

INTERNATIONAL CONFERENCE

RECENT ADVANCES IN

NATURAL LANGUAGE PROCESSING

P R O C E E D I N G S

Edited by
Galia Angelova, Kalina Bontcheva, Ruslan Mitkov

Hissar, Bulgaria
9–11 September, 2013

INTERNATIONAL CONFERENCE
RECENT ADVANCES IN
NATURAL LANGUAGE PROCESSING'2013

PROCEEDINGS

Hissar, Bulgaria
7—13 September 2013

ISSN 1313-8502

Designed and Printed by INCOMA Ltd.
Shoumen, BULGARIA

Preface

Welcome to the 9th International Conference on “Recent Advances in Natural Language Processing” (RANLP 2013) in Hissar, Bulgaria, 9–11 September 2013. The main objective of the conference is to give researchers the opportunity to present new results in Natural Language Processing (NLP) based on modern theories and methodologies.

The conference is preceded by two days of tutorials (7-8 September 2013) and the lecturers are:

- Preslav Nakov (Qatar Computing Research Institute, Qatar Foundation)
- Vivi Nastase (Fondazione Bruno Kessler)
- Diarmuid Ó Séaghdha (Cambridge University)
- Stan Szpakowicz (University of Ottawa)
- Iryna Gurevych (Technical University Darmstadt)
- Judith Eckle-Kohler (Technical University Darmstadt)
- Violeta Seretan (University of Geneva)
- Dekai Wu (Hong Kong University of Science & Technology)

The conference keynote speakers are:

- Nicoletta Calzolari (Institute of Computational Linguistics “Antonio Zampolli”, Pisa)
- Iryna Gurevych (Technical University Darmstadt)
- Horacio Saggion (University Pompeu Fabra, Barcelona)
- Violeta Seretan (University of Geneva)
- Mark Stevenson (University of Sheffield)
- Dekai Wu (Hong Kong University of Science & Technology)

This year 22 regular papers, 36 short papers, and 41 posters have been accepted for presentation at the conference. In 2013 RANLP hosts 3 workshops on influential NLP topics, such as NLP for medicine and biology, Linked Open Data (LOD) for NLP, semantic web and information extraction, and adaptation of language resources.

The proceedings cover a wide variety of NLP topics: part of speech tagging, language resources, semantics, opinion mining and sentiment analysis, multilingual NLP, language modelling, word sense disambiguation, information extraction, term extraction, parsing, text summarisation, machine translation, question answering, temporal processing, text simplification, named entity recognition, text generation, text categorisation, NLP for special languages, morphology and syntax, etc.

We would like to thank all members of the Programme Committee and all additional reviewers. Together they have ensured that the best papers were included in the proceedings and have provided invaluable comments for the authors.

Finally, special thanks go to the University of Wolverhampton, the Bulgarian Academy of Sciences, the ACOMIN European project, Ontotext, the Association for Computational Linguistics – Bulgaria for their generous support for RANLP.

Welcome to Hissar and we hope that you enjoy the conference!

The RANLP 2013 Organisers

The International Conference RANLP–2013 is organised by:

Research Group in Computational Linguistics, University of Wolverhampton, UK

Linguistic Modelling Department, Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences, Bulgaria

RANLP–2013 is partially supported by:

AComIn (Advanced Computing for Innovation, FP7 Capacity grant 316087)

Ontotext AD

Programme Committee Chair:

Ruslan Mitkov, University of Wolverhampton

Organising Committee Chair:

Galia Angelova, Bulgarian Academy of Sciences

Workshop Coordinator:

Kiril Simov, Bulgarian Academy of Sciences

Publication Chair:

Kalina Bontcheva, University of Sheffield

Tutorial Coordinator:

Preslav Nakov, Qatar Computing Research Institute

Proceedings Printing:

Nikolai Nikolov, Association for Computational Linguistics, Bulgaria

Programme Committee Coordinators:

Ivelina Nikolova, Bulgarian Academy of Sciences

Irina Temnikova, Bulgarian Academy of Sciences

Natalia Konstantinova, University of Wolverhampton

Program Committee:

Guadalupe Aguado de Cea (Polytechnic University Madrid, Spain)
Roberto Basili (University of Roma, Tor Vergata, Italy)
Jerome Bellegarda (Apple Inc., USA)
Chris Biemann (TU Darmstadt, Germany)
Kalina Bontcheva (University of Sheffield, UK)
Svetla Boytcheva (American University in Bulgaria, Bulgaria)
António Branco (University of Lisbon, Portugal)
Jill Burstein (Educational Testing Service, USA)
Nicoletta Calzolari (National Research Council, Italy)
Kevin Bretonnel Cohen (University of Colorado School of Medicine, USA)
Ken Church (The Johns Hopkins University, IBM Research, USA)
Dan Cristea (“Al. I. Cuza” University of Iasi, Romania)
Ido Dagan (Bar Ilan University, Israel)
Anne De Roeck (The Open University, UK)
Richard Evans (University of Wolverhampton, UK)
Antonio Ferrández Rodríguez (University of Alicante, Spain)
Joey Frazee (University of Texas at Austin, USA)
Fumiyo Fukumoto (Yamanashi University, Japan)
Alexander Gelbukh (Nat. Polytechnic Inst., Mexico)
Ralph Grishman (New York University, USA) Patrick Hanks (University of the West of England and University of Wolverhampton, UK)
Kris Heylen (University of Leuven, Belgium)
Graeme Hirst (Univ. of Toronto, Canada)
Veronique Hoste (University College Ghent, Belgium)
Mans Hulden (University of Helsinki, Finland)
Diana Inkpen (University of Ottawa, Canada)
Hitoshi Isahara (Toyohashi University of Technology, Japan)
Ali Jaoua (Qatar University, Qatar)
Mijail Kabadjov (DaXtra Technologies Ltd., UK)
Dimitar Kazakov (University of York, UK)
Alma Kharrat (Microsoft, USA)
Udo Kruschwitz (University of Essex, UK)
Hristo Krushkov (University of Plovdiv, Bulgaria)
Sandra Kuebler (Indiana University, USA)
Lori Lamel (LIMSI - CNRS, France)
Chew Lim Tan (National University of Singapore, Singapore)
Qun Liu (Chinese Academy of Sciences, China)
Suresh Manandhar (University of York, UK)
Yusuke Miyao (National Institute of Informatics, Japan)
Johanna Monti (University of Sassari, Italy)
Alessandro Moschitti (University of Trento, Italy)
Rafael Muñoz Guillena (University of Alicante, Spain)
Preslav Nakov (QCRI, Qatar)
Roberto Navigli (University di Roma La Sapienza, Italy)
Vincent Ng (The University of Texas at Dallas, USA)
Kemal Oflazer (Carnegie Mellon University, Qatar)
Constantin Orasan (University of Wolverhampton, UK)

Sebastian Pado (University of Heidelberg, Germany)
Karel Pala (Masaryk University, Czech Republic)
Martha Palmer (University of Colorado, USA)
Stelios Piperidis (ILSP, Greece)
Simone Paolo Ponzetto (University of Heidelberg, Germany)
Gábor Prószéky (Pázmány University & MorphoLogic, Hungary)
Allan Ramsay (Univ. of Manchester, UK)
Horacio Rodriguez (Universitat Politècnica de Catalunya, Spain)
Paolo Rosso (University of Valencia, Spain)
Vasile Rus (University of Memphis, USA)
Horacio Saggion (Universitat Pompeu Fabra, Spain)
Patrick Saint-Dizier (IRIT-CNRS, France)
Satoshi Sakine (New York University, USA)
Doaa Samy (University Autonomous of Madrid, Spain)
Violeta Seretan (University of Geneva, Switzerland)
Khaled Shaalan (Cairo University, Egypt)
Kiril Simov (Bulgarian Academy of Sciences, Bulgaria)
Keh-Yih Su (Behavior Design Corp., Taiwan)
Stan Szpakowicz (University of Ottawa, Canada)
John Tait (Johntait.net Limited)
Josef van Genabith (Dublin City University, Ireland)
Dan Tufis (RIAI, Romanian Academy, Romania)
L. Alfonso Ureña López (University of Jaen, Spain)
Paola Velardi (University of Roma “La Sapienza”, Italy)
Suzan Verberne (Radboud University Nijmegen, The Netherlands)
Piek Vossen (VU University Amsterdam, The Netherlands)
Yorick Wilks (Univ. of Sheffield, UK)
Dekai Wu (HKUST, Hong Kong)
Torsten Zesch (TU Darmstadt, Germany)
Min Zhang (University of Michigan, USA)

Additional Reviewers:

Karteek Addanki (HKUST, Hong Kong)
Itziar Aldabe (Univ. of Basque Country, Spain)
Hadi Amiri (National University of Singapore)
Marilisa Amoia (Saarland University, Germany)
Wilker Aziz (University of Wolverhampton, UK)
Nguyen Bach (Carnegie Mellon University, USA)
Daniel Bär (TU Darmstadt, Germany)
Eduard Barbu (Universiy of Jaén, Spain)
Leonor Becerra (Laboratoire Hubert Curien, France)
Cosmin Bejan (University of Washington, USA)
Asma Ben Abacha (CRP Henri Tudor, Luxembourg)
Boryana Bratanova (University of Veliko Turnovo, Bulgaria)
Erik Cambria (National University of Singapore, Singapore)
Marie Candito (Univ Paris Diderot - INRIA, France)
Miranda Chong (University of Wolverhampton, UK)
Marta R. Costa-Jussa (Barcelona Media Innovation Center, Spain)

Eugeniu Costetchi (CRP Henri Tudor, Luxembourg)
Raquel Criado (University of Murcia, Spain)
Noa Cruz (University of Huelva, Spain)
Daniel Dahlmeier (National University of Singapore, Singapore)
Kareem Darwish (QCRI, Qatar Foundation, Qatar)
Orphee De Clercq (University College Ghent, Belgium)
Gerard de Melo (ICSI Berkeley, USA)
Leon Derczynski (University of Sheffield, UK)
Liviu Dinu (University of Bucharest, Romania)
Son Doan (UC San Diego, USA)
Iustin Dornescu (University of Wolverhampton, UK)
Brett Drury (LIAAD-INESC, Portugal)
Kevin Duh (Nara Institute of Science and Technology, Japan)
Isabel Durán Muñoz (University of Wolverhampton, UK)
Chris Dyer (Carnegie Mellon University, USA)
Ismail El Maarouf (University of Wolverhampton, UK)
Maria Eskevich (Dublin City University, Ireland)
Mariano Felice (Cambridge University, UK)
Mark Fishel (University of Zurich, Switzerland)
Wei Gao (QCRI, Qatar Foundation, Qatar)
Albert Gatt (University of Malta, Malta)
Matthew Gerber (University of Virginia, USA)
Goran Glavaš (University of Zagreb, Croatia)
José Miguel Goñi-Menoyo (Politechnical University of Madrid, Spain)
Brian Harrington (University of Toronto Scarborough, Canada)
Laura Hasler (University of Strathclyde, UK)
Hany Hassan (Microsoft Research, USA)
Kai Hong (University of Pennsylvania, USA)
Ales Horak (Masaryk University, Czech Republic)
Young-Sook Hwang (SK Telecom, South Korea)
Iustina Ilisei (University of Wolverhampton, UK)
Sujay Kumar Jauhar (Carnegie Mellon University, USA)
Minwoo Jeong (Microsoft, USA)
Kristiina Jokinen (University of Helsinki, Finland)
David Kauchak (Middlebury College, USA)
Jin-Dong Kim (Database Center for Life Science, Japan)
Natalia Konstantinova (University of Wolverhampton, UK)
Zornitsa Kozareva (USC Information Sciences Institute, USA)
Laska Laskova (Sofia University, Bulgaria)
Junyi Li (University of Pennsylvania, USA)
Maria Liakata (University of Warwick, UK)
Ting Liu (Google, USA)
Elena Lloret (University of Alicante, Spain)
Chi-kiu LO (HKUST, Hong Kong)
Oier Lopez de Lacalle (Basque Foundation for Science, Spain and University of Edinburgh, Scotland)
Annie Louis (University of Pennsylvania, USA)
Wei Lu (University of Illinois at Urbana-Champaign, USA)

Yapomo Manuela (University of Strasbourg, France)
Maite Martin (Univeristy of Jaén, Spain)
Eugenio Martinez-Camara (University of Jaén, Spain)
Bonan Min (New York University, USA)
Wolfgang Minker (Ulm University, Germany)
Olga Mitrofanova (St. Petersburg State University, Russia)
Makoto Miwa (National Centre for Text Mining, University of Manchester, UK)
Behrang Mohit (Carnegie Mellon University, Qatar)
Michael Mohler (University of North-Texas, USA)
Manuel Montes (INAOE, Mexico)
Vlad Niculae (University of Wolverhampton, UK)
Ivelina Nikolova (Bulgarian Academy of Sciences, Bulgaria)
Petya Osenova (Sofia University and ICT-BAS, Bulgaria)
Diarmuid Ó Séaghdha (University of Cambridge, UK)
Georgios Paltoglou (University of Wolverhampton, UK)
Alexander Panchenko (Universite catholique de Louvain, Belgium)
Katherin Pérez (University of Wolverhampton, UK)
Vinodkumar Prabhakaran (Columbia University, USA)
Carlos Ramisch (Université Joseph Fourier, France)
Luz Rello (Universitat Pompeu Fabra, Spain)
Miguel Angel Rios Gaona (University of Wolverhampton, UK)
Raphael Rubino (Dublin City University, Symantec, Ireland)
Pavel Rychlý (Masaryk University, Czech Republic)
Gerold Schneider (University of Zurich, Switzerland)
Lane Schwartz (Air Force Research Laboratory, USA)
Avirup Sil (Temple University, USA)
Yvonne Skalban (University of Wolverhampton, UK)
Jan Snajder (University of Zagred, Croatia)
Sanja Stajner (University of Wolverhampton, UK)
Ekaterina Stambolieva (euroscript Luxembourg S.à. r.l., Luxembourg)
Sebastian Stüker (Karlsruhe Institute of Technology)
Ang Sun (inome Inc, USA)
Yoshimi Suzuki (University of Yamanashi, Japan)
Irina Temnikova (Bulgarian Academy of Sciences, Bulgaria)
Joel Tetreault (Nuance Communications, USA)
Katerina Raisa Timonera (University of Wolverhamtpon, UK)
Maria Cristina Toledo Baez (University of Murcia, Spain)
Marco Turchi (Fondazione Bruno Kessler, Italy)
Paola Valli (University of Trieste, Italy)
Andrea Varga (The University Of Sheffield, UK)
Aline Villavicencio (Federal University of Rio Grande do Sul, Brazil)
Veronika Vincze (University of Szeged, Hungary)
Haifeng Wang (Baidu, China)
Stephanie Weiser (Knowbel Technologies, Belgium)
Sandra Williams (The Open University, UK)
Victoria Yaneva (University of Wolverhampton, UK)
Heng Yu (Chinese Academy of Sciences, China)
Wajdi Zaghouni (Carnegie Mellon University, Qatar)

Table of Contents

<i>ASMA: A System for Automatic Segmentation and Morpho-Syntactic Disambiguation of Modern Standard Arabic</i>	
Muhammad Abdul-Mageed, Mona Diab and Sandra Kübler	1
<i>Optimising Tree Edit Distance with Subtrees for Textual Entailment</i>	
Maytham Alabbas and Allan Ramsay	9
<i>Opinion Learning from Medical Forums</i>	
Tanveer Ali, Marina Sokolova, David Schramm and Diana Inkpen	18
<i>Annotating events, Time and Place Expressions in Arabic Texts</i>	
Hassina Aliane, Wassila Guendouzi and Amina Mokrani	25
<i>A Semi-supervised Learning Approach to Arabic Named Entity Recognition</i>	
Maha Althobaiti, Udo Kruschwitz and Massimo Poesio	32
<i>An NLP-based Reading Tool for Aiding Non-native English Readers</i>	
Mahmoud Azab, Ahmed Salama, Kemal Oflazer, Hideki Shima, Jun Araki and Teruko Mitamura	41
<i>Improving Sentiment Analysis in Twitter Using Multilingual Machine Translated Data</i>	
Alexandra Balahur and Marco Turchi	49
<i>Domain Adaptation for Parsing</i>	
Eric Baucom, Levi King and Sandra Kübler	56
<i>Towards a Structured Representation of Generic Concepts and Relations in Large Text Corpora</i>	
Archana Bhattarai and Vasile Rus	65
<i>Authorship Attribution in Health Forums</i>	
Victoria Bobicev, Marina Sokolova, Khaled El Emam and Stan Matwin	74
<i>TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text</i>	
Kalina Bontcheva, Leon Derczynski, Adam Funk, Mark Greenwood, Diana Maynard and Niraj Aswani	83
<i>A unified lexical processing framework based on the Margin Infused Relaxed Algorithm. A case study on the Romanian Language</i>	
Tiberiu Boros	91
<i>Automatic extraction of contextual valence shifters.</i>	
Noémi Boubel, Thomas François and Hubert Naets	98
<i>Grammar-Based Lexicon Extension for Aligning German Radiology Text and Images</i>	
Claudia Bretschneider, Sonja Zillner and Matthias Hammon	105
<i>Recognising and Interpreting Named Temporal Expressions</i>	
Matteo Brucato, Leon Derczynski, Hector Llorens, Kalina Bontcheva and Christian S. Jensen .	113
<i>Unsupervised Improving of Sentiment Analysis Using Global Target Context</i>	
Tomáš Brychcín and Ivan Habernal	122

<i>An Agglomerative Hierarchical Clustering Algorithm for Labelling Morphs</i> Burcu Can and Suresh Manandhar	129
<i>Temporal Text Classification for Romanian Novels set in the Past</i> Alina Maria Ciobanu, Liviu P. Dinu, Octavia-Maria Şulea, anca dinu and Vlad Niculae	136
<i>A Dictionary-Based Approach for Evaluating Orthographic Methods in Cognates Identification</i> Alina Maria Ciobanu and Liviu Petrisor Dinu	141
<i>A Pilot Study on the Semantic Classification of Two German Prepositions: Combining Monolingual and Multilingual Evidence</i> Simon Clematide and Manfred Klenner	148
<i>Semantic Relations between Events and their Time, Locations and Participants for Event Coreference Resolution</i> Agata Cybulska and Piek Vossen	156
<i>Sense Clustering Using Wikipedia</i> Bharath Dandala, Chris Hokamp, Rada Mihalcea and Razvan Bunescu	164
<i>Effective Spell Checking Methods Using Clustering Algorithms</i> Renato Cordeiro de Amorim and Marcos Zampieri	172
<i>Normalization of Dutch User-Generated Content</i> Orphee De Clercq, Sarah Schulz, Bart Desmet, Els Lefever and Veronique Hoste	179
<i>Linguistic Profiling of Texts Across Textual Genres and Readability Levels. An Exploratory Study on Italian Fictional Prose</i> Felice Dell’Orletta, Simonetta Montemagni and Giulia Venturi	189
<i>Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data</i> Leon Derczynski, Alan Ritter, Sam Clark and Kalina Bontcheva	198
<i>Weighted maximum likelihood loss as a convenient shortcut to optimizing the F-measure of maximum entropy classifiers</i> Georgi Dimitroff, Laura Tolosi, Borislav Popov and Georgi Georgiev	207
<i>Sequence Tagging for Verb Conjugation in Romanian</i> Liviu Dinu, Octavia-Maria Şulea and Vlad Niculae	215
<i>A Tagging Approach to Identify Complex Constituents for Text Simplification</i> Iustin Dornescu, Richard Evans and Constantin Orasan	221
<i>Automatic Evaluation Metric for Machine Translation that is Independent of Sentence Length</i> Hiroshi Echizen’ya, Kenji Araki and Eduard Hovy	230
<i>Acronym recognition and processing in 22 languages</i> Maud Ehrmann, Leo della Rocca, Ralf Steinberger and Hristo Tanev	237
<i>An Evaluation Summary Method Based on a Combination of Content and Linguistic Metrics</i> Samira Ellouze, Maher Jaoua and Lamia Hadrich Belguith	245
<i>Hierarchy Identification for Automatically Generating Table-of-Contents</i> Nicolai Erbs, Iryna Gurevych and Torsten Zesch	252

<i>Temporal Relation Classification in Persian and English contexts</i> Mahbaneh Eshaghzadeh Torbati, Gholamreza Ghassem-sani, Seyed Abolghasem Mirroshandel, Yadollah Yaghoobzadeh and Negin Karimi Hosseini	261
<i>The Extended Lexicon: Language Processing as Lexical Description</i> Roger Evans	270
<i>Did I really mean that? Applying automatic summarisation techniques to formative feedback</i> Debra Field, Stephen Pulman, Nicolas Van Labeke, Denise Whitelock and John Richardson .	277
<i>Matching sets of parse trees for answering multi-sentence questions</i> Boris Galitsky, Dmitry Ilvovsky, Sergei O. Kuznetsov and Fedor Strok	285
<i>Realization of common statistical methods in computational linguistics with functional automata</i> Stefan Gerdjikov, Petar Mitankin and Vladislav Nenchev	294
<i>Mining Fine-grained Opinion Expressions with Shallow Parsing</i> Sucheta Ghosh, Sara Tonelli and Richard Johansson	302
<i>Justifying Corpus-Based Choices in Referring Expression Generation</i> Helmut Horacek	311
<i>A Boosting-based Algorithm for Classification of Semi-Structured Text using the Frequency of Substructures</i> Tomoya Iwakura	319
<i>Headerless, Quoteless, but not Hopeless? Using Pairwise Email Classification to Disentangle Email Threads</i> Emily Jamison and Iryna Gurevych	327
<i>Using Parallel Corpora for Word Sense Disambiguation</i> Dimitar Kazakov and Ahmad R. Shahid	336
<i>Recognizing semantic relations within Polish noun phrase: A rule-based approach</i> Paweł Kędzia and Marek Maziarz	342
<i>Unsupervised Induction of Arabic Root and Pattern Lexicons using Machine Learning</i> Bilal Khaliq and John Carroll	350
<i>Towards Domain Adaptation for Parsing Web Data</i> Mohammad Khan, Markus Dickinson and Sandra Kübler	357
<i>Capturing Anomalies in the Choice of Content Words in Compositional Distributional Semantic Space</i> Ekaterina Kochmar and Ted Briscoe	365
<i>Incremental and Predictive Dependency Parsing under Real-Time Conditions</i> Arne Köhn and Wolfgang Menzel	373
<i>Rationale, Concepts, and Current Outcome of the Unit Graphs Framework</i> Maxime Lefrançois and Fabien Gandon	382
<i>The Unit Graphs Framework: Foundational Concepts and Semantic Consequence</i> Maxime Lefrançois and Fabien Gandon	389
<i>Confidence Estimation for Knowledge Base Population</i> Xiang Li and Ralph Grishman	396

<i>Towards Fine-grained Citation Function Classification</i>	
Xiang Li, Yifan He, Adam Meyers and Ralph Grishman	402
<i>Supervised Morphology Generation Using Parallel Corpus</i>	
Alireza Mahmoudi, Mohsen Arabsorkhi and Hesham Faili	408
<i>Sentiment Analysis of Reviews: Should we analyze writer intentions or reader perceptions?</i>	
Isa Maks and Piek Vossen	415
<i>Revisiting the Old Kitchen Sink: Do we Need Sentiment Domain Adaptation?</i>	
Riham Mansour, Nesma Refaei, Michael Gamon, Ahmed Abdul-Hamid and Khaled Sami	420
<i>Evaluation of baseline information retrieval for Polish open-domain Question Answering system</i>	
Michał Marcińczuk, Adam Radziszewski, Maciej Piasecki, Dominik Piasecki and Marcin Ptak	428
<i>WCCL Relation — a Toolset for Rule-based Recognition of Semantic Relations Between Named Entities</i>	
Michał Marcińczuk	436
<i>Beyond the Transfer-and-Merge Wordnet Construction: plWordNet and a Comparison with WordNet</i>	
Marek Maziarz, Maciej Piasecki, Ewa Rudnicka and Stan Szpakowicz	443
<i>History Based Unsupervised Data Oriented Parsing</i>	
Mohsen Mesgar and Gholamreza Ghasem-Sani	453
<i>Contrasting and Corroborating Citations in Journal Articles</i>	
Adam Meyers	460
<i>CCG Categories for Distributional Semantic Models</i>	
Paramita Mirza and Raffaella Bernardi	467
<i>Discourse-aware Statistical Machine Translation as a Context-sensitive Spell Checker</i>	
Behzad Mirzababaei, Hesham Faili and Nava Ehsan	475
<i>Cross-Lingual Information Retrieval and Semantic Interoperability for Cultural Heritage Repositories</i>	
Johanna Monti, Mario Monteleone, Maria Pia di Buono and Federica Marano	483
<i>Improving Web 2.0 Opinion Mining Systems Using Text Normalisation Techniques</i>	
Alejandro Mosquera and Paloma Moreda Pozo	491
<i>Identifying Social and Expressive Factors in Request Texts Using Transaction/Sequence Model</i>	
Dasa Munkova, Michal Munk and Zuzana Fraterova	496
<i>Parameter Optimization for Statistical Machine Translation: It Pays to Learn from Hard Examples</i>	
Preslav Nakov, Fahad Al Obaidli, Francisco Guzman and Stephan Vogel	504
<i>Automatic Cloze-Questions Generation</i>	
Annamaneni Narendra, Manish Agarwal and Rakshit shah	511
<i>High-Accuracy Phrase Translation Acquisition Through Battle-Royale Selection</i>	
Lionel Nicolas, Egon W. Stemle, Klara Kranebitter and Verena Lyding	516
<i>Enriching Patent Search with External Keywords: a Feasibility Study</i>	
Ivelina Nikolova, Irina Temnikova and Galia Angelova	525

<i>A clustering approach for translationese identification</i> Sergiu Nisioi and Liviu P. Dinu	532
<i>PurePos 2.0: a hybrid tool for morphological disambiguation</i> György Orosz and Attila Novak	539
<i>More than Bag-of-Words: Sentence-based Document Representation for Sentiment Analysis</i> Georgios Paltoglou and Mike Thelwall	546
<i>Information Spreading in Expanding Wordnet Hypernymy Structure</i> Maciej Piasecki, Radosław Ramocki and Michał Kaliński	553
<i>Context Independent Term Mapper for European Languages</i> Mārcis Pinnis	562
<i>Semi-supervised vs. Cross-domain Graphs for Sentiment Analysis</i> Natalia Ponomareva and Mike Thelwall	571
<i>Towards a Hybrid Rule-based and Statistical Arabic-French Machine Translation System</i> fatiha sadat	579
<i>Segmenting vs. Chunking Rules: Unsupervised ITG Induction via Minimum Conditional Description Length</i> Markus Saers, Karteek Addanki and Dekai Wu	584
<i>A Combined Pattern-based and Distributional Approach for Automatic Hypernym Detection in Dutch.</i> Gwendolijn Schropp, Els Lefever and Veronique Hoste	593
<i>Exploiting Synergies Between Open Resources for German Dependency Parsing, POS-tagging, and Morphological Analysis</i> Rico Sennrich, Martin Volk and Gerold Schneider	601
<i>Using a Weighted Semantic Network for Lexical Semantic Relatedness</i> Reda Sibliini and Leila Kosseim	610
<i>A New Approach to the POS Tagging Problem Using Evolutionary Computation</i> Ana Paula Silva, Arlindo Silva and Irene Rodrigues	619
<i>How Joe and Jane Tweet about Their Health: Mining for Personal Health Information on Twitter</i> Marina Sokolova, Stan Matwin, Yasser Jafer and David Schramm	626
<i>What Sentiments Can Be Found in Medical Forums?</i> Marina Sokolova and Victoria Bobicev	633
<i>Automated learning of everyday patients' language for medical blogs analytics</i> Giovanni Stilo, Moreno De Vincenzi, Alberto E. Tozzi and Paola Velardi	640
<i>How Symbolic Learning Can Help Statistical Learning (and vice versa)</i> Isabelle Tellier and Yoann Dupont	649
<i>Measuring Closure Properties of Patent Sublanguages</i> Irina Temnikova, Negacy Hailu, Galia Angelova and K. Bretonnel Cohen	659
<i>Closure Properties of Bulgarian Clinical Text</i> Irina Temnikova, Ivelina Nikolova, William A. Baumgartner, Galia Angelova and K. Bretonnel Cohen	667

<i>Analyzing the Use of Character-Level Translation with Sparse and Noisy Datasets</i> Jörg Tiedemann and Preslav Nakov	676
<i>A Feature Induction Algorithm with Application to Named Entity Disambiguation</i> Laura Tolosi, Valentin Zhikov, Georgi Georgiev and Borislav Popov	685
<i>Introducing a Corpus of Human-Authored Dialogue Summaries in Portuguese</i> Norton Trevisan Roman, Paul Piwek, Ariadne M. B. Rizzoni Carvalho and Alexandre Rossi Alvares	692
<i>Wikipedia as an SMT Training Corpus</i> Dan Tufiş, Radu Ion, Stefan Dumitrescu and Dan Stefanescu	702
<i>DutchSemCor: in quest of the ideal sense-tagged corpus</i> Piek Vossen, Ruben Izquierdo and Attila Görög	710
<i>Towards detecting anomalies in the content of standardized LMF dictionaries</i> Wafa WALI, Bilel Gargouri and Abdelmajid BEN HAMADOU	719
<i>Edit Distance: A New Data Selection Criterion for Domain Adaptation in SMT</i> Longyue Wang, Derek F. Wong, Lidia S. Chao, Junwen Xing, Yi Lu and Isabel Trancoso	727
<i>Automatic Enhancement of LTAG Treebank</i> Farzaneh Zarei, Ali Basirat, Hesham Faili and Maryam S.Mirian	733
<i>Inductive and deductive inferences in a Crowdsourced Lexical-Semantic Network</i> Manel Zarrouk, Mathieu Lafourcade and Alain Joubert	740
<i>Machine Learning for Mention Head Detection in Multilingual Coreference Resolution</i> Desislava Zhekova and Sandra Kübler	747
<i>Combining POS Tagging, Dependency Parsing and Coreferential Resolution for Bulgarian</i> Valentin Zhikov, Georgi Georgiev, Kiril Simov and Petya Osenova	755
<i>magyarlanc: A Tool for Morphological and Dependency Parsing of Hungarian</i> János Zsibrita, Veronika Vincze and Richárd Farkas	763

Conference Programme

Monday September 9, 2013

Session 1a: (10:10 - 11:10) Part-of-Speech Tagging

Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data

Leon Derczynski, Alan Ritter, Sam Clark and Kalina Bontcheva

Exploiting Synergies Between Open Resources for German Dependency Parsing, POS-tagging, and Morphological Analysis

Rico Sennrich, Martin Volk and Gerold Schneider

Session 1b: (10:10 - 11:10) Language Resources

DutchSemCor: in Quest of the Ideal Sense-tagged Corpus

Piek Vossen, Rubén Izquierdo and Attila Görög

Beyond the Transfer-and-Merge Wordnet Construction: plWordNet and a Comparison with WordNet

Marek Maziarz, Maciej Piasecki, Ewa Rudnicka and Stan Szpakowicz

Session 1c: (10:10 - 11:10) Semantics

Using a Weighted Semantic Network for Lexical Semantic Relatedness

Reda Sibli and Leila Kosseim

Capturing Anomalies in the Choice of Content Words in Compositional Distributional Semantic Space

Ekaterina Kochmar and Ted Briscoe

Monday September 9, 2013 (continued)

Session 2a: (11:40 - 13:00) Opinion Mining and Sentiment Analysis

Semi-supervised vs. Cross-domain Graphs for Sentiment Analysis

Natalia Ponomareva and Mike Thelwall

Sentiment Analysis of Reviews: Should we Analyze Writer Intentions or Reader Perceptions?

Isa Maks and Piek Vossen

Improving Web 2.0 Opinion Mining Systems Using Text Normalisation Techniques

Alejandro Mosquera and Paloma Moreda Pozo

Mining Fine-grained Opinion Expressions with Shallow Parsing

Sucheta Ghosh, Sara Tonelli and Richard Johansson

Session 2b: (11:40 - 13:00) Multilingual NLP

Acronym Recognition and Processing in 22 Languages

Maud Ehrmann, Leonida della Rocca, Ralf Steinberger and Hristo Tannev

Sense Clustering Using Wikipedia

Bharath Dandala, Chris Hokamp, Rada Mihalcea and Razvan Bunescu

A Pilot Study on the Semantic Classification of Two German Prepositions: Combining Monolingual and Multilingual Evidence

Simon Clematide and Manfred Klenner

Temporal Relation Classification in Persian and English contexts

Mahbaneh Eshaghzadeh Torbati, Gholamreza Ghassem-sani, Seyed Abolghasem Mirroshandel, Yadollah Yaghoobzadeh and Negin Karimi Hosseini

Session 2c: (11:40 - 13:00) Language Modelling

Segmenting vs. Chunking Rules: Unsupervised ITG Induction via Minimum Conditional Description Length

Markus Saers, Kartteek Addanki and Dekai Wu

CCG Categories for Distributional Semantic Models

Paramita Mirza and Raffaella Bernardi

Identifying Social and Expressive Factors in Request Texts Using Transaction/Sequence Model

Daša Munková, Michal Munk and Zuzana Fráterová

Realization of Common Statistical Methods in Computational Linguistics with Functional Automata

Stefan Gerdjikov, Petar Mitankin and Vladislav Nenchev

Monday September 9, 2013 (continued)

Session 3a: (15:30 - 16:30) Information Extraction

Confidence Estimation for Knowledge Base Population

Xiang Li and Ralph Grishman

TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text

Kalina Bontcheva, Leon Derczynski, Adam Funk, Mark Greenwood, Diana Maynard and Niraj Aswani

Session 3b: (15:30 - 16:30) Term Extraction

Context Independent Term Mapper for European Languages

Mārcis Pinnis

Automated Learning of Everyday Patients' Language for Medical blogs Analytics

Giovanni Stilo, Moreno De Vincenzi, Alberto E. Tozzi and Paola Velardi

Session 3c: (15:30 - 16:30) Parsing

magyarlanc: A Tool for Morphological and Dependency Parsing of Hungarian

János Zsibrita, Veronika Vincze and Richárd Farkas

Incremental and Predictive Dependency Parsing under Real-Time Conditions

Arne Köhn and Wolfgang Menzel

Monday September 9, 2013 (continued)

Session 4: (16:30 - 18:30) Methods, Resources and Language Processing Tasks (Posters)

Towards Domain Adaptation for Parsing Web Data

Mohammad Khan, Markus Dickinson and Sandra Kübler

Annotating events, Time and Place Expressions in Arabic Texts

Hassina Aliane, Wassila Guendouzi and Amina Mokrani

Semantic Relations between Events and their Time, Locations and Participants for Event Coreference Resolution

Agata Cybulska and Piek Vossen

Machine Learning for Mention Head Detection in Multilingual Coreference Resolution

Desislava Zhekova and Sandra Kübler

Contrasting and Corroborating Citations in Journal Articles

Adam Meyers

A New Approach to the POS Tagging Problem Using Evolutionary Computation

Ana Paula Silva, Arlindo Silva and Irene Rodrigues

The Extended Lexicon: Language Processing as Lexical Description

Roger Evans

Supervised Morphology Generation Using Parallel Corpus

Alireza Mahmoudi, Mohsen Arabsorkhi and Hesham Faili

Enriching Patent Search with External Keywords: a Feasibility Study

Ivelina Nikolova, Irina Temnikova and Galia Angelova

Rationale, Concepts, and Current Outcome of the Unit Graphs Framework

Maxime Lefrançois and Fabien Gandon

The Unit Graphs Framework: Foundational Concepts and Semantic Consequence

Maxime Lefrançois and Fabien Gandon

Automatic Enhancement of LTAG Treebank

Farzaneh Zarei, Ali Basirat, Hesham Faili and Maryam S.Mirian

An Agglomerative Hierarchical Clustering Algorithm for Labelling Morphs

Burcu Can and Suresh Manandhar

Automatic Evaluation Metric for Machine Translation that is Independent of Sentence Length

Hiroshi Echizen'ya, Kenji Araki and Eduard Hovy

Monday September 9, 2013 (continued)

Inductive and deductive inferences in a Crowdsourced Lexical-Semantic Network

Manel Zarrouk, Mathieu Lafourcade and Alain Joubert

Semantic Relation Recognition within Polish Noun Phrase: A Rule-based Approach

Paweł Kedzia and Marek Maziarz

Using Parallel Corpora for Word Sense Disambiguation

Dimitar Kazakov and Ahmad R. Shahid

Unsupervised Induction of Arabic Root and Pattern Lexicons using Machine Learning

Bilal Khaliq and John Carroll

A Unified Lexical Processing Framework based on the Margin Infused Relaxed Algorithm.

A Case Study on the Romanian Language

Tiberiu Boros

An Evaluation Summary Method Based on a Combination of Content and Linguistic Metrics

Samira Ellouze, Maher Jaoua and Lamia Hadrach Belguith

A Dictionary-Based Approach for Evaluating Orthographic Methods in Cognates Identification

Alina Maria Ciobanu and Liviu Petrisor Dinu

Tuesday September 10, 2013

Session 5a: (10:00 - 11:00) Text Summarisation and Tables-of-Content Generation

Introducing a Corpus of Human-Authored Dialogue Summaries in Portuguese

Norton Trevisan Roman, Paul Piwek, Ariadne M. B. Rizzoni Carvalho and Alexandre Rossi Alvares

Hierarchy Identification for Automatically Generating Table-of-Contents

Nicolai Erbs, Iryna Gurevych and Torsten Zesch

Session 5b: (10:00 - 11:00) Translation Technology / MT-inspired Approaches

High-Accuracy Phrase Translation Acquisition Through Battle-Royale Selection

Lionel Nicolas, Egon W. Stemle, Klara Kranebitter and Verena Lyding

Normalization of Dutch User-Generated Content

Orphée De Clercq, Sarah Schulz, Bart Desmet, Els Lefever and Véronique Hoste

Session 5c: (10:00 - 11:00) POS Tagging and Parsing

Combining POS Tagging, Dependency Parsing and Coreferential Resolution for Bulgarian

Valentin Zhikov, Georgi Georgiev, Kiril Simov and Petya Osenova

How Symbolic Learning Can Help Statistical Learning (and vice versa)

Isabelle Tellier and Yoann Dupont

Tuesday September 10, 2013 (continued)

Session 6a: (11:30 - 12:50) Machine Translation and Question Answering

Wikipedia as an SMT Training Corpus

Dan Tufiş, Radu Ion, Stefan Dumitrescu and Dan Stefanescu

Edit Distance: A New Data Selection Criterion for Domain Adaptation in SMT

Longyue Wang, Derek F. Wong, Lidia S. Chao, Junwen Xing, Yi Lu and Isabel Trancoso

Analyzing the Use of Character-Level Translation with Sparse and Noisy Datasets

Jörg Tiedemann and Preslav Nakov

Matching Sets of Parse Trees for Answering Multi-Sentence Questions

Boris Galitsky, Dmitry Ilvovsky, Sergei O. Kuznetsov and Fedor Strok

Session 6b: (11:30 - 12:50) Language Resources and Opinion Mining

Information Spreading in Expanding Wordnet Hypernymy Structure

Maciej Piasecki, Radosław Ramocki and Michał Kaliński

Towards Detecting Anomalies in the Content of Standardized LMF Dictionaries

Wafa Wali, Bilel Gargouri and Abdelmajid Ben Hamadou

A Combined Pattern-based and Distributional Approach for Automatic Hypernym Detection in Dutch.

Gwendolijn Schropp, Els Lefever and Véronique Hoste

Revisiting the Old Kitchen Sink: Do we Need Sentiment Domain Adaptation?

Riham Mansour, Nesma Refaei, Michael Gamon, Ahmed Abdul-Hamid and Khaled Sami

Wednesday September 11, 2013

Session 7a: (10:00 - 11:00) NLP Applications (Text Simplification, Ssentiment Analysis)

A Tagging Approach to Identify Complex Constituents for Text Simplification

Justin Dornescu, Richard Evans and Constantin Orasan

Unsupervised Improving of Sentiment Analysis Using Global Target Context

Tomáš Brychcín and Ivan Habernal

Session 7b: (10:00 - 11:10) NLP Tasks (Temporal Processing, NER, Text Generation)

Recognising and Interpreting Named Temporal Expressions

Matteo Brucato, Leon Derczynski, Hector Llorens, Kalina Bontcheva and Christian S. Jensen

A Semi-supervised Learning Approach to Arabic Named Entity Recognition

Maha Althobaiti, Udo Kruschwitz and Massimo Poesio

Justifying Corpus-Based Choices in Referring Expression Generation

Helmut Horacek

Session 7c: (10:00 - 11:10) NLP Methods (Optimisations, Representations)

Weighted Maximum Likelihood Loss as a Convenient Shortcut to Optimizing the F-measure of Maximum Entropy Classifiers

Georgi Dimitroff, Laura Tološi, Borislav Popov and Georgi Georgiev

Optimising Tree Edit Distance with Subtrees for Textual Entailment

Maytham Alabbas and Allan Ramsay

Towards a Structured Representation of Generic Concepts and Relations in Large Text Corpora

Archana Bhattarai and Vasile Rus

Wednesday September 11, 2013 (continued)

Session 8a: (11:40 - 13:00) Text categorisation

Headerless, Quoteless, but not Hopeless? Using Pairwise Email Classification to Disentangle Email Threads

Emily Jamison and Iryna Gurevych

A Boosting-based Algorithm for Classification of Semi-Structured Text using the Frequency of Substructures

Tomoya Iwakura

Authorship Attribution in Health Forums

Victoria Bobicev, Marina Sokolova, Khaled El Emam and Stan Matwin

Linguistic Profiling of Texts Across Textual Genres and Readability Levels. An Exploratory Study on Italian Fictional Prose

Felice Dell'Orletta, Simonetta Montemagni and Giulia Venturi

Session 8b: (11:40 - 13:00) NLP for Specialised Languages

Closure Properties of Bulgarian Clinical Text

Irina Temnikova, Ivelina Nikolova, William A. Baumgartner, Galia Angelova and K. Bretonnel Cohen

Measuring Closure Properties of Patent Sublanguages

Irina Temnikova, Negacy Hailu, Galia Angelova and K. Bretonnel Cohen

Grammar-Based Lexicon Extension for Aligning German Radiology Text and Images

Claudia Bretschneider, Sonja Zillner and Matthias Hammon

Domain Adaptation for Parsing

Eric Baucom, Levi King and Sandra Kübler

Session 8c: (11:40 - 13:00) Morphology and Syntax

ASMA: A System for Automatic Segmentation and Morpho-Syntactic Disambiguation of Modern Standard Arabic

Muhammad Abdul-Mageed, Mona Diab and Sandra Kübler

PurePos 2.0: a Hybrid Tool for Morphological Disambiguation

György Orosz and Attila Novák

Sequence Tagging for Verb Conjugation in Romanian

Liviu Dinu, Octavia-Maria Şulea and Vlad Niculae

History Based Unsupervised Data Oriented Parsing

Mohsen Mesgar and Gholamreza Ghasem-Sani

Wednesday September 11, 2013 (continued)

Session 9: (15:30 - 17:30) Applications (Posters)

More than Bag-of-Words: Sentence-based Document Representation for Sentiment Analysis

Georgios Paltoglou and Mike Thelwall

How Joe and Jane Tweet about Their Health: Mining for Personal Health Information on Twitter

Marina Sokolova, Stan Matwin, Yasser Jafer and David Schramm

What Sentiments Can Be Found in Medical Forums?

Marina Sokolova and Victoria Bobicev

Opinion Learning from Medical Forums

Tanveer Ali, Marina Sokolova, David Schramm and Diana Inkpen

Towards Fine-grained Citation Function Classification

Xiang Li, Yifan He, Adam Meyers and Ralph Grishman

Automatic Extraction of Contextual Valence Shifters.

Noémi Boubel, Thomas François and Hubert Naets

Improving Sentiment Analysis in Twitter Using Multilingual Machine Translated Data

Alexandra Balahur and Marco Turchi

Evaluation of Baseline Information Retrieval for Polish Open-domain Question Answering System

Michał Marcińczuk, Adam Radziszewski, Maciej Piasecki, Dominik Piasecki and Marcin Ptak

Automatic Cloze-Questions Generation

Annamaneni Narendra, Manish Agarwal and Rakshit shah

An NLP-based Reading Tool for Aiding Non-native English Readers

Mahmoud Azab, Ahmed Salama, Kemal Oflazer, Hideki Shima, Jun Araki and Teruko Mitamura

Did I Really Mean That? Applying Automatic Summarisation Techniques to Formative Feedback

Debora Field, Stephen Pulman, Nicolas Van Labeke, Denise Whitelock and John Richardson

A Feature Induction Algorithm with Application to Named Entity Disambiguation

Laura Toloşi, Valentin Zhikov, Georgi Georgiev and Borislav Popov

A Clustering Approach for Translationese Identification

Sergiu Nisioi and Liviu P. Dinu

Wednesday September 11, 2013 (continued)

Parameter Optimization for Statistical Machine Translation: It Pays to Learn from Hard Examples

Preslav Nakov, Fahad Al Obaidli, Francisco Guzman and Stephan Vogel

Cross-Lingual Information Retrieval and Semantic Interoperability for Cultural Heritage Repositories

Johanna Monti, Mario Monteleone, Maria Pia di Buono and Federica Marano

Discourse-aware Statistical Machine Translation as a Context-sensitive Spell Checker

Behzad Mirzababaei, Hesham Faili and Nava Ehsan

Towards a Hybrid Rule-based and Statistical Arabic-French Machine Translation System

Fatiha Sadat

WCCL Relation — a Toolset for Rule-based Recognition of Semantic Relations Between Named Entities

Michał Marcińczuk

Temporal Text Classification for Romanian Novels set in the Past

Alina Maria Ciobanu, Liviu P. Dinu, Octavia-Maria Șulea, Anca Dinu and Vlad Niculae

Effective Spell Checking Methods Using Clustering Algorithms

Renato Cordeiro de Amorim and Marcos Zampieri

