitkov, Le An Ha, Andrea Varga and Luz Rello

12:45–13:45 Lunch break

Session 3

13:45–14:10 *Paraphrase Assessment in Structured Vector Space: Exploring Parameters and Datasets*
Katrin Erk and Sebastian Padó

14:10–14:35 *SVD Feature Selection for Probabilistic Taxonomy Learning*
Francesca Fallucchi and Fabio Massimo Zanzotto

14:35–15:00 *Unsupervised and Constrained Dirichlet Process Mixture Models for Verb Clustering*
Andreas Vlachos, Anna Korhonen and Zoubin Ghahramani

15:00–15:25 *A Non-negative Tensor Factorization Model for Selectional Preference Induction*
Tim Van de Cruys

Short Presentations

15:25–15:40 *A Graph-Theoretic Algorithm for Automatic Extension of Translation Lexicons*
Beate Dorow, Florian Laws, Lukas Michelbacher, Christian Scheible and Jason Utt

15:40–15:55 *Handling Sparsity for Verb Noun MWE Token Classification*
Mona Diab and Madhav Krishna

16:00–16:30 Coffee break

Tuesday, March 31, 2009 (continued)

Session 4

16:30–16:55 *Semantic Density Analysis: Comparing Word Meaning across Time and Phonetic Space*
Eyal Sagi, Stefan Kaufmann and Brady Clark

16:55–17:20 *Context-theoretic Semantics for Natural Language: an Overview*
Daoud Clarke

Panel

17:20–18:00 Discussion and Concluding Remarks