ACL 2010

GEMS 2010

2010 Workshop on
GEometrical Models of Natural Language Semantics

Proceedings of the Workshop
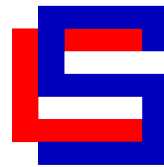
16 July 2010
Uppsala University
Uppsala, Sweden

Endorsed by:

ACL SIGLEX

ACL SIGSEM

Introduction

This volume includes the papers presented at the GEMS 2010 Workshop - Geometrical Models for Natural Language Semantics, held on the 16th July 2010, jointly with the Conference of the Association for Computational Linguistics, ACL 2010, and endorsed by SigLex and SigSem.

Distributional models and semantic spaces represent a core topic in contemporary computational linguistics for their impact on advanced tasks and on other knowledge fields (such as social science and the humanities). Semantic spaces based on simple contextual units have been early used in information retrieval, showing dramatic impact on accuracy and scalability of many tasks. Later on, more linguistically principled spaces have been introduced for large-scale natural language learning problems, such as the acquisition of lexical taxonomies, word sense discrimination, pattern acquisition and conceptual clustering. More recently, specialized distributional models have been successfully applied to solve complex NLP tasks such as question answering, textual entailment and sentiment analysis. Cutting-edge applications include the adoption of semantic spaces as models of rich lexical semantic resources (e.g. lexical networks and lexicalized meaning theories, as frame semantics), and of machine learning approaches based on kernel methods.

The GEMS 2010 Workshops builds on the successful first edition held in Athens in 2009 jointly with the Conference of the European Chapter of the Association for Computational Linguistics, which counted more than 40 registered people and a large audience. As a follow-up of the GEMS 2009 Workshop, the Special Issue of the *Journal of Natural Engineering* dedicated to "*Distributional Lexical Semantics*"[1] is a further proof of the high interest in this research area. The 2010 edition aims at consolidating GEMS' contribution to the field, by stimulating research on semantic spaces and distributional methods for NLP, by pushing for an interdisciplinary view, and by amplifying exchange of ideas, results and resources among often independent communities.

The Workshop has successfully gathered high quality contributions to problems of meaning representation, acquisition and use, including a total of 15 paper submissions. After a peer-review phase, the program committee has selected 8 papers to be presented at the workshop, all of which have been included in these proceedings. The papers are representative of the current state of the art in distributional semantics, including:

- cutting edge researches on geometric techniques and machine learning, such as tensor analysis, non-linear embeddings, kernel methods, and latent topic models;

- applications of semantic space models to lexical acquisition tasks;

- novel optimization techniques for efficient and scalable distributional methods.

We would like to thank all the authors for the hard work dedicated to the submissions, and the members of the program committee for their precious reviewing. A special thanks goes to Katrin Erk for her invited contribution that provides a challenging and inspiring vision on the topic. Finally, we acknowledge the ACL 2010 organization and mostly the workshop chairs, Pushpak Bhattacharyia and David Weir, for their constant support across all the preparatory work.

Roberto Basili, University of Roma, *Tor Vergata*, Italy
Marco Pennacchiotti, *Yahoo! Inc*, Sunnyvale, USA.

June, 2010

---

[1] http://art.uniroma2.it/jnle

Organizers:

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

Program Committee:

Enrique Alfonseca (Google Research, US)
Marco Baroni (University of Trento, Italy)
Paul Buitelaar (DFKI, Germany)
John A. Bullinaria (University of Birmingham, UK)
Carlotta Domeniconi (George Mason University, US)
Katrin Erk (University of Texas, US)
Stefan Evert (University of Osnabruck, Germany)
Alfio Massimiliano Gliozzo (STLab - ISTC-CNR, Italy)
Gregory Grefenstette (Exalead S.A., France)
Alpa Jain (Yahoo! Labs, US)
Jussi Karlgren (Swedish Institute of Computer Science)
Alessandro Lenci (University of Pisa, Italy)
Alessandro Moschitti (University of Trento, Italy)
Sebastian Pado (Stanford University, US)
Ted Pedersen (University of Minnesota, US)
Yves Peirsman (University of Leuven, Belgium)
Ana-Maria Popescu (Yahoo! Labs, US)
Magnus Sahlgren (Swedish Institute of Computer Science, Sweden)
Sabine Schulte im Walde (University of Stuttgart, Germany)
Hristo Tanev (Yahoo! UK, UK)
Tim Van de Cruys (University of Groningen, The Netherlands)
Peter D. Turney (National Research Council, Canada)
Yorick Wilks (University of Sheffield, UK)
Fabio Massimo Zanzotto (University of Roma, Tor Vergata, Italy)

Invited Speaker:

Katrin Erk, University of Texas at Austin, US

Table of Contents

Workshop Program

July 16, 2010

9:25–9:30   Welcome and Opening

Session: Geometry and Semantics

9:30–10:00   *Capturing Nonlinear Structure in Word Spaces through Dimensionality Reduction*
David Jurgens and Keith Stevens

10:00–10:30   *Manifold Learning for the Semi-Supervised Induction of FrameNet Predicates: An Empirical Investigation*
Danilo Croce and Daniele Previtali

10:30–11:00   Coffee Break

Invited Talk

11:00–12:10   *What Is Word Meaning, Really? (And How Can Distributional Models Help Us Describe It?)*
Katrin Erk

Session: Lexical Acquisition (1)

12:10–12:40   *Relatedness Curves for Acquiring Paraphrases*
Georgiana Dinu and Grzegorz Chrupala

12:40–13:10   *A Regression Model of Adjective-Noun Compositionality in Distributional Semantics*
Emiliano Guevara

13:10–14:30   Lunch

July 16, 2010 (continued)

Session: Lexical Acquisition (2)

14:30–15:00   *Semantic Composition with Quotient Algebras*
Daoud Clarke, Rudi Lutz and David Weir

15:00–15:30   *Expectation Vectors: A Semiotics Inspired Approach to Geometric Lexical-Semantic Representation*
Justin Washtell

15:30–16:00   Coffee Break

Session: Computational Aspects

16:00–16:30   *Sketch Techniques for Scaling Distributional Similarity to the Web*
Amit Goyal, Jagadeesh Jagaralamudi, Hal Daumé III and Suresh Venkatasubramanian

16:30–17:00   *Active Learning for Constrained Dirichlet Process Mixture Models*
Andreas Vlachos, Zoubin Ghahramani and Ted Briscoe

Panel

17:00–17:55   GEMS panel

17:55–18:00   Closing Remarks