ACL 2013

51st Annual Meeting of the Association for Computational Linguistics

Proceedings of the Conference System Demonstrations

August 4-9, 2013
Sofia, Bulgaria

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53704 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

Preface

Welcome to the proceedings of the system demonstration session. This volume contains the papers of the system demonstrations presented at the 51st Annual Meeting of the Association for Computational Linguistics, held in Sofia, Bulgaria, on August 5th, 2013.

The system demonstrations program offers the presentation of early research prototypes as well as interesting mature systems. The system demonstration chairs and the members of the program committee received 64 submissions, of which 34 were selected for inclusion in the program (acceptance rate of 53%) after review by two members of the program committee.

We would like to thank the members of the program committee for their timely help in reviewing the submissions.

Co-Chairs:

Miriam Butt, University of Konstanz
Sarmad Hussain, Al-Khawarizmi Institute of Computer Science, University of Engineering and Technology, Lahore

Program Committee:

Timothy Baldwin, University of Melbourne
Thierry Declerck, DFKI, Saarbrücken
Koenrad de Smedt, University of Bergen
Mark Dras, Macquarie University
Martin Forst, Netbase
Tracy King, eBay Inc.
Adam Przepiorkowski, Institute of Computer Science, Polish Academy of Sciences
Frédérique Segond, Viseo Reseach and Development
Heike Zinsmeister, IMS, University of Stuttgart
Ruvan Weerasinghe, University of Colombo
Mirna Adriani, University of Indonesia
Sivaji Bandyopadhyay, Jadavpur University
Ying Chen, China Agriculture University
Christopher Culy, University of Tübingen
Alexander Fraser, IMS, University of Stuttgart
Usman Ghani, Al-Khawarizmi Institute of Computer Science, University of Engineering and Technology, Lahore
Annette Hautli, University of Konstanz
Gurpreet Lehal, Punjabi University
Shoushan Li, Soochow University
Katja Markert, University of Leeds
Thomas Mayer, University of Marburg
Christian Rohrdantz, University of Konstanz
Virach Sornlertlamvanich, NECTEC
Alexandros Tantos, Aristotle University of Thessaloniki

Table of Contents

WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations
Seid Muhie Yimam, Iryna Gurevych, Richard Eckart de Castilho and Chris Biemann 1

A Stacking-based Approach to Twitter User Geolocation Prediction
Bo Han, Paul Cook and Timothy Baldwin 7

An Open Source Toolkit for Quantitative Historical Linguistics
Johann-Mattis List and Steven Moran 13

AnnoMarket: An Open Cloud Platform for NLP
Valentin Tablan, Kalina Bontcheva, Ian Roberts, Hamish Cunningham and Marin Dimitrov 19

Detecting Event-Related Links and Sentiments from Social Media Texts
Alexandra Balahur and Hristo Tanev 25

DISSECT - DIStributional SEmantics Composition Toolkit
Georgiana Dinu, Nghia The Pham and Marco Baroni 31

DKPro WSD: A Generalized UIMA-based Framework for Word Sense Disambiguation
Tristan Miller, Nicolai Erbs, Hans-Peter Zorn, Torsten Zesch and Iryna Gurevych 37

Extending an interoperable platform to facilitate the creation of multilingual and multimodal NLP applications
Georgios Kontonatsios, Paul Thompson, Riza Theresa Batista-Navarro, Claudiu Mihăilă, Ioannis Korkontzelos and Sophia Ananiadou 43

FudanNLP: A Toolkit for Chinese Natural Language Processing
Xipeng Qiu, Qi Zhang and Xuanjing Huang 49

ICARUS – An Extensible Graphical Search Tool for Dependency Treebanks
Markus Gärtner, Gregor Thiele, Wolfgang Seeker, Anders Björkelund and Jonas Kuhn 55

Meet EDGAR, a tutoring agent at MONSERRATE
Pedro Fialho, Luísa Coheur, Sérgio Curto, Pedro Cláudio, Ângela Costa, Alberto Abad, Hugo Meinedo and Isabel Trancoso 61

PAL: A Chatterbot System for Answering Domain-specific Questions
Yuanchao Liu, Ming Liu, Xiaolong Wang, Limin Wang and Jingjing Li 67

PhonMatrix: Visualizing co-occurrence constraints of sounds
Thomas Mayer and Christian Rohrdantz 73

QuEst - A translation quality estimation framework
Lucia Specia, , , Kashif Shah, Jose G.C. de Souza and Trevor Cohn 79

SORT: An Interactive Source-Rewriting Tool for Improved Translation
Shachar Mirkin, Sriram Venkatapathy, Marc Dymetman and Ioan Calapodescu 85

Travatar: A Forest-to-String Machine Translation Engine based on Tree Transducers
Graham Neubig 91

PLIS: a Probabilistic Lexical Inference System
Eyal Shnarch, Erel Segal-haLevi, Jacob Goldberger and Ido Dagan 97

A Java Framework for Multilingual Definition and Hypernym Extraction
Stefano Faralli and Roberto Navigli 103

A Visual Analytics System for Cluster Exploration
Andreas Lamprecht, Annette Hautli, Christian Rohrdantz and Tina Bögel 109

Development and Analysis of NLP Pipelines in Argo
Rafal Rak, Andrew Rowley, Jacob Carter and Sophia Ananiadou 115

DKPro Similarity: An Open Source Framework for Text Similarity
Daniel Bär, Torsten Zesch and Iryna Gurevych 121

Fluid Construction Grammar for Historical and Evolutionary Linguistics
Pieter Wellens, Remi van Trijp, Katrien Beuls and Luc Steels 127

HYENA-live: Fine-Grained Online Entity Type Classification from Natural-language Text
Mohamed Amir Yosef, Sandro Bauer, Johannes Hoffart, Marc Spaniol and Gerhard Weikum . . 133

Linggle: a Web-scale Linguistic Search Engine for Words in Context
Joanne Boisson, Ting-Hui Kao, Jian-Cheng Wu, Tzu-Hsi Yen and Jason S. Chang 139

ParaQuery: Making Sense of Paraphrase Collections
Lili Kotlerman, Nitin Madnani and Aoife Cahill 145

PATHS: A System for Accessing Cultural Heritage Collections
Eneko Agirre, Nikolaos Aletras, Paul Clough, Samuel Fernando, Paula Goodale, Mark Hall, Aitor Soroa and Mark Stevenson 151

Propminer: A Workflow for Interactive Information Extraction and Exploration using Dependency Trees
Alan Akbik, Oresti Konomi and Michail Melnikov 157

SEMILAR: The Semantic Similarity Toolkit
Vasile Rus, Mihai Lintean, Rajendra Banjade, Nobal Niraula and Dan Stefanescu 163

Tag2Blog: Narrative Generation from Satellite Tag Data
Kapila Ponnamperuma, Advaith Siddharthan, Cheng Zeng, Chris Mellish and René van der Wal 169

TransDoop: A Map-Reduce based Crowdsourced Translation for Complex Domain
Anoop Kunchukuttan, Rajen Chatterjee, Shourya Roy, Abhijit Mishra and Pushpak Bhattacharyya 175

tSEARCH: Flexible and Fast Search over Automatic Translations for Improved Quality/Error Analysis
Meritxell Gonzàlez, Laura Mascarell and Lluís Màrquez 181

VSEM: An open library for visual semantics representation
Elia Bruni, Ulisse Bordignon, Adam Liska, Jasper Uijlings and Irina Sergienya 187

Docent: A Document-Level Decoder for Phrase-Based Statistical Machine Translation
Christian Hardmeier, Sara Stymne, Jörg Tiedemann and Joakim Nivre 193

Mr. MIRA: Open-Source Large-Margin Structured Learning on MapReduce
Vladimir Eidelman, Ke Wu, Ferhan Ture, Philip Resnik and Jimmy Lin 199

Conference Program

Monday August 5, 2013

(18:30 - 19:45) System Demonstrations A

WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations
Seid Muhie Yimam, Iryna Gurevych, Richard Eckart de Castilho and Chris Biemann

A Stacking-based Approach to Twitter User Geolocation Prediction
Bo Han, Paul Cook and Timothy Baldwin

An Open Source Toolkit for Quantitative Historical Linguistics
Johann-Mattis List and Steven Moran

AnnoMarket: An Open Cloud Platform for NLP
Valentin Tablan, Kalina Bontcheva, Ian Roberts, Hamish Cunningham and Marin Dimitrov

Detecting Event-Related Links and Sentiments from Social Media Texts
Alexandra Balahur and Hristo Tanev

DISSECT - DIStributional SEmantics Composition Toolkit
Georgiana Dinu, Nghia The Pham and Marco Baroni

DKPro WSD: A Generalized UIMA-based Framework for Word Sense Disambiguation
Tristan Miller, Nicolai Erbs, Hans-Peter Zorn, Torsten Zesch and Iryna Gurevych

Extending an interoperable platform to facilitate the creation of multilingual and multimodal NLP applications
Georgios Kontonatsios, Paul Thompson, Riza Theresa Batista-Navarro, Claudiu Mihăilă, Ioannis Korkontzelos and Sophia Ananiadou

FudanNLP: A Toolkit for Chinese Natural Language Processing
Xipeng Qiu, Qi Zhang and Xuanjing Huang

ICARUS – An Extensible Graphical Search Tool for Dependency Treebanks
Markus Gärtner, Gregor Thiele, Wolfgang Seeker, Anders Björkelund and Jonas Kuhn

Meet EDGAR, a tutoring agent at MONSERRATE
Pedro Fialho, Luísa Coheur, Sérgio Curto, Pedro Cláudio, Ângela Costa, Alberto Abad, Hugo Meinedo and Isabel Trancoso

Monday August 5, 2013 (continued)

PAL: A Chatterbot System for Answering Domain-specific Questions
Yuanchao Liu, Ming Liu, Xiaolong Wang, Limin Wang and Jingjing Li

PhonMatrix: Visualizing co-occurrence constraints of sounds
Thomas Mayer and Christian Rohrdantz

QuEst - A translation quality estimation framework
Lucia Specia, , , Kashif Shah, Jose G.C. de Souza and Trevor Cohn

SORT: An Interactive Source-Rewriting Tool for Improved Translation
Shachar Mirkin, Sriram Venkatapathy, Marc Dymetman and Ioan Calapodescu

Travatar: A Forest-to-String Machine Translation Engine based on Tree Transducers
Graham Neubig

PLIS: a Probabilistic Lexical Inference System
Eyal Shnarch, Erel Segal-haLevi, Jacob Goldberger and Ido Dagan

(19:45 - 21:00) System Demonstrations B

A Java Framework for Multilingual Definition and Hypernym Extraction
Stefano Faralli and Roberto Navigli

A Visual Analytics System for Cluster Exploration
Andreas Lamprecht, Annette Hautli, Christian Rohrdantz and Tina Bögel

Development and Analysis of NLP Pipelines in Argo
Rafal Rak, Andrew Rowley, Jacob Carter and Sophia Ananiadou

DKPro Similarity: An Open Source Framework for Text Similarity
Daniel Bär, Torsten Zesch and Iryna Gurevych

Fluid Construction Grammar for Historical and Evolutionary Linguistics
Pieter Wellens, Remi van Trijp, Katrien Beuls and Luc Steels

HYENA-live: Fine-Grained Online Entity Type Classification from Natural-language Text
Mohamed Amir Yosef, Sandro Bauer, Johannes Hoffart, Marc Spaniol and Gerhard Weikum

Monday August 5, 2013 (continued)

Linggle: a Web-scale Linguistic Search Engine for Words in Context
Joanne Boisson, Ting-Hui Kao, Jian-Cheng Wu, Tzu-Hsi Yen and Jason S. Chang

ParaQuery: Making Sense of Paraphrase Collections
Lili Kotlerman, Nitin Madnani and Aoife Cahill

PATHS: A System for Accessing Cultural Heritage Collections
Eneko Agirre, Nikolaos Aletras, Paul Clough, Samuel Fernando, Paula Goodale, Mark Hall, Aitor Soroa and Mark Stevenson

Propminer: A Workflow for Interactive Information Extraction and Exploration using Dependency Trees
Alan Akbik, Oresti Konomi and Michail Melnikov

SEMILAR: The Semantic Similarity Toolkit
Vasile Rus, Mihai Lintean, Rajendra Banjade, Nobal Niraula and Dan Stefanescu

Tag2Blog: Narrative Generation from Satellite Tag Data
Kapila Ponnamperuma, Advaith Siddharthan, Cheng Zeng, Chris Mellish and René van der Wal

TransDoop: A Map-Reduce based Crowdsourced Translation for Complex Domain
Anoop Kunchukuttan, Rajen Chatterjee, Shourya Roy, Abhijit Mishra and Pushpak Bhattacharyya

tSEARCH: Flexible and Fast Search over Automatic Translations for Improved Quality/Error Analysis
Meritxell Gonzàlez, Laura Mascarell and Lluís Màrquez

VSEM: An open library for visual semantics representation
Elia Bruni, Ulisse Bordignon, Adam Liska, Jasper Uijlings and Irina Sergienya

Docent: A Document-Level Decoder for Phrase-Based Statistical Machine Translation
Christian Hardmeier, Sara Stymne, Jörg Tiedemann and Joakim Nivre

Mr. MIRA: Open-Source Large-Margin Structured Learning on MapReduce
Vladimir Eidelman, Ke Wu, Ferhan Ture, Philip Resnik and Jimmy Lin