International Conference on Recent Advances in Natural Language Processing 2009 (RANLP 2009)

Borovets, Bulgaria
14 - 16 September 2009

Editors:

Galia Angelova
Kalina Bontcheva
Ruslan Mitkov
Nicolas Nicolov
Nikolai Nikolov

ISBN: 978-1-5108-1327-4

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© (2009) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2016)

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

Unsupervised Relation Extraction for Automatic Generation of Multiple-Choice Questions
Naveed Afzal and Viktor Pekar 1

Summary Generation for Toponym-referenced Images using Object Type Language Models
Ahmet Aker and Robert Gaizauskas 6

Prepositional Phrase Attachment in Shallow Parsing
Vincent Van Asch and Walter Daelemans 12

A Comparative Study of Open Domain and Opinion Question Answering Systems for Factual and Opinionated Queries
Alexandra Balahur, Ester Boldrini, Andrés Montoyo and Patricio Martínez-Barco 18

Acquisition of Common Sense Knowledge for Basic Level Concepts
Eduard Barbu 23

Unsupervised Knowledge Extraction for Taxonomies of Concepts from Wikipedia
Eduard Barbu and Massimo Poesio 28

Exploring Treebank Transformations in Dependency Parsing
Kepa Bengoetxea and Koldo Gojenola 33

Contextual Salience in Query-based Summarization
Wauter Bosma 39

Integrating Document Structure into a Multi-Document Summarizer
Aurélien Bossard and Thierry Poibeau 45

Cross-Linguistic Sentiment Analysis: From English to Spanish
Julian Brooke, Milan Tofiloski and Maite Taboada 50

The Influence of Text Pre-processing on Plagiarism Detection
Zdenek Ceska and Chris Fox 55

Combining Finite State and Corpus-based Techniques for Unknown Word Prediction
Kostadin Cholakov and Gertjan van Noord 60

Prototype-based Active Learning for Lemmatization
Walter Daelemans, Hendrik J. Groenewald and Gerhard B. van Huyssteen 65

From Partial toward Full Parsing
Heshaam Faili 71

Grouping Synonyms by Definitions
Ingrid Falk, Claire Gardent, Evelyne Jacquey and Fabienne Venant 76

Singular Value Decomposition for Feature Selection in Taxonomy Learning
Francesca Fallucchi and Fabio Massimo Zanzottto 82

Improving Text Segmentation by Combining Endogenous and Exogenous Methods
Olivier Ferret 88

Edlin: an Easy to Read Linear Learning Framework
Kuzman Ganchev and Georgi Georgiev 94

Exploiting the Use of Prior Probabilities for Passage Retrieval in Question Answering
Surya Ganesh and Vasudeva Varma 99

Exploiting Structure and Content of Wikipedia for Query Expansion in the Context
Surya Ganesh and Vasudeva Varma 103

Text Content and Task Performance in the Evaluation of a Natural Language Generation System
Albert Gatt and François Portet 107

Feature-Rich Named Entity Recognition for Bulgarian Using Conditional Random Fields
Georgi Georgiev, Preslav Nakov, Kuzman Ganchev, Petya Osenova and Kiril Simov 113

Uncertainty Detection for Information Extraction
Bénédicte Goujon 118

Learning to Identify Educational Materials
Samer Hassan and Rada Mihalcea 123

Lexicalized Semi-incremental Dependency Parsing
Hany Hassan, Khalil Sima'an and Andy Way 128

Identification of Parallel Text Pairs Using Fingerprints
Martin Hassel and Hercules Dalianis 135

Stochastic Definite Clause Grammars
Christian Theil Have 139

Topic-based Multi-Document Summarization with Probabilistic Latent Semantic Analysis
Leonhard Hennig 144

Detection of Opinions and Facts. A Cognitive Approach
Yann Vigile Hoareau, Adil El-Ghali and Charles Tijus 150

Evaluating the Impact of Morphosyntactic Ambiguity in Grammatical Error Detection
Arantza Díaz de Ilarraza, Koldo Gojenola and Maite Oronoz 155

Fast Boosting-based Part-of-Speech Tagging and Text Chunking with Efficient Rule Representation for Sequential Labeling
Tomoya Iwakura 161

Cross-document Event Extraction and Tracking: Task, Evaluation, Techniques and Challenges
Heng Ji, Ralph Grishman, Zheng Chen and Prashant Gupta 166

Co-Parsing with Competitive Models
Lidia Khmylko, Kilian A. Foth and Wolfgang Menzel 173

Robust Compositional Polarity Classification
Manfred Klenner, Stefanos Petrakis and Angela Fahrni 180

Feature Subset Selection in Conditional Random Fields for Named Entity Recognition
Roman Klinger and Christoph M. Friedrich 185

User's Choice of Precision and Recall in Named Entity Recognition
Roman Klinger and Christoph M. Friedrich 192

Semi-Supervised Learning for Word Sense Disambiguation: Quality vs. Quantity
Sandra Kübler and Desislava Zhekova 197

Treelex Meets Adjectival Tables
Anna Kupść 203

Integrating WordNet and FrameNet using a Knowledge-based Word Sense Disambiguation Algorithm
Egoitz Laparra and German Rigau 208

Sampling-based Multilingual Alignment
Adrien Lardilleux and Yves Lepage 214

Using Semantic Networks to Identify Temporal Expressions from Semantic Roles
Hector Llorens, Borja Navarro and Estela Saquete 219

The Design of an Experiment in Anaphora Resolution for Referring Expressions Generation
Diego Jesus de Lucena and Ivandré Paraboni 225

A Model for the Cross-Modal Influence of Visual Context upon Language Procesing
Patrick McCrae 230

Bimodal Corpora Terminology Extraction: Another Brick in the Wall
Claudiu Mihăilă and Dalila Mekhaldi 236

Exploiting Latent Semantic Relations in Highly Linked Hypertext for Information Retrieval in Wikis
Tristan Miller, Bertin Klein and Elisabeth Wolf 241

Large Vocabulary Continuous Speech Recognition for Bulgarian
Petar Mitankin, Stoyan Mihov and Tinko Tinchev 246

Diacritization for Real-World Arabic Texts
Emad Mohamed and Sandra Kübler 251

Multi-entity Sentiment Scoring
Karo Moilanen and Stephen Pulman 258

A Morphological and Syntactic Wide-coverage Lexicon for Spanish: The Leffe
Miguel A. Molinero, Benoît Sagot and Lionel Nicolas 264

How Limited is the Limit?
Prakash Mondal 270

Dependency Parsing and Semantic Role Labeling as a Single Task
Roser Morante, Vincent Van Asch and Antal van den Bosch 275

Structured Output Learning with Polynomial Kernel
Hajime Morita, Hiroya Takamura and Manabu Okumura 281

Unsupervised Word Sense Induction from Multiple Semantic Spaces with Locality Sensitive Hashing
Claire Mouton, Guillaume Pitel, Gaël de Chalendar and Anne Vilnat 287

Unsupervised Extraction of False Friends from Parallel Bi-Texts Using the Web as a Corpus
Svetlin Nakov, Preslav Nakov and Elena Paskaleva 292

Evaluating Term Extraction
Adeline Nazarenko and Haïfa Zargayouna 299

Question Answering over Structured Data: an Entailment-Based Approach to Question Analysis
Matteo Negri and Milen Kouylekov 305

A Semi-supervised Approach for Generating a Table-of-Contents
Viet Cuong Nguyen, Le Minh Nguyen and Akira Shimazu 312

Towards Efficient Production of Linguistic Resources: the Victoria Project
Lionel Nicolas, Miguel A. Molinero, Benoît Sagot, Elena Trigo, Éric De la Clergerie, Miguel Pardo, Jacques Farré and Joan Miquel Vergés 318

A Classification-driven Approach to Document Planning
Rafael Oliveira, Eder Novais, Roberto Araujo and Ivandré Paraboni 324

Interactive Machine Translation Based on Partial Statistical Phrase-based Alignments
Daniel Ortiz-Martínez, Ismael García-Varea and Francisco Casacuberta 330

Topic Modeling of Research Fields: An Interdisciplinary Perspective
Michael Paul and Roxana Girju 337

An Interaction Grammar of Interrogative and Relative Clauses in French
Guy Perrier 343

Comparing Statistical Similarity Measures for Stylistic Multivariate Analysis
Marius Popescu and Liviu P. Dinu 349

From Bag of Languages to Family Trees From Noisy Corpus
Taraka Rama and Anil Kumar Singh 355

Language-Independent Sentiment Analysis Using Subjectivity and Positional Information
Veselin Raychev and Preslav Nakov 360

All Words Unsupervised Semantic Category Labeling for Hindi
Siva Reddy, Abhilash Inumella, Rajeev Sangal and Soma Paul 365

Sentiment Analysis of Figurative Language using a Word Sense Disambiguation Approach
Vassiliki Rentoumi, George Giannakopoulos, Vangelis Karkaletsis and George A. Vouros 370

Global Evaluation of Random Indexing through Swedish Word Clustering Compared to the Peoples Dictionary of Synonyms
Magnus Rosell, Martin Hassel and Viggo Kann 376

Identifying Semantic Relations in Context: Near-misses and Overlaps
Alla Rozovskaya and Roxana Girju 381

Statistical Confidence Measures for Probabilistic Parsing
Ricardo Sánchez-Sáez, Joan-Andreu Sánchez and José Miguel Benedí Ruíz 388

Exploring the Vector Space Model for Finding Verb Synonyms in Portuguese
Luís Sarmento, Paula Carvalho and Eugénio Oliveira 393

A Unified Method for Extracting Simple and Multiword Verbs with Valence Information and Application for Hungarian
Bálint Sass 399

Combining Lexical Resources for Contextual Synonym Expansion
Ravi Sinha and Rada Mihalcea 404

String Distance-Based Stemming of the Highly Inflected Croatian Language
Jan Šnajder and Bojana Dalbelo Bašić 411

Classification of Emotion Words in Russian and Romanian Languages
Marina Sokolova and Victoria Bobicev 416

Classification of Opinions with Non-affective Adverbs and Adjectives
Marina Sokolova and Guy Lapalme 421

Amharic Part-of-Speech Tagger for Factored Language Modeling
Martha Yifiru Tachbelie and Wolfgang Menzel 428

Improving Unsegmented Statistical Dialogue Act Labelling
Vicent Tamarit, Carlos-D. Martínez-Hinarejos and José Miguel Benedí Ruíz 434

Three Issues in Cross-Language Frame Information Transfer
Sara Tonelli and Emanuele Pianta 441

A Study on Linking Wikipedia Categories to Wordnet Synsets using Text Similarity
Antonio Toral, Óscar Ferrández, Eneko Agirre and Rafael Muñoz 449

Ontology Engineering and Knowledge Extraction for Cross-Lingual Retrieval
Jantine Trapman and Paola Monachesi 455

A Method to Restrict the Blow-up of Hypotheses of a Non-disambiguated
Jernej Vičič, Petr Homola and Vladislav Kuboň 460

Sources of Performance in CRF Transfer Training: a Business Name-tagging Case Study
Marc Vilain, Jonathan Huggins and Ben Wellner 465

Extracting Synonyms from Dictionary Definitions
Tong Wang and Graeme Hirst 471

Instance Sampling Methods for Pronoun Resolution
Holger Wunsch, Sandra Kübler and Rachael Cantrell 478

Approximate Matching for Evaluating Keyphrase Extraction
Torsten Zesch and Iryna Gurevych 484

Too Many Mammals: Improving the Diversity of Automatically Recognized Terms
Ziqi Zhang, Lei Xia, Mark A. Greenwood and José Iria 490