GWC 2018

Proceedings of the 9th Global Wordnet Conference

Francis Bond, Takayuki Kuribayashi,
Christiane Fellbaum, Piek Vossen (Eds.)

8–12 January, 2018
Nanyang Technological University (NTU)
Singapore

ISBN 978-981-11-7087-4

Foreword

The Ninth Global Wordnet Conference was held at Nanyang Technological University, Singapore from 9–12th January 2018.

The program combined the main conference with a special day on wordnets and word-embeddings and finished with a half day workshop on technology enhanced learning (TEL). There were 4 invited talks, 41 full papers, 15 posters and 4 invited talks on TEL. Including the papers on embeddings, there were 15 rejections: the acceptance rate for full papers was 58% a sign of the consistently high quality of papers submitted to the conference. Copyrights for the papers reside with the original authors.

The invited papers were *One Million Sense-Tagged Instances for Word Sense Disambiguation and Induction* by Ng Hwee Tou (National University of Singapore), *How are you two related? Corpus-based Learning of Lexical Semantic Relations* by Vered Shwartz (Bar-Ilan University), *Inducing Interpretable Word Senses for WSD and Enrichment of Lexical Resources* by Alexander Panchenko (University of Hamburg) and *Using a Grammar Implementation to Teach Writing Skills* by Dan Flickinger (Stanford). As well as many papers on distributional semantics, there were some on extending the coverage of existing wordnets, linking wordnets to new resources (especially in the medical domain), using wordnets for teaching and many other topics. There were papers from 24 different countries with every continent except Antarctica represented.

The conference and workshops were partially supported by the NTU Centre for Liberal Arts and Social Sciences (CLASS) and the Singapore MOE TRF *Grant Syntactic Well-Formedness Diagnosis and Error-Based Coaching in Computer Assisted Language Learning using Machine Translation Technology*. Support for students came from the Global Wordnet Association. We would like to thank the programme committee for their thoughtful and timely reviews.

The conference homepage is http://compling.hss.ntu.edu.sg/events/2018-gwc/.

Francis Bond, Nanyang Technological University

Takayuki Kuribayashi, Nanyang Technological University

Christiane Fellbaum, Princeton University

Piek Vossen, VU University Amsterdam

January 2018

Organizers

Local Chairs Francis Bond and Takayuki Kuribayashi

Program Chairs Christiane Fellbaum, Francis Bond and Piek Vossen

Workshop Chairs Christiane Fellbaum, Francis Bond and Erhard Hinrichs

Program Committee

Adam Pease	Articulate Software, San Jose, USA
Ales Horak	Masaryk University
Alexandre Rademaker	IBM Research Brazil and EMAp/FGV
Antoni Oliver	Universitat Oberta de Catalunya
Bolette Pedersen	University of Copenhagen
Corina Forăscu	Alexandru Ioan Cuza University of Iaşi
Dan Cristea	Alexandru Ioan Cuza University of Iaşi
Darja Fiser	University of Ljubljana
Diptesh Kanojia	IIT Bombay
Emiel van Miltenburg	Vrije Universiteit Amsterdam
Eneko Agirre	University of the Basque Country
Erhard Hinrichs	University of Tübingen
Ewa Rudnicka	Wroclaw University of Technology
Gerard de Melo	Rutgers University
German Rigau	University of the Basque Country (EPV/EHU)
Haldur Oim	University of Tartu
Heili Orav	University of Tartu
Horacio Rodriguez	Universitat Politècnica de Catalunya
Isahara Hitoshi	Toyohashi University of Technology
Janos Csirik	University of Szeged
John P. Mccrae	National University of Ireland, Galway
Kadri Vider	University of Tartu
Kevin Scannell	Saint Louis University
Kyoko Kanzaki	Toyohashi University of Technology
Maciej Piasecki	Wrocław University of Technology
Magnini Bernardo	FBK-IRST
Marten Postma	VU University Amsterdam
Mikhail Khodak	Princeton University
Rada Mihalcea	University of North Texas / Oxford University
Roxane Segers	VU University Amsterdam
Shan Wang	The Education University of Hong Kong
Shu-Kai Hsieh	National Taiwan Normal University

Sonja Bosch	Department of African Languages, University of South Africa
Ted Pedersen	University of Minnesota, Duluth
Timothy Baldwin	The University of Melbourne
Tomaž Erjavec	Dept. of Knowledge Technologies, Jožef Stefan Institute
Verginica Mititelu	Romanian Academy Research Institute for Artificial Intelligence

Invited Speakers

- Ng Hwee Tou, National University of Singapore
- Vered Shwartz, Bar-Ilan University
- Alexander Panchenko, University of Hamburg
- Dan Flickinger, Stanford

Invited Talks

Ng Hwee Tou: One Million Sense-Tagged Instances for Word Sense Disambiguation and Induction

Supervised word sense disambiguation (WSD) systems have achieved the best performance when evaluated on standard benchmark datasets. However, the lack of large amounts of sense-tagged data poses a major hurdle to scaling up supervised WSD systems to disambiguate all words of English. In this talk, I will present a semi- automatic approach to extract and annotate a large sense-tagged corpus. This one-million-word sense-tagged corpus has been publicly released since 2015 and has been used by other researchers working on automated WSD. When trained on this one- million-word sense-tagged corpus, the open source IMS (It Makes Sense) WSD system created in my research group achieves good performance on standard WSD tasks and another word sense induction task.

Vered Shwartz: How are you two related? Corpus-based Learning of Lexical Semantic Relations

Recognizing lexical semantic relations between words is an essential component in semantic applications such as question answering and recognizing textual entailment. In order to overcome lexical variability, such systems traditionally relied heavily on lexical resources such as WordNet.

In the main part of the talk I will discuss our work on automatic detection of lexical semantic relations from free text. This task stems from the limited coverage of lexical resources, both in terms of missing lexical items (proper names, new words) and missing relations between existing items. Typical approaches to address this task are either distributional, i.e. based on the word embeddings of the two target words, or path-based (pattern-based) approach, based on the words co-occurrences in the corpus. I will present our integrated path-based and distributional method for recognizing lexical semantic relations, which is currently the state-of-the-art in this task.

In the second part, I will raise some questions about the interplay of WordNet and word embeddings: is external lexical knowledge obsolete in the deep learning era? And if it isn't, then how can lexical knowledge from WordNet and other resources be incorporated into neural models for semantic applications?

Alexander Panchenko: Inducing Interpretable Word Senses for WSD and Enrichment of Lexical Resources

In this talk, we will discuss induction of sparse and dense word sense representations using graph-based approaches and distributional models. Induced senses are represented by a vector, but also a set of hypernyms, images, and usage examples, derived in an unsupervised and knowledge-free manner, which ensure interpretability of the discovered senses by humans. We showcase the usage of the induced representations for the tasks of word sense disambiguation and enrichment of lexical resources, such as WordNet.

Dan Flickinger: Using a Grammar Implementation to Teach Writing Skills

This paper presents an approach to grammar checking, using a large-scale HPSG grammar of English. The system has been used in a Language Arts & Writing course for McGraw-Hill Education in U.S. classrooms for the past ten years. It has helped over 50,000 primary school students, mostly native English speakers. We have given feedback on over 10 million sentences. The feedback is generated using mal-rules that identify errors with high precision. We are currently looking at extending the system to non-native speakers' English.

Table of Contents