International Conference on Recent Advances in Natural Language Processing 2015 (RANLP 2015)

Hissar, Bulgaria
7 – 9 September 2015

Editors:

Galia Angelova
Kalina Bontcheva
Ruslan Mitkov

ISBN: 978-1-5108-1328-1

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© (2015) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2015)

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

POS Tagging for Arabic Tweets
Fahad Albogamy and Allan Ramsay 1

Lexicon-based Sentiment Analysis for Persian Text
Fatemeh Amiri, Simon Scerri and Mohammadhassan Khodashahi 9

Automatic Construction of a TMF Terminological Database using a Transducer Cascade
Chihebeddine Ammar, Kais Haddar and Laurent Romary 17

A Statistical Model for Measuring Structural Similarity between Webpages
Zhenisbek Assylbekov, Assulan Nurkas and Inês Russinho Mouga 24

Predicting the quality of questions on Stackoverflow
Antoaneta Baltadzhieva and Grzegorz Chrupała 32

How Topic Biases Your Results? A Case Study of Sentiment Analysis and Irony Detection in Italian
Francesco Barbieri, Francesco Ronzano and Horacio Saggion 41

Arabic Named Entity Recognition Process using Transducer Cascade and Arabic Wikipedia
Fatma BEN MESMIA, Kais HADDAR, Denis Maurel and Nathalie Friburger 48

DanProof: Pedagogical Spell and Grammar Checking for Danish
Eckhard Bick 55

Maximal Repeats Enhance Substring-based Authorship Attribution
Romain Brixtel 63

Improving Event Detection with Active Learning
Kai Cao, Xiang Li, Miao Fan and Ralph Grishman 72

Improving Event Detection with Dependency Regularization
Kai Cao, Xiang Li and Ralph Grishman 78

Authorship Verification, Average Similarity Analysis
Daniel Castro Castro, Yaritza Adame Arcia, María Pelaez Brioso and Rafael Muñoz Guillena . . 84

Coreference Resolution to Support IE from Indian Classical Music Forums
Joe Cheri and Pushpak Bhattacharyya 91

Readability Assessment of Translated Texts
Alina Maria Ciobanu, Liviu P. Dinu and Flaviu Pepelea 97

Processing and Normalizing Hashtags
Thierry Declerck and Piroska Lendvai 104

Tune Your Brown Clustering, Please
Leon Derczynski, Sean Chester and Kenneth Bøgh 110

Temporal Relation Classification using a Model of Tense and Aspect
Leon Derczynski and Robert Gaizauskas 118

Efficient Named Entity Annotation through Pre-empting
Leon Derczynski and Kalina Bontcheva 123

A Joint Model of Product Properties, Aspects and Ratings for Online Reviews
Ying Ding and Jing Jiang 131

Towards Opinion Summarization from Online Forums
Ying Ding and Jing Jiang 138

Cross-lingual Synonymy Overlap
Anca Dinu, Liviu P. Dinu and Ana Sabina Uban 147

Barbecued Opakapaka: Using Semantic Preferences for Ontology Population
Ismail El Maarouf, Georgiana Marsic and Constantin Orasan 153

Towards a Lexicon-grammar based Framework for NLP: an Opinion Mining Application
Annibale Elia, Serena Pelosi, Alessandro Maisto and Raffaele Guarasci 160

Using the Textual Content of the LMF-Normalized Dictionaries for Identifying and Linking the Syntactic Behaviors to the Meanings
Imen ELLEUCH, Bilel GARGOURI and Abdelmajid BEN HAMADOU 168

Weakly Supervised Definition Extraction
Luis Espinosa Anke, Horacio Saggion and Francesco Ronzano 176

Jointly Embedding Relations and Mentions for Knowledge Population
Miao Fan, Kai Cao, Yifan He and Ralph Grishman 186

Distributional Semantics for Resolving Bridging Mentions
Tim Feuerbach, Martin Riedl and Chris Biemann 192

Text Classification into Abstract Classes Based on Discourse Structure
Boris Galitsky, Dmitry Ilvovsky and Sergey O. Kuznetsov 200

Enriching Word Sense Embeddings with Translational Context
Mehdi Ghanimifard and Richard Johansson 208

Automatic Acquisition of Artifact Nouns in French
Xiaoqin Hu and Pierre-André Buvet 216

Feature-Rich Part-Of-Speech Tagging Using Deep Syntactic and Semantic Analysis
Luchezar Jackov 224

The Application of Constraint Rules to Data-driven Parsing
Sardar Jaf and Allan Ramsay 232

Part-of-Speech Tagging for Code-Mixed English-Hindi Twitter and Facebook Chat Messages
Anupam Jamatia, Björn Gambäck and Amitava Das 239

Sentiment Analysis in Twitter for Macedonian
Dame Jovanoski, Veno Pachovski and Preslav Nakov 249

About Emotion Identification in Visual Sentiment Analysis
Olga Kanishcheva and Galia Angelova 258

Fine-Grained Sentiment Analysis for Movie Reviews in Bulgarian
Borislav Kapukaranov and Preslav Nakov 266

Structural Alignment for Comparison Detection
Wiltrud Kessler and Jonas Kuhn 275

Recognition of Polish Temporal Expressions
Jan Kocoń and Michał Marcińczuk 282

Domain Adaptation with Filtering for Named Entity Extraction of Japanese Anime-Related Words
Kanako Komiya, Daichi EDAMURA, Ryuta TAMURA, Minoru SASAKI, Hiroyuki SHINNOU and Yoshiyuki KOTANI 291

Feature Extraction for Native Language Identification Using Language Modeling
Vincent Kríž, Martin Holub and Pavel Pecina 298

Learning Agglutinative Morphology of Indian Languages with Linguistically Motivated Adaptor Grammars
Arun Kumar, Lluís Padró and Antoni Oliver 307

Statistical Sandhi Splitter and its Effect on NLP Applications
Prathyusha Kuncham, Kovida Nelakuditi and Radhika Mamidi 313

Automatically Identifying Periodic Social Events from Twitter
Florian Kunneman and Antal Van den Bosch 320

Collecting and Evaluating Lexical Polarity with A Game With a Purpose
Mathieu Lafourcade, Alain Joubert and Nathalie Le Brun 329

Medical imaging report indexing: enrichment of index through an algorithm of spreading over a lexico-semantic network
Mathieu Lafourcade and Lionel Ramadier 338

Taxonomy Beats Corpus in Similarity Identification, but Does It Matter?
Minh Le and Antske Fokkens 346

Semantic Parsing via ℓ_0-norm-based Alignment
Zhihua Liao, Qixian Zeng and Qiyun Wang 355

A Supervised Semantic Parsing with Lexicon Extension and Syntactic Constraint
Zhihua Liao, Qixian Zeng and Qiyun Wang 362

Predicting the Level of Text Standardness in User-generated Content
Nikola Ljubešić, Darja Fišer, Tomaž Erjavec, Jaka Čibej, Dafne Marko, Senja Pollak and Iza Škrjanec 371

Predicting Inflectional Paradigms and Lemmata of Unknown Words for Semi-automatic Expansion of Morphological Lexicons
Nikola Ljubešić, Miquel Esplà-Gomis, Filip Klubička and Nives Mikelić Preradović 379

Predicting Correlations Between Lexical Alignments and Semantic Inferences
Simone Magnolini and Bernardo Magnini 388

Learning the Impact of Machine Translation Evaluation Metrics for Semantic Textual Similarity
Simone Magnolini, Ngoc Phuoc An Vo and Octavian Popescu 398

Norwegian Native Language Identification
Shervin Malmasi, Mark Dras and Irina Temnikova 404

Automatic construction of complex features in Conditional Random Fields for Named Entities Recognition
Michał Marcińczuk 413

Pattern Construction for Extracting Domain Terminology
Yusney Marrero García, Paloma Moreda Pozo and Rafael Muñoz-Guillena 420

A Procedural Definition of Multi-word Lexical Units
Marek Maziarz, Stan Szpakowicz and Maciej Piasecki 427

Semi-Supervised Never-Ending Learning in Rhetorical Relation Identification
Erick Galani Maziero, Graeme Hirst and Thiago Alexandre Salgueiro Pardo 436

Exposing Paid Opinion Manipulation Trolls
Todor Mihaylov, Ivan Koychev, Georgi Georgiev and Preslav Nakov 443

Extractive Summarization by Aggregating Multiple Similarities
Olof Mogren, Mikael Kågebäck and Devdatt Dubhashi 451

Statistical Machine Translation Improvement based on Phrase Selection
Cyrine Nasri, Chiraz Latiri and Kamel Smaili 458

A Simple and Efficient Method to Generate Word Sense Representations
Luis Nieto Piña and Richard Johansson 465

TBXTools: A Free, Fast and Flexible Tool for Automatic Terminology Extraction
Antoni Oliver and Mercè Vàzquez 473

Enlarging the Croatian WordNet with WN-Toolkit and Cro-Deriv
Antoni Oliver, Krešimir Šojat and Matea Srebačić 480

A Comparative Study of Different Sentiment Lexica for Sentiment Analysis of Tweets
Canberk Ozdemir and Sabine Bergler 488

Classification of Attributes in a Natural Language Query into Different SQL Clauses
Ashish Palakurthi, Ruthu S M, Arjun Akula and Radhika Mamidi 497

Classifying Idiomatic and Literal Expressions Using Vector Space Representations
Jing Peng, Anna Feldman and Hamza Jazmati 507

Extraction of the Multi-word Lexical Units in the Perspective of the Wordnet Expansion
Maciej Piasecki, Michał Wendelberger and Marek Maziarz 512

A New Approach to Automated Text Readability Classification based on Concept Indexing with Integrated Part-of-Speech n-gram Features
Abigail Razon and John Barnden 521

Classification of Lexical Collocation Errors in the Writings of Learners of Spanish
Sara Rodríguez-Fernández, Roberto Carlini and Leo Wanner 529

Measuring Semantic Similarity for Bengali Tweets Using WordNet
Dwijen Rudrapal, Amitava Das and Baby Bhattacharya 537

Ordering adverbs by their scaling effect on adjective intensity
Josef Ruppenhofer, Jasper Brandes, Petra Steiner and Michael Wiegand 545

bRol: The Parser of Syntactic and Semantic Dependencies for Basque
Haritz Salaberri, Olatz Arregi and Beñat Zapirain 555

Error Analysis and Improving Speech Recognition for Latvian Language
Askars Salimbajevs and Jevgenijs Strigins 563

Towards the Unsupervised Acquisition of Implicit Semantic Roles
Niko Schenk, Christian Chiarcos and Maria Sukhareva 570

Evaluating the Impact of Using a Domain-specific Bilingual Lexicon on the Performance of a Hybrid Machine Translation Approach
Nasredine Semmar, Othman Zennaki and Meriama Laib 579

Hierarchical Topic Structuring: From Dense Segmentation to Topically Focused Fragments via Burst Analysis
Anca-Roxana Simon, Pascale Sébillot and Guillaume Gravier 588

Improving Word Sense Disambiguation with Linguistic Knowledge from a Sense Annotated Treebank
Kiril Simov, Alexander Popov and Petya Osenova 596

Learning Relationship between Authors' Activity and Sentiments: A case study of online medical forums
Marina Sokolova and Victoria Bobicev 604

Translating from Original to Simplified Sentences using Moses: When does it Actually Work?
Sanja Štajner and Horacio Saggion 611

Automatic Text Simplification for Spanish: Comparative Evaluation of Various Simplification Strategies
Sanja Štajner, Iacer Calixto and Horacio Saggion 618

Towards Multilingual Event Extraction Evaluation: A Case Study for the Czech Language
Josef Steinberger and Hristo Tanev 627

A VSM-based Statistical Model for the Semantic Relation Interpretation of Noun-Modifier Pairs
Nitesh Surtani and Soma Paul 636

An LDA-based Topic Selection Approach to Language Model Adaptation for Handwritten Text Recognition
Jafar Tanha, Jesse de Does and Katrien Depuydt 646

Training Automatic Transliteration Models on DBPedia Data
Velislava Todorova and Kiril Simov 654

Arabic-English Semantic Word Class Alignment to Improve Statistical Machine Translation
Ines Turki Khemakhem, Salma Jamoussi and Abdelmajid Ben Hamadou 663

Detection and Fine-Grained Classification of Cyberbullying Events
Cynthia Van Hee, Els Lefever, Ben Verhoeven, Julie Mennes, Bart Desmet, Guy De Pauw, Walter Daelemans and Veronique Hoste 672

A New Approach for Idiom Identification Using Meanings and the Web
Rakesh Verma and Vasanthi Vuppuluri 681

Learning the Impact and Behavior of Syntactic Structure: A Case Study in Semantic Textual Similarity
Ngoc Phuoc An Vo and Octavian Popescu 688

Six Good Predictors of Autistic Text Comprehension
Victoria Yaneva and Richard Evans 697

User Name Disambiguation in Community Question Answering
Baoguo Yang and Suresh Manandhar 707

Discovering Causal Relations in Textual Instructions
Kristina Yordanova 714

A Large Wordnet-based Sentiment Lexicon for Polish
Monika Zaśko-Zielińska, Maciej Piasecki and Stan Szpakowicz 721

Named Entity Recognition of Persons' Names in Arabic Tweets
Omnia Zayed and Samhaa El-Beltagy 731

One Tree is not Enough: Cross-lingual Accumulative Structure Transfer for Semantic Indeterminacy
Patrick Ziering and Lonneke van der Plas 739

Lost in Discussion? Tracking Opinion Groups in Complex Political Discussions by the Example of the FOMC Meeting Transcriptions
Cäcilia Zirn, Robert Meusel and Heiner Stuckenschmidt 747