International Conference on Recent Advances in Natural Language Processing 2011

(RANLP 2011)

Hissar, Bulgaria
12-14 September 2011

Editors:

Galia Angelova
Kalina Bontcheva
Ruslan Mitkov
Nikolai Nikolov

ISBN: 978-1-62276-511-9

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2011) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Extracting STRIPS Representations of Actions and Events
Avirup Sil and Alexander Yates 1

Acquiring Topic Features to improve Event Extraction: in Pre-selected and Balanced Collections
Shasha Liao and Ralph Grishman 9

Minimally Supervised Rule Learning for the Extraction of Biographic Information from Various Social Domains
Hong Li, Feiyu Xu and Hans Uszkoreit 17

Extracting Relations Within and Across Sentences
Kumutha Swampillai and Mark Stevenson 25

Knowledge-Poor Approach to Shallow Parsing: Contribution of Unsupervised Part-of-Speech Induction
Marie Guégan and Claude de Loupy 33

Fast Domain Adaptation for Part of Speech Tagging for Dialogues
Sandra Kübler and Eric Baucom 41

Using a Morphological Database to Increase the Accuracy in POS Tagging
Hrafn Loftsson, Sigrún Helgadóttir and Eiríkur Rögnvaldsson 49

Actions Speak Louder than Words: Evaluating Parsers in the Context of Natural Language Understanding Systems for Human-Robot Interaction
Sandra Kübler, Rachael Cantrell and Matthias Scheutz 56

Constructing Linguistically Motivated Structures from Statistical Grammars
Ali Basirat and Heshaam Faili 63

An Open Source Punjabi Resource Grammar
Shafqat Mumtaz Virk, Muhammad Humayoun and Aarne Ranta 70

Multi-Document Summarization by Capturing the Information Users are Interested in
Elena Lloret, Laura Plaza and Ahmet Aker 77

Efficient algorithm for Context Sensitive Aggregation in Natural Language generation
Hemanth Sagar Bayyarapu 84

Enriching a statistical machine translation system trained on small parallel corpora with rule-based bilingual phrases
Víctor M. Sánchez-Cartagena, Felipe Sánchez-Martínez and Juan Antonio Pérez-Ortiz 90

Assessing the Post-Editing Effort for Automatic and Semi-Automatic Translations of DVD Subtitles
Sheila C. M. de Sousa, Wilker Aziz and Lucia Specia 97

JRC-NAMES: A Freely Available, Highly Multilingual Named Entity Resource
Ralf Steinberger, Bruno Pouliquen, Mijail Kabadjov, Jenya Belyaeva and Erik van der Goot . . . 104

MDL-based Models for Alignment of Etymological Data
Hannes Wettig, Suvi Hiltunen and Roman Yangarber 111

Building a Multilingual Named Entity-Annotated Corpus Using Annotation Projection
Maud Ehrmann, Marco Turchi and Ralf Steinberger 118

Bilingual lexicon extraction from comparable corpora for closely related languages
Darja Fišer and Nikola Ljubešić 125

Sentiments and Opinions in Health-related Web messages
Marina Sokolova and Victoria Bobicev 132

An Exploration into the Use of Contextual Document Clustering for Cluster Sentiment Analysis
Niall Rooney, Hui Wang, Fiona Browne, Fergal Monaghan, Jann Müller, Alan Sergeant, Zhiwei Lin, Philip Taylor and Vladimir Dobrynin 140

Pause and Stop Labeling for Chinese Sentence Boundary Detection
Hen-Hsen Huang and Hsin-Hsi Chen 146

Multilabel Tagging of Discourse Relations in Ambiguous Temporal Connectives
Yannick Versley 154

Noun Compound and Named Entity Recognition and their Usability in Keyphrase Extraction
István Nagy T., Gábor Berend and Veronika Vincze 162

A Named Entity Recognition Method using Rules Acquired from Unlabeled Data
Tomoya Iwakura 170

An Incremental Entity-Mention Model for Coreference Resolution with Restrictive Antecedent Accessibility
Manfred Klenner and Don Tuggener 178

Cross-Domain Dutch Coreference Resolution
Orphée De Clercq, Véronique Hoste and Iris Hendrickx 186

Finding the Best Approach for Multi-lingual Text Summarisation: A Comparative Analysis
Elena Lloret and Manuel Palomar 194

Automatically Creating General-Purpose Opinion Summaries from Text
Veselin Stoyanov and Claire Cardie 202

Exploring the Usefulness of Cross-lingual Information Fusion for Refining Real-time News Event Extraction: A Preliminary Study
Jakub Piskorski, Jenya Belayeva and Martin Atkinson 210

Temporal Relation Extraction Using Expectation Maximization
Seyed Abolghasem Mirroshandel and Gholamreza Ghassem-Sani 218

Improving Chunk-based Semantic Role Labeling with Lexical Features
Wilker Aziz, Miguel Rios and Lucia Specia 226

Improving WSD using ISR-WN with Relevant Semantic Trees and SemCor Senses Frequency
Yoan Gutiérrez, Sonia Vázquez and Andrés Montoyo 233

Investigating Advanced Techniques for Document Content Similarity Applied to External Plagiarism Analysis
Daniel Micol, Rafael Muñoz and Óscar Ferrández 240

Using Cognates in a French-Romanian Lexical Alignment System: A Comparative Study
Mirabela Navlea and Amalia Todiraşcu 247

Highly Multilingual Coreference Resolution Exploiting a Mature Entity Repository
Josef Steinberger, Jenya Belyaeva, Jonathan Crawley, Leonida Della-Rocca, Mohamed Ebrahim, Maud Ehrmann, Mijail Kabadjov, Ralf Steinberger and Erik Van-der-Goot 254

Singletons and Coreference Resolution Evaluation
Sandra Kübler and Desislava Zhekova 261

Modelling Entity Instantiations
Andrew McKinlay and Katja Markert 268

A New Scheme for Annotating Semantic Relations between Named Entities in Corpora
Mani Ezzat and Thierry Poibeau 275

Prototypical Opinion Holders: What We can Learn from Experts and Analysts
Michael Wiegand and Dietrich Klakow 282

Multiword Expressions and Named Entities in the Wiki50 Corpus
Veronika Vincze, István Nagy T. and Gábor Berend 289

Towards the Automatic Merging of Lexical Resources: Automatic Mapping
Muntsa Padró, Núria Bel and Silvia Necsulescu 296

Unsupervised Learning for Persian WordNet Construction
Mortaza Montazery and Heshaam Faili 302

Domain Independent Authorship Attribution without Domain Adaptation
Rohith Menon and Yejin Choi 309

Cultural Configuration of Wikipedia: measuring Autoreferentiality in Different Languages
Marc Miquel Ribé and Horacio Rodríguez 316

Combining Relational and Attributional Similarity for Semantic Relation Classification
Preslav Nakov and Zornitsa Kozareva 323

In Search of Missing Arguments: A Linguistic Approach
Josef Ruppenhofer, Philip Gorinski and Caroline Sporleder 331

Enlarging Monolingual Dictionaries for Machine Translation with Active Learning and Non-Expert Users
Miquel Esplà-Gomis, Víctor M. Sánchez-Cartagena and Juan Antonio Pérez-Ortiz 339

Morphological Analysis of Biomedical Terminology with Analogy-Based Alignment
Vincent Claveau and Ewa Kijak 347

Adaptability of Lexical Acquisition for Large-scale Grammars
Kostadin Cholakov, Gertjan van Noord, Valia Kordoni and Yi Zhang 355

Integration of Data from a Syntactic Lexicon into Generative and Discriminative Probabilistic Parsers
Anthony Sigogne, Matthieu Constant and Éric Laporte 363

Pattern Learning for Event Extraction using Monolingual Statistical Machine Translation
Marco Turchi, Vanni Zavarella and Hristo Tanev 371

META-DARE: Monitoring the Minimally Supervised ML of Relation Extraction Rules
Hong Li, Feiyu Xu and Hans Uszkoreit 378

Mining Transliterations from Wikipedia using Dynamic Bayesian Networks
Peter Nabende 385

Detecting Opinions Using Deep Syntactic Analysis
Caroline Brun 392

Using Visual Information to Predict Lexical Preference
Shane Bergsma and Randy Goebel 399

Systematic Knowledge Acquisition for Question Analysis
Dat Quoc Nguyen, Dai Quoc Nguyen and Son Bao Pham 406

A Semi-Automatic, Iterative Method for Creating a Domain-Specific Treebank
Corina Dima and Erhard Hinrichs 413

Determining Immediate Constituents of Compounds in GermaNet
Verena Henrich and Erhard Hinrichs 420

Segmentation and Clustering of Textual Sequences: a Typological Approach
Christelle Cocco, Raphaël Pittier, François Bavaud and Aris Xanthos 427

A Contextual Classification Strategy for Polarity Analysis of Direct Quotations from Financial News
Brett Drury, Gaël Dias and Luís Torgo 434

On the Contribution of MWE-based Features to a Readability Formula for French as a Foreign Language
Thomas François and Patrick Watrin 441

Exploiting Hidden Morphophonemic Constraints for Finding the Underlying Forms of 'weak' Arabic Verbs
Allan Ramsay and Hanady Mansour 448

A Confidence Model for Syntactically-Motivated Entailment Proofs
Asher Stern and Ido Dagan 455

Learning Script Participants from Unlabeled Data
Michaela Regneri, Alexander Koller, Josef Ruppenhofer and Manfred Pinkal 463

Towards Minimal Recursion Semantics over Bulgarian Dependency Parsing
Kiril Simov and Petya Osenova 471

Discourse Structures to Reduce Discourse Incoherence in Blog Summarization
Shamima Mithun and Leila Kosseim 479

Parallel Suffix Arrays for Linguistic Pattern Search
Johannes Goller 487

A Mechanism to Restrict the Scope of Clause-Bounded Quantifiers in 'Continuation' Semantics
Anca Dinu 495

A Support Tool for Deriving Domain Taxonomies from Wikipedia
Lili Kotlerman, Zemer Avital, Ido Dagan, Amnon Lotan and Ofer Weintraub 503

Barrier Features for Classification of Semantic Relations
Anita Alicante and Anna Corazza 509

A Reflective View on Text Similarity
Daniel Bär, Torsten Zesch and Iryna Gurevych 515

Evaluating the Robustness of EmotiBlog for Sentiment Analysis and Opinion Mining
Ester Boldrini, Javi Fernández, José Manuel Gómez and Patricio Martínez-Barco 521

Hybrid System For Plagiarism Detection
Javier R. Bru, Patricio Martínez-Barco and Rafael Muñoz 527

Data-Driven Approach Using Semantics for Recognizing and Classifying TimeML Events in Italian
Tommaso Caselli, Hector Llorens, Borja Navarro-Colorado and Estela Saquete 533

Can Alternations Be Learned? A Machine Learning Approach To Romanian Verb Conjugation
Liviu P. Dinu, Emil Ionescu, Vlad Niculae and Octavia-Maria Şulea 539

A New Representation Model for the Automatic Recognition and Translation of Arabic Named Entities with NooJ
Héla Fehri, Kais Haddar and Abdelmajid Ben Hamadou 545

Training Data in Statistical Machine Translation - the More, the Better?
Monica Gavrila and Cristina Vertan 551

Towards a Corpus-based Approach to Modelling Language Production of Foreign Language Learners in Communicative Contexts
Voula Gotsoulia and Bessie Dendrinos 557

Parsing a Polysynthetic Language
Petr Homola 562

An algorithm of Identifying Semantic Arguments of a Verb From Structured Data
Minhua Huang and Robert M. Haralick 568

Construction of an HPSG Grammar for the Arabic Relative Sentences
Ines Zalila and Kais Haddar 574

Automatically Selected Skip Edges in Conditional Random Fields for Named Entity Recognition
Roman Klinger 580

Negation Naive Bayes for Categorization of Product Pages on the Web
Kanako Komiya, Naoto Sato, Koji Fujimoto and Yoshiyuki Kotani 586

A Hybrid Approach for Event Extraction and Event Actor Identification
Anup Kumar Kolya, Asif Ekbal and Sivaji Bandyopadhyay 592

Evaluating Human Correction Quality for Machine Translation from Crowdsourcing
Shasha Liao, Cheng Wu and Juan Huerta 598

Multi-class SVM for Relation Extraction from Clinical Reports
Anne-Lyse Minard, Anne-Laure Ligozat and Brigitte Grau 604

Discovering coreference using image-grounded verb models
Amitabha Mukerjee, Kruti Neema and Sushobhan Nayak 610

Word and Phrase Learning based on Prior Semantics
Amitabha Mukerjee and Nikhil Joshi 616

Domain-Dependent Identification of Multiword Expressions
István Nagy T., Veronika Vincze and Gábor Berend 622

Robust Semantic Analysis for Unseen Data in FrameNet
Alexis Palmer, Afra Alishahi and Caroline Sporleder 628

Studying Translationese at the Character Level
Marius Popescu 634

Linear Transduction Grammars and Zipper Finite-State Transducers
Markus Saers and Dekai Wu 640

Finding Negative Key Phrases for Internet Advertising Campaigns using Wikipedia
Martin Scaiano and Diana Inkpen 648

Establishing Implementation Priorities in Aiding Writers of Controlled Crisis Management Texts
Irina Temnikova 654

TechWatchTool: Innovation and Trend Monitoring
Hong Li, Feiyu Xu and Hans Uszkoreit 660

"Yes we can?": Subjectivity Annotation and Tagging for the Health Domain
Muhammad Abdul-Mageed, Mohammed Korayem and Ahmed YoussefAgha 666

Wordnets: State of the Art and Perspectives. Case Study: the Romanian Wordnet
Verginica Barbu Mititelu 672

Creation and Development of the Romanian Lexical Resources
Elena Boian, Constantin Ciubotaru, Svetlana Cojocaru, Alexandru Colesnicov, Ludmila Malahov and Mircea Petic 678

Analyses Tools for Non-head Structures
Sirine Boukedi and Kais Haddar 686

Visualization for Coreference Annotation
Andre Burkovski and Gunther Heidemann 692

The RST Spanish Treebank On-line Interface
Iria da Cunha, Juan-Manuel Torres-Moreno, Gerardo Sierra, Luis Adrián Cabrera-Diego, Brenda Gabriela Castro Rolón and Juan Miguel Rolland Bartilotti 698

Lexical Generalisation for Word-level Matching in Plagiarism Detection
Miranda Chong and Lucia Specia 704

Multiple Evidence for Term Extraction in Broad Domains
Boris Dobrov and Natalia Loukachevitch 710

Language Modeling for Document Selection in Question Answering
Nicolas Foucault, Gilles Adda and Sophie Rosset 716

Evaluating Various Linguistic Features on Semantic Relation Extraction
Marcos Garcia and Pablo Gamallo 721

Automatic titling of Articles Using Position and Statistical Information
Cédric Lopez, Violaine Prince and Mathieu Roche 727

Unsupervised Domain Adaptation based on Text Relatedness
Georgios Petasis 733

Bilingual Experiments with an Arabic-English Corpus for Opinion Mining
Mohammed Rushdi-Saleh, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López and José M. Perea-Ortega 740

Experiments on Term Extraction using Noun Phrase Subclassifications
Merley da Silva Conrado, Walter Koza, Josuka Díaz-Labrador, Joseba Abaitua, Solange Oliveira Rezende, Thiago Pardo and Zulema Solana 746

Adaptive Feedback Message Generation for Second Language Learners of Arabic
Khaled Shaalan and Marwa Magdy 752

Building a Patient-based Ontology for User-written Web Messages
Marina Sokolova and David Schramm 758

Recognition and Classification of Numerical Entities in Basque
Ander Soraluze, Iñaki Alegria, Olatz Ansa, Olatz Arregi and Xabier Arregi 764

Multilingual Entity-Centered Sentiment Analysis Evaluated by Parallel Corpora
Josef Steinberger, Polina Lenkova, Mijail Kabadjov, Ralf Steinberger and Erik van der Goot . . 770

Term Validation for Vocabulary Construction and Key Term Extraction
Alexander Ulanov and Andrey Simanovsky 776

Agreement: How to Reach it? Defining Language Features Leading to Agreement in Discourse
Tatiana Zidraşco, Victoria Bobicev, Shun Shiramatsu, Tadachika Ozono and Toramatsu Shintani 781