

COLING 2014

**The 25th International Conference  
on Computational Linguistics**

**Proceedings of the Conference  
System Demonstrations**

**Editors**

**Dr. Lamia Tounsi, CNGL, Dublin City University  
Dr. Rafal Rak, NaCTeM, University of Manchester**

August 23-29, 2014  
Dublin, Ireland

©2014 The Authors

The papers in this volume are licensed by the authors under a Creative Commons Attribution 4.0 International License.

ISBN 978-1-941643-27-3

## Preface

This volume contains papers from the system demonstration session of the 25th International Conference on Computational Linguistics (COLING 2014) held in Dublin, Ireland. The conference is organized by the Centre for Global Intelligent Content (CNGL) and held at the Helix Conference Centre at Dublin City University (DCU) from 25 to 29 August 2014, under the auspices of the International Committee on Computational Linguistics (ICCL).

The demonstration session complements the conference's presentation and poster sessions and is focused on working software systems that are the tangible outcomes of research on computational linguistics.

As a result of a rigorous review process, we accepted 28 papers out of 45 submissions. The program committee consisted of 35 members and two chairs from both academia and industry. Each member evaluated two or three papers, which amounted to two reviews per paper. The acceptance criteria we followed during the selection process included the quality of work as well as the utility and demonstrability potential of the presented systems. Consequently, most of the accepted systems are user-interactive and feature rich graphical user interfaces.

First and foremost we would like to thank the program committee for their hard work and dedication to help make this event a success. Our special thanks also go to the people who made COLING 2014 and this volume possible. We thank Programme Chairs, Prof. Junichi Tsujii (Microsoft Research) and Prof. Jan Hajic (Charles University), General Chairs, Prof. Josef van Genabith (Universität des Saarlandes/DFKI) and Professor Andy Way (CNGL, DCU), the chairs of the local organizing committee, Dr. Cara Green (CNGL, DCU) and Dr. John Judge (CNGL/NCLT, DCU), and Publications Chairs, Dr. Joachim Wagner (CNGL, DCU), Dr. Liadh Kelly (CNGL, DCU) and Dr. Lorraine Goeriot (CNGL, DCU), for their tireless work.

Lamia Tounsi and Rafal Rak

COLING 2014 Demonstration Programme Co-chairs

25 July 2014



## **Organizers:**

### **Demonstration Chairs**

Lamia Tounsi, CNGL, Dublin City University, Ireland  
Rafal Rak, NaCTeM, University of Manchester, UK

### **Program Committee:**

Michiel Bacchiani, Google Inc.  
Kay Berkling, Cooperative State University, Karlsruhe  
Ann Bies, Linguistic Data Consortium  
William Black, University of Manchester  
Francis Bond, Nanyang Technological University  
Chris Brew, Nuance Communications  
Aoife Cahill, Educational Testing Service  
Vittorio Castelli, IBM  
Md. Faisal Mahbub Chowdhury, IBM  
Léa Deleris, IBM  
Martin Emms, Trinity College Dublin  
Guillaume Gravier, IRISA and INRIA Rennes  
Keith Hall, Google Research  
Derrick Higgins, Educational Testing Service  
Keikichi Hirose, University of Tokyo  
Frank Hopfgartner, Technische Universität Berlin  
Daxin Jiang, Microsoft STC-A  
John Kelleher, Trinity College Dublin  
Adam Kilgarriff, Lexical Computing Ltd  
BalaKrishna Kolluru, Toshiba  
Seamus Lawless, Trinity College Dublin  
Saturnino Luz, Trinity College Dublin  
Nitin Madnani, Educational Testing Service  
Hilary McDonald, Trinity College Dublin  
Helen Meng, Chinese University of Hong Kong  
Peter Mika, Yahoo Labs  
Tony O'Dowd, KantanMT  
Florian Pinel, IBM  
Johann Roturier, Symantec  
Andrew Rowley, University of Manchester  
Frédérique Segond, Viseo Research  
Swapna Somasundaran, Educational Testing Services  
Tomoki Toda, Nara Institute of Science and Technology  
Xinglong Wang, Brandwatch  
Jason Williams, Microsoft Research



## Table of Contents

<i>An Error Analysis Tool for Natural Language Processing and Applied Machine Learning</i> Apoorv Agarwal, Ankit Agarwal and Deepak Mittal .....	1
<i>Claims on demand – an initial demonstration of a system for automatic detection and polarity identification of context dependent claims in massive corpora</i> Noam Slonim, Ehud Aharoni, Carlos Alzate, Roy Bar-Haim, Yonatan Bilu, Lena Dankin, Iris Eiron, Daniel Hershcovich, Shay Hummel, Mitesh Khapra, Tamar Lavee, Ran Levy, Paul Matchen, Anatoly Polnarov, Vikas Raykar, Ruty Rinott, Amrita Saha, Naama Zwerdling, David Konopnicki and Dan Gutfreund .....	6
<i>Copa 2014 FrameNet Brasil: a frame-based trilingual electronic dictionary for the Football World Cup</i> Tiago Torrent, Maria Margarida Salomão, Fernanda Campos, Regina Braga, Ely Matos, Maucha Gamonal, Julia Gonçalves, Bruno Souza, Daniela Gomes and Simone Peron .....	10
<i>Creating Custom Taggers by Integrating Web Page Annotation and Machine Learning</i> Srikrishna Raamadhurai, Oskar Kohonen and Teemu Ruokolainen .....	15
<i>How to deal with students’ writing problems? Process-oriented writing support with the digital Writing Aid Dutch</i> Lieve De Wachter, Serge Verlinde, Margot D’Hertefelt and Geert Peeters .....	20
<i>Processing Discourse in Dislog on the TextCoop Platform</i> patrick saint-dizier .....	25
<i>UIMA Ruta Workbench: Rule-based Text Annotation</i> Peter Kluegl, Martin Toepfer, Philip-Daniel Beck, Georg Fette and Frank Puppe .....	29
<i>Discourse Relations in the Prague Dependency Treebank 3.0</i> Jiří Mírovský, Pavlína Jínová and Lucie Poláková .....	34
<i>Lightweight Client-Side Chinese/Japanese Morphological Analyzer Based on Online Learning</i> Masato Hagiwara and Satoshi Sekine .....	39
<i>MultiDPS – A multilingual Discourse Processing System</i> Daniel Anechitei .....	44
<i>Sanskrit Linguistics Web Services</i> Gérard Huet and Amba Kulkarni .....	48
<i>TICCLops: Text-Induced Corpus Clean-up as online processing system</i> Martin Reynaert .....	52
<i>Trameur: A Framework for Annotated Text Corpora Exploration</i> Serge Fleury and Maria Zimina .....	57
<i>TweetGenie: Development, Evaluation, and Lessons Learned</i> Dong Nguyen, Dolf Trieschnigg and Theo Meder .....	62
<i>A Sentence Judgment System for Grammatical Error Detection</i> Lung-Hao Lee, Liang-Chih Yu, Kuei-Ching Lee, Yuen-Hsien Tseng, Li-Ping Chang and Hsin-Hsi Chen .....	67

<i>CLAM: Quickly deploy NLP command-line tools on the web</i> Maarten van Gompel and Martin Reynaert .....	71
<i>CRAB 2.0: A text mining tool for supporting literature review in chemical cancer risk assessment</i> Yufan Guo, Diarmuid Ó Séaghdha, Ilona Silins, Lin Sun, Johan Högberg, Ulla Stenius and Anna Korhonen .....	76
<i>Nerdle: Topic-Specific Question Answering Using Wikia Seeds</i> Umar Maqsood, Sebastian Arnold, Michael Hülfenhaus and Alan Akbik .....	81
<i>NTU-MC Toolkit: Annotating a Linguistically Diverse Corpus</i> Liling Tan and Francis Bond .....	86
<i>RDF Triple Stores and a Custom SPARQL Front-End for Indexing and Searching (Very) Large Semantic Networks</i> Milen Kouylekov and Stephan Oepen .....	90
<i>What or Who is Multilingual Watson?</i> Keith Cortis, Urvesh Bhowan, Ronan Mac an tSaoir, D.J. McCloskey, Mikhail Sogrin and Ross Cadogan .....	95
<i>A Marketplace for Web Scale Analytics and Text Annotation Services</i> Johannes Kirschnick, Torsten Kiliyas, Holmer Hensen, Alexander Löser, Peter Adolphs, Heiko Ehrig and Holger Düwiger .....	100
<i>DKPro Agreement: An Open-Source Java Library for Measuring Inter-Rater Agreement</i> Christian M. Meyer, Margot Mieskes, Christian Stab and Iryna Gurevych .....	105
<i>Distributional Semantics in R with the wordspace Package</i> Stefan Evert .....	110
<i>Method51 for Mining Insight from Social Media Datasets</i> Simon Wibberley, David Weir and Jeremy Reffin .....	115
<i>MT-EQuAl: a Toolkit for Human Assessment of Machine Translation Output</i> Christian Girardi, Luisa Bentivogli, Mohammad Amin Farajian and Marcello Federico .....	120
<i>OpenSoNaR: user-driven development of the SoNaR corpus interfaces</i> Martin Reynaert, Matje van de Camp and Menno van Zaanen .....	124
<b>THE MATECAT TOOL</b> Marcello Federico, Nicola Bertoldi, Mauro Cettolo, Matteo Negri, Marco Turchi, Marco Trombetti, Alessandro Cattelan, Antonio Farina, Domenico Lupinetti, Andrea Martines, Alberto Massidda, Holger Schwenk, Loïc Barrault, Frederic Blain, Philipp Koehn, Christian Buck and Ulrich Germann .....	129



# Conference Program

25/08/2014

## (10:15 - 12:25) Demo session 1

*An Error Analysis Tool for Natural Language Processing and Applied Machine Learning*

Apoorv Agarwal, Ankit Agarwal and Deepak Mittal

*Claims on demand – an initial demonstration of a system for automatic detection and polarity identification of context dependent claims in massive corpora*

Noam Slonim, Ehud Aharoni, Carlos Alzate, Roy Bar-Haim, Yonatan Bilu, Lena Dankin, Iris Eiron, Daniel Hershcovich, Shay Hummel, Mitesh Khapra, Tamar Lavee, Ran Levy, Paul Matchen, Anatoly Polnarov, Vikas Raykar, Ruty Rinott, Amrita Saha, Naama Zwerdling, David Konopnicki and Dan Gutfreund

*Copa 2014 FrameNet Brasil: a frame-based trilingual electronic dictionary for the Football World Cup*

Tiago Torrent, Maria Margarida Salomão, Fernanda Campos, Regina Braga, Ely Matos, Maucha Gamonal, Julia Gonçalves, Bruno Souza, Daniela Gomes and Simone Peron

*Creating Custom Taggers by Integrating Web Page Annotation and Machine Learning*

Srikrishna Raamadhurai, Oskar Kohonen and Teemu Ruokolainen

*How to deal with students' writing problems? Process-oriented writing support with the digital Writing Aid Dutch*

Lieve De Wachter, Serge Verlinde, Margot D'Hertefelt and Geert Peeters

*Processing Discourse in Dislog on the TextCoop Platform*

patrick saint-dizier

*UIMA Ruta Workbench: Rule-based Text Annotation*

Peter Kluegl, Martin Toepfer, Philip-Daniel Beck, Georg Fette and Frank Puppe

## (15:15 - 17:25) Demo session 2

*Discourse Relations in the Prague Dependency Treebank 3.0*

Jiří Mírovský, Pavlína Jínová and Lucie Poláková

*Lightweight Client-Side Chinese/Japanese Morphological Analyzer Based on Online Learning*

Masato Hagiwara and Satoshi Sekine

*MultiDPS – A multilingual Discourse Processing System*

Daniel Anechitei

**25/08/2014 (continued)**

*Sanskrit Linguistics Web Services*

Gérard Huet and Amba Kulkarni

*TICCLops: Text-Induced Corpus Clean-up as online processing system*

Martin Reynaert

*Trameur: A Framework for Annotated Text Corpora Exploration*

Serge Fleury and Maria Zimina

*TweetGenie: Development, Evaluation, and Lessons Learned*

Dong Nguyen, Dolf Trieschnigg and Theo Meder

**26/08/2014**

**(10:15 - 12:25) Demo session 3**

*A Sentence Judgment System for Grammatical Error Detection*

Lung-Hao Lee, Liang-Chih Yu, Kuei-Ching Lee, Yuen-Hsien Tseng, Li-Ping Chang and Hsin-Hsi Chen

*CLAM: Quickly deploy NLP command-line tools on the web*

Maarten van Gompel and Martin Reynaert

*CRAB 2.0: A text mining tool for supporting literature review in chemical cancer risk assessment*

Yufan Guo, Diarmuid Ó Séaghdha, Ilona Silins, Lin Sun, Johan Högberg, Ulla Stenius and Anna Korhonen

*Nerdle: Topic-Specific Question Answering Using Wikia Seeds*

Umar Maqsud, Sebastian Arnold, Michael Hülftenhaus and Alan Akbik

*NTU-MC Toolkit: Annotating a Linguistically Diverse Corpus*

Liling Tan and Francis Bond

*RDF Triple Stores and a Custom SPARQL Front-End for Indexing and Searching (Very) Large Semantic Networks*

Milen Kouylekov and Stephan Oepen

*What or Who is Multilingual Watson?*

Keith Cortis, Urvesh Bhowan, Ronan Mac an tSaoir, D.J. McCloskey, Mikhail Sogrin and Ross Cadogan

26/08/2014 (continued)

**(15:15 - 17:25) Demo Session 4**

*A Marketplace for Web Scale Analytics and Text Annotation Services*

Johannes Kirschnick, Torsten Kiliyas, Holmer Hemsén, Alexander Löser, Peter Adolphs, Heiko Ehrig and Holger Düwiger

*DKPro Agreement: An Open-Source Java Library for Measuring Inter-Rater Agreement*

Christian M. Meyer, Margot Mieskes, Christian Stab and Iryna Gurevych

*Distributional Semantics in R with the wordSpace Package*

Stefan Evert

*Method51 for Mining Insight from Social Media Datasets*

Simon Wibberley, David Weir and Jeremy Reffin

*MT-EQuAl: a Toolkit for Human Assessment of Machine Translation Output*

Christian Girardi, Luisa Bentivogli, Mohammad Amin Farajian and Marcello Federico

*OpenSoNaR: user-driven development of the SoNaR corpus interfaces*

Martin Reynaert, Matje van de Camp and Menno van Zaanen

***THE MATECAT TOOL***

Marcello Federico, Nicola Bertoldi, Mauro Cettolo, Matteo Negri, Marco Turchi, Marco Trombetti, Alessandro Cattelan, Antonio Farina, Domenico Lupinetti, Andrea Martines, Alberto Massidda, Holger Schwenk, Loïc Barrault, Frederic Blain, Philipp Koehn, Christian Buck and Ulrich Germann

