24th International Conference on Computational Linguistics

(COLING 2012)

Mumbai, India
8-15 December 2012

Volume 1 of 4

ISBN: 978-1-62748-389-6

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© (2012) 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

Multi-Dimensional Feature Merger for Question Answering
Apoorv Agarwal, J William Murdock, Jennifer Chu-Carroll, Adam Lally and Aditya Kalyanpur 1

Unsupervised Discovery of Relations and Discriminative Extraction Patterns
Alan Akbik, Larysa Visengeriyeva, Priska Herger, Holmer Hemsen and Alexander Löser 17

Automatic Detection of Point of View Differences in Wikipedia
Khalid Al Khatib, Hinrich Schütze and Cathleen Kantner 33

SpeedRead: A Fast Named Entity Recognition Pipeline
Rami Al-Rfou' and Steven Skiena 51

Experiments with Term Translation
Mihael Arcan, Christian Federmann and Paul Buitelaar 67

The Floating Arabic Dictionary: An Automatic Method for Updating a Lexical Database through the Detection and Lemmatization of Unknown Words
Mohammed Attia, Younes Samih, Khaled Shaalan and Josef van Genabith 83

Contribution of Complex Lexical Information to Solve Syntactic Ambiguity in Basque
Aitziber Atutxa, Eneko Agirre and Kepa Sarasola 97

Comparative Quality Estimation: Automatic Sentence-Level Ranking of Multiple Machine Translation Outputs
Eleftherios Avramidis 115

Constructing Reference Semantic Predictions from Biomedical Knowledge Sources
Demeke Ayele, Jean-Pierre Chevallet, Million Meshesha and Getnet Kassie 133

Translation Quality-Based Supplementary Data Selection by Incremental Update of Translation Models
Pratyush Banerjee, Sudip Kumar Naskar, Johann Roturier, Andy Way and Josef van Genabith 149

Text Reuse Detection using a Composition of Text Similarity Measures
Daniel Bär, Torsten Zesch and Iryna Gurevych 167

Deriving Paraphrases for Highly Inflected Languages from Comparable Documents
Kfir Bar and Nachum Dershowitz 185

Harvesting Parallel Text in Multiple Languages with Limited Supervision
Luciano Barbosa, Vivek Kumar Rangarajan Sridhar, Mahsa Yarmohammadi and Srinivas Bangalore 201

An Evaluation of Statistical Post-Editing Systems Applied to RBMT and SMT Systems
Hanna Béchara, Raphaël Rubino, Yifan He, Yanjun Ma and Josef van Genabith 215

Prague Dependency Treebank 2.5 – a Revisited Version of PDT 2.0
Eduard Bejček, Jarmila Panevová, Jan Popelka, Pavel Straňák, Magda Ševčíková, Jan Štěpánek and Zdeněk Žabokrtský 231

Deriving a Lexicon for a Precision Grammar from Language Documentation Resources: A Case Study of Chintang
Emily M. Bender, Robert Schikowski and Balthasar Bickel 247

Quantifying Semantics using Complex Network Analysis
Chris Biemann, Stefanie Roos and Karsten Weihe 263

Improvements to Training an RNN parser
Richard Billingsley and James Curran 279

Thread Specific Features are Helpful for Identifying Subjectivity Orientation of Online Forum Threads
Prakhar Biyani, Sumit Bhatia, Cornelia Caragea and Prasenjit Mitra 295

Natural Language Generation for Nature Conservation: Automating Feedback to Help Volunteers Identify Bumblebee Species
Steven Blake, Advaith Siddharthan, Hien Nguyen, Nirwan Sharma, Anne-Marie Robinson, Elaine O'Mahony, Ben Darvill, Chris Mellish and Rene van der Wal 311

Studying the Effect of Input Size for Bayesian Word Segmentation on the Providence Corpus
Benjamin Börschinger, Katherine Demuth and Mark Johnson 325

Bayesian Language Modelling of German Compounds
Jan A. Botha, Chris Dyer and Phil Blunsom 341

Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish
Stefan Bott, Luz Rello, Biljana Drndarevic and Horacio Saggion 357

Identification of Social Acts in Dialogue
David Bracewell, Marc Tomlinson and Hui Wang 375

Robust, Lexicalized Native Language Identification
Julian Brooke and Graeme Hirst 391

Identifying Urdu Complex Predication via Bigram Extraction
Miriam Butt, Tina Bögel, Annette Hautli, Sebastian Sulger and Tafseer Ahmed 409

Native Language Identification using Recurring n-grams – Investigating Abstraction and Domain Dependence
Serhiy Bykh and Detmar Meurers 425

Analysis and Enhancement of Wikification for Microblogs with Context Expansion
Taylor Cassidy, Heng Ji, Lev-Arie Ratinov, Arkaitz Zubiaga and Hongzhao Huang ... 441

On the Effectiveness of using Sentence Compression Models for Query-Focused Multi-Document Summarization
Yllias Chali and Sadid A. Hasan 457

Towards Automatic Topical Question Generation
Yllias Chali and Sadid A. Hasan 475

Adjective Deletion for Linguistic Steganography and Secret Sharing
Ching-Yun Chang and Stephen Clark 493

The Secret's in the Word Order: Text-to-Text Generation for Linguistic Steganography
Ching-Yun Chang and Stephen Clark 511

Joint Modeling for Chinese Event Extraction with Rich Linguistic Features
Chen Chen and Vincent Ng 529

A Simplification-Translation-Restoration Framework for Cross-Domain SMT Applications
Han-Bin Chen, Hen-Hsen Huang, Hsin-Hsi Chen and Ching-Ting Tan 545

A Semi-Supervised Bayesian Network Model for Microblog Topic Classification
Yan Chen, Zhoujun Li, Liqiang Nie, Xia Hu, Xiangyu Wang, Tat-Seng Chua and Xiaoming Zhang 561

A System for Multilingual Sentiment Learning On Large Data Sets
Alex Cheng and Oles Zhulyn 577

Extraction of Russian Sentiment Lexicon for Product Meta-Domain
Ilia Chetviorkin and Natalia Loukachevitch 593

Problems in Evaluating Grammatical Error Detection Systems
Martin Chodorow, Markus Dickinson, Ross Israel and Joel Tetreault 611

Unsupervised and Semi-Supervised Morphological Analysis for Information Retrieval in the Biomedical Domain
Vincent Claveau 629

A Hybrid Approach to Finding Phenotype Candidates in Genetic Texts
Nigel Collier, Mai-Vu Tran, Hoang-Quynh Le, Anika Oellrich, Ai Kawazoe, Martin Hall-May and Dietrich Rebholz-Schuhmann 647

Using Argumentative Zones for Extractive Summarization of Scientific Articles
Danish Contractor, Yufan Guo and Anna Korhonen 663

Annotation Tools and Knowledge Representation for a Text-To-Scene System
Bob Coyne, Alex Klapheke, Masoud Rouhizadeh, Richard Sproat and Daniel Bauer . 679

Towards Efficient HPSG Generation for German, a Non-Configurational Language
Berthold Crysmann and Woodley Packard 695

A Corpus-Based Study of Edit Categories in Featured and Non-Featured Wikipedia Articles
Johannes Daxenberger and Iryna Gurevych 711

A Computational Cognitive Model for Semantic Sub-Network Extraction from Natural Language Queries
Suman Deb Roy and Wenjun Zeng 727

Extraction of Domain-Specific Bilingual Lexicon from Comparable Corpora: Compositional Translation and Ranking
Estelle Delpech, Béatrice Daille, Emmanuel Morin and Claire Lemaire 745

Twitter Topic Summarization by Ranking Tweets using Social Influence and Content Quality
Yajuan Duan, Zhumin Chen, Furu Wei, Ming Zhou and Heung-Yeung Shum 763

S-Restricted Monotone Alignments: Algorithm, Search Space, and Applications
Steffen Eger 781

Mining Words in the Minds of Second Language Learners: Learner-Specific Word Difficulty
Yo Ehara, Issei Sato, Hidekazu Oiwa and Hiroshi Nakagawa 799

Jointly Disambiguating and Clustering Concepts and Entities with Markov Logic
Angela Fahrni and Michael Strube 815

Flexible Structural Analysis of Near-Meet-Semilattices for Typed Unification-Based Grammar Design
Rouzbeh Farahmand and Gerald Penn 833

Stacking of Dependency and Phrase Structure Parsers
Richárd Farkas and Bernd Bohnet 849

Semantic Cohesion Model for Phrase-Based SMT
Minwei Feng, Weiwei Sun and Hermann Ney 867

Comparing Taxonomies for Organising Collections of Documents
Samuel Fernando, Mark Hall, Eneko Agirre, Aitor Soroa, Paul Clough and Mark Stevenson 879

Modeling the Complexity of Manual Annotation Tasks: a Grid of Analysis
Karën Fort, Adeline Nazarenko and Sophie Rosset 895

Extractive Multi-Document Summarization with Integer Linear Programming and Support Vector Regression
Dimitrios Galanis, Gerasimos Lampouras and Ion Androutsopoulos 911

Cross-Lingual Topical Relevance Models
Debasis Ganguly, Johannes Leveling and Gareth Jones 927

Structured Term Recognition in Medical Text
Michael Glass and Alfio Gliozzo 943

A Dynamic Oracle for Arc-Eager Dependency Parsing
Yoav Goldberg and Joakim Nivre 959

Statistical Mechanical Analysis of Semantic Orientations on Lexical Network
Takuma Goto, Yoshiyuki Kabashima and Hiroya Takamura 977

Finding Thoughtful Comments from Social Media
Swapna Gottipati and Jing Jiang 995

A Distributed Platform for Sanskrit Processing
Pawan Goyal, Gérard Huet, Amba Kulkarni, Peter Scharf and Ralph Bunker 1011

Understanding the Performance of Statistical MT Systems: A Linear Regression Framework
Francisco Guzman and Stephan Vogel 1029

Geolocation Prediction in Social Media Data by Finding Location Indicative Words
Bo Han, Paul Cook and Timothy Baldwin 1045

Readability Classification for German using Lexical, Syntactic, and Morphological Features
Julia Hancke, Sowmya Vajjala and Detmar Meurers 1063

Walk-based Computation of Contextual Word Similarity
Kazuo Hara, Ikumi Suzuki, Masashi Shimbo and Yuji Matsumoto 1081

Flexible Japanese Sentence Compression by Relaxing Unit Constraints
Jun Harashima and Sadao Kurohashi 1097

Approximating Theoretical Linguistics Classification in Real Data: the Case of German "nach" Particle Verbs
Boris Haselbach, Kerstin Eckart, Wolfgang Seeker, Kurt Eberle and Ulrich Heid 1113

Bridging the Gap between Intrinsic and Perceived Relevance in Snippet Generation
Jing He, Pablo Duboue and Jian-Yun Nie 1129

A Comparison and Improvement of Online Learning Algorithms for Sequence Labeling
Zhengyan He and Houfeng Wang 1147

Creating an Extended Named Entity Dictionary from Wikipedia
Ryuichiro Higashinaka, Kugatsu Sadamitsu, Kuniko Saito, Toshiro Makino and Yoshihiro Matsuo 1163

Statistical Method of Building Dialect Language Models for ASR Systems
Naoki Hirayama, Shinsuke Mori and Hiroshi G. Okuno 1179

Tailored Feature Extraction for Lexical Disambiguation of English Verbs Based on Corpus Pattern Analysis
Martin Holub, Vincent Kríž, Silvie Cinková and Eckhard Bick 1195

Method Mention Extraction from Scientific Research Papers
Hospice Houngbo and Robert E. Mercer 1211

Context-Enhanced Personalized Social Summarization
Po Hu, Donghong Ji, Chong Teng and Yujing Guo 1223

Tweet Ranking Based on Heterogeneous Networks
Hongzhao Huang, Arkaitz Zubiaga, Heng Ji, Hongbo Deng, Dong Wang, Hieu Le, Tarek Abdelzaher, Jiawei Han, Alice Leung, John Hancock and Clare Voss 1239

Improved Combinatory Categorial Grammar Induction with Boundary Words and Bayesian Inference
Yun Huang, Min Zhang and Chew-Lim Tan 1257

Mining Rules for Rewriting States in a Transition-based Dependency Parser for English
Akihiro Inokuchi and Ayumu Yamaoka 1275

Coreference Resolution with ILP-based Weighted Abduction
Naoya Inoue, Ekaterina Ovchinnikova, Kentaro Inui and Jerry Hobbs 1291

N-gram Fragment Sequence Based Unsupervised Domain-Specific Document Readability
Shoaib Jameel, Xiaojun Qian and Wai Lam 1309

Using Knowledge and Constraints To Find the Best Antecedent
Prateek Jindal and Dan Roth 1327

Towards a Generic and Flexible Citation Classifier Based on a Faceted Classification Scheme
Charles Jochim and Hinrich Schütze 1343

Semantics-Based Machine Translation with Hyperedge Replacement Grammars
Bevan Jones, Jacob Andreas, Daniel Bauer, Karl Moritz Hermann and Kevin Knight 1359

Answering Yes/No Questions via Question Inversion
Hiroshi Kanayama, Yusuke Miyao and John Prager 1377

Improving Topic Classification for Highly Inflective Languages
Jurgita Kapociute-Dzikiene, Frederik Vaassen, Walter Daelemans and Algis Krupavičius .. 1393

Generating "A for Alpha" When There Are Thousands of Characters
Hiroaki Kawasaki, Ryohei Sasano, Hiroya Takamura and Manabu Okumura 1411

A Machine Learning Approach for Phenotype Name Recognition
Maryam Khordad, Robert E Mercer and Peter Rogan 1425

Improving Combinatory Categorial Grammar Parse Reranking with Dependency Grammar Features
Sunghwan Mac Kim, Dominick Ng, Mark Johnson and James Curran 1441

Inducing Crosslingual Distributed Representations of Words
Alexandre Klementiev, Ivan Titov and Binod Bhattarai 1459

Exploring Local and Global Semantic Information for Event Pronoun Resolution
Fang Kong and Guodong Zhou 1475

Semantic Processing of Compounds in Indian Languages
Amba Kulkarni, Soma Paul, Malhar Kulkarni, Anil Kumar and Nitesh Surtani 1489

Unsupervised Japanese-Chinese Opinion Word Translation using Dependency Distance and Feature-Opinion Association Weight
Guo-Hau Lai, Ying-Mei Guo and Richard Tzong-Han Tsai 1503

On-line Trend Analysis with Topic Models: #twitter Trends Detection Topic Model Online
Jey Han Lau, Nigel Collier and Timothy Baldwin 1519

Learning Compositional Semantics for Open Domain Semantic Parsing
Phong Le and Willem Zuidema 1535

Evaluating Different Methods for Automatically Collecting Large General Corpora for Basque from the Web
Igor Leturia 1553

Approximate Sentence Retrieval for Scalable and Efficient Example-Based Machine Translation
Johannes Leveling, Debasis Ganguly, Sandipan Dandapat and Gareth Jones 1571

Improving Text Normalization using Character-Blocks Based Models and System Combination
Chen Li and Yang Liu 1587

Update Summarization using a Multi-level Hierarchical Dirichlet Process Model
Jiwei Li, Sujian Li, Xun Wang, Ye Tian and Baobao Chang 1603

Employing Morphological Structures and Sememes for Chinese Event Extraction
Peifeng Li and Guodong Zhou 1619

Joint Modeling of Trigger Identification and Event Type Determination in Chinese Event Extraction
Peifeng Li, Qiaoming Zhu, Hongjun Diao and Guodong Zhou 1635

Integrating Surface and Abstract Features for Robust Cross-Domain Chinese Word Segmentation
Xiaoqing Li, Kun Wang, Chengqing Zong and Keh-Yih Su 1653

Code-Switch Language Model with Inversion Constraints for Mixed Language Speech Recognition
Ying Li and Pascale Fung 1671

A Separately Passive-Aggressive Training Algorithm for Joint POS Tagging and Dependency Parsing
Zhenghua Li, Min Zhang, Wanxiang Che and Ting Liu 1681

Graph-Based Multi-Tweet Summarization using Social Signals
Xiaohua Liu, Yitong Li, Furu Wei and Ming Zhou 1699

Topical Word Trigger Model for Keyphrase Extraction
Zhiyuan Liu, Chen Liang and Maosong Sun 1715

Easy-First Chinese POS Tagging and Dependency Parsing
Ji Ma, Tong Xiao, Jingbo Zhu and Feiliang Ren 1731

Recognizing Personal Characteristics of Readers using Eye-Movements and Text Features
Pascual Martínez-Gómez, Tadayoshi Hara and Akiko Aizawa 1747

To Exhibit is not to Loiter: A Multilingual, Sense-Disambiguated Wiktionary for Measuring Verb Similarity
Christian M. Meyer and Iryna Gurevych 1763

Using Distributional Similarity for Lexical Expansion in Knowledge-based Word Sense Disambiguation
Tristan Miller, Chris Biemann, Torsten Zesch and Iryna Gurevych 1781

Revising the Compositional Method for Terminology Acquisition from Comparable Corpora
Emmanuel Morin and Béatrice Daille 1797

Is Bad Structure Better Than No Structure?: Unsupervised Parsing for Realisation Ranking
Yasaman Motazedi, Mark Dras and François Lareau 1811

Analysis of Linguistic Style Accommodation in Online Debates
Arjun Mukherjee and Bing Liu 1831

Sentiment Analysis in Twitter with Lightweight Discourse Analysis
Subhabrata Mukherjee and Pushpak Bhattacharyya 1847

YouCat: Weakly Supervised Youtube Video Categorization System from Meta Data & User Comments using WordNet & Wikipedia
Subhabrata Mukherjee and Pushpak Bhattacharyya 1865

Constrained Decoding for Text-Level Discourse Parsing
Philippe Muller, Stergos Afantenos, Pascal Denis and Nicholas Asher 1883

Incremental Learning of Affix Segmentation
Wondwossen Mulugeta, Michael Gasser and Baye Yimam 1901

Semi-Supervised Noun Compound Analysis with Edge and Span Features
Yugo Murawaki and Sadao Kurohashi 1915

Learning Effective and Interpretable Semantic Models using Non-Negative Sparse Embedding
Brian Murphy, Partha Talukdar and Tom Mitchell 1933

Combining Wordnet and Morphosyntactic Information in Terminology Clustering
Agnieszka Mykowiecka and Malgorzata Marciniak 1951

Alignment by Bilingual Generation and Monolingual Derivation
Toshiaki Nakazawa and Sadao Kurohashi 1963

Optimizing for Sentence-Level BLEU+1 Yields Short Translations
Preslav Nakov, Francisco Guzman and Stephan Vogel 1979

Grammarless Parsing for Joint Inference
Jason Naradowsky, Tim Vieira and David Smith 1995

Error Mining with Suspicion Trees: Seeing the Forest for the Trees
Shashi Narayan and Claire Gardent 2011

Structure-Driven Lexicalist Generation
Shashi Narayan and Claire Gardent 2027

A Comparison of Syntactic Reordering Methods for English-German Machine Translation
Jiri Navratil, Karthik Visweswariah and Ananthakrishnan Ramanathan 2043

Grounded Language Acquisition: A Minimal Commitment Approach
Sushobhan Nayak and Amitabha Mukerjee 2059

Bayesian Text Segmentation for Index Term Identification and Keyphrase Extraction
David Newman, Nagendra Koilada, Jey Han Lau and Timothy Baldwin 2077

Exploiting Category-Specific Information for Multi-Document Summarization
Jun-Ping Ng, Praveen Bysani, Ziheng Lin, Min-Yen Kan and Chew-Lim Tan 2093

Improved Temporal Relation Classification using Dependency Parses and Selective Crowdsourced Annotations
Jun-Ping Ng and Min-Yen Kan 2109

Accurate Unbounded Dependency Recovery using Generalized Categorial Grammars
Luan Nguyen, Marten Van Schijndel and William Schuler 2125

Tibetan Base Noun Phrase Identification Framework Based on Chinese-Tibetan Sentence Aligned Corpus
Ming Hua Nuo, Hui Dan Liu, Wei Na Zhao, Long Long Ma, Jian Wu and Zhi Ming Ding .. 2141

A Pipeline Arabic Named Entity Recognition using a Hybrid Approach
Mai Oudah and Khaled Shaalan 2159

Attribute Extraction from Conjectural Queries
Marius Pasca 2177

A Comprehensive Analysis of Constituent Coordination for Grammar Engineering
Agnieszka Patejuk and Adam Przepiórkowski 2191

Simple and Effective Parameter Tuning for Domain Adaptation of Statistical Machine Translation
Pavel Pecina, Antonio Toral and Josef van Genabith 2209

A Supervised Aggregation Framework for Multi-Document Summarization
Yulong Pei, Wenpeng Yin, Qifeng Fan and Lian'en Huang 2225

Collective Search for Concept Disambiguation
Anja Pilz and Gerhard Paaß 2243

Who's (Really) the Boss? Perception of Situational Power in Written Interactions
Vinodkumar Prabhakaran, Owen Rambow and Mona Diab 2259

Bilingual Lexicon Construction from Comparable Corpora via Dependency Mapping
Longhua Qian, Hongling Wang, Guodong Zhou and Qiaoming Zhu 2275

A MWE Acquisition and Lexicon Builder Web Service
Valeria Quochi, Francesca Frontini and Francesco Rubino 2291

A Diverse Dirichlet Process Ensemble for Unsupervised Induction of Syntactic Categories
Roi Reichart, Gal Elidan and Ari Rappoport 2307

From Finite-State to Inversion Transductions: Toward Unsupervised Bilingual Grammar Induction
Markus Saers, Karteek Addanki and Dekai Wu 2325

Underspecified Query Refinement via Natural Language Question Generation
Hassan Sajjad, Patrick Pantel and Michael Gamon 2341

Joint English Spelling Error Correction and POS Tagging for Language Learners Writing
Keisuke Sakaguchi, Tomoya Mizumoto, Mamoru Komachi and Yuji Matsumoto 2357

Automatic Detection of Psychological Distress Indicators and Severity Assessment from Online Forum Posts
Shirin Saleem, Rohit Prasad, Shiv Vitaladevuni, Maciej Pacula, Michael Crystal, Brian Marx, Denise Sloan, Jennifer Vasterling and Theodore Speroff 2375

Ant Colony Algorithm for the Unsupervised Word Sense Disambiguation of Texts: Comparison and Evaluation
Didier Schwab, Jérôme Goulian, Andon Tchechmedjiev and Hervé Blanchon 2389

Learnability-Based Syntactic Annotation Design
Roy Schwartz, Omri Abend and Ari Rappoport 2405

Improving Supervised Sense Disambiguation with Web-Scale Selectors
H. Andrew Schwartz, Fernando Gomez and Lyle Ungar 2423

The French Social Media Bank: a Treebank of Noisy User Generated Content
Djamé Seddah, Benoit Sagot, Marie Candito, Virginie Mouilleron and Vanessa Combet . . . 2441

Initial Explorations on using CRFs for Turkish Named Entity Recognition
Gökhan Akın Şeker and Gülşen Eryiğit 2459

Differential Evolution Based Feature Selection and Classifier Ensemble for Named Entity Recognition
Utpal Kumar Sikdar, Asif Ekbal and Sriparna Saha 2475

Noun Group and Verb Group Identification for Hindi
Smriti Singh, Om P. Damani and Vaijayanthi M. Sarma 2491

Named Entity Recognition System for Urdu
UmrinderPal Singh, Vishal Goyal and Gurpreet Singh Lehal 2507

Easy-first Coreference Resolution
Veselin Stoyanov and Jason Eisner 2519

Modeling Leadership and Influence in Multi-party Online Discourse
Tomek Strzalkowski, Samira Shaikh, Ting Liu, George Aaron Broadwell, Jenny Stromer-Galley, Sarah Taylor, Umit Boz, Veena Ravishankar and Xiaoai Ren 2535

NEER: An Unsupervised Method for Named Entity Evolution Recognition
Nina Tahmasebi, Gerhard Gossen, Nattiya Kanhabua, Helge Holzmann and Thomas Risse 2553

Evaluating the Translation Accuracy of a Novel Language-Independent MT Methodology
George Tambouratzis, Sokratis Sofianopoulos and Marina Vassiliou 2569

Native Tongues, Lost and Found: Resources and Empirical Evaluations in Native Language Identification
Joel Tetreault, Daniel Blanchard, Aoife Cahill and Martin Chodorow 2585

Inverse Document Density: A Smooth Measure for Location-Dependent Term Irregularities
Dennis Thom, Harald Bosch and Thomas Ertl 2603

Efficient Discrimination Between Closely Related Languages
Jörg Tiedemann and Nikola Ljubešić 2619

Semi-Supervised Semantic Role Labeling: Approaching from an Unsupervised Perspective
Ivan Titov and Alexandre Klementiev 2635

Hunting for Entailing Pairs in the Penn Discourse Treebank
Sara Tonelli and Elena Cabrio 2653

Implicitness of Discourse Relations
Fatemeh Torabi Asr and Vera Demberg 2669

Combining Statistical Translation Techniques for Cross-Language Information Retrieval
Ferhan Ture, Jimmy Lin and Douglas Oard 2685

Multi-way Tensor Factorization for Unsupervised Lexical Acquisition
Tim van de Cruys, Laura Rimell, Thierry Poibeau and Anna Korhonen 2703

Sub-corpora Sampling with an Application to Bilingual Lexicon Extraction
Ivan Vulić and Marie-Francine Moens 2721

The Utility of Discourse Structure in Identifying Resolved Threads in Technical User Forums
Li Wang, Su Nam Kim and Timothy Baldwin 2739

Implicit Discourse Relation Recognition by Selecting Typical Training Examples
Xun Wang, Sujian Li, Jiwei Li and Wenjie Li 2757

Chinese Evaluative Information Analysis
Yiou Wang, Jun'ichi Kazama, Takuya Kawada and Kentaro Torisawa 2773

Harnessing the CRF Complexity with Domain-Specific Constraints. The Case of Morphosyntactic Tagging of a Highly Inflected Language
Jakub Waszczuk 2789

A Latent Discriminative Model for Compositional Entailment Relation Recognition using Natural Logic
Yotaro Watanabe, Junta Mizuno, Eric Nichols, Naoaki Okazaki and Kentaro Inui... 2805

Strategies for Mixed-Initiative Conversation Management using Question-Answer Pairs
Wilson Wong, Lawrence Cavedon, John Thangarajah and Lin Padgham 2821

Factored Language Model based on Recurrent Neural Network
Youzheng Wu, Xugang Lu, Hitoshi Yamamoto, Shigeki Matsuda, Chiori Hori and Hideki Kashioka..... 2835

Multi-View AdaBoost for Multilingual Subjectivity Analysis
Min Xiao and Yuhong Guo 2851

Semi-supervised Representation Learning for Domain Adaptation using Dynamic Dependency Networks
Min Xiao, Yuhong Guo and Alexander Yates 2867

Unsupervised Discriminative Induction of Synchronous Grammar for Machine Translation
Xinyan Xiao, Deyi Xiong, Yang Liu, Qun Liu and Shouxun Lin 2883

Paraphrasing for Style
Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman and Colin Cherry..... 2899

Modeling ESL Word Choice Similarities By Representing Word Intensions and Extensions
Huichao Xue and Rebecca Hwa..... 2915

ISO-TimeML Event Extraction in Persian Text
Yadollah Yaghoobzadeh, Gholamreza Ghassem-Sani, Seyed Abolghassem Mirroshandel and Mahbaneh Eshaghzadeh 2931

Measuring the Similarity between TV Programs using Semantic Relations
Ichiro Yamada, Masaru Miyazaki, Hideki Sumiyoshi, Atsushi Matsui, Hironori Furumiya and Hideki Tanaka..... 2945

RelationListwise for Query-Focused Multi-Document Summarization
Wenpeng Yin, Lifu Huang, Yulong Pei and Lian'en Huang 2961

SentTopic-MultiRank: a Novel Ranking Model for Multi-Document Summarization
Wenpeng Yin, Yulong Pei, Fan Zhang and Lian'en Huang..... 2977

Language Modeling for Spoken Dialogue System based on Filtering using Predicate-Argument Structures
Koichiro Yoshino, Shinsuke Mori and Tatsuya Kawahara 2993

Detecting Word Ordering Errors in Chinese Sentences for Learning Chinese as a Foreign Language
Chi-Hsin Yu and Hsin-Hsi Chen 3003

Machine Translation by Modeling Predicate-Argument Structure Transformation
Feifei Zhai, Jiajun Zhang, Yu Zhou and Chengqing Zong 3019

Tree-based Translation without using Parse Trees
Feifei Zhai, Jiajun Zhang, Yu Zhou and Chengqing Zong 3037

Constructing Chinese Abbreviation Dictionary: A Stacked Approach
Longkai Zhang, Sujian Li, Houfeng Wang, Ni Sun and Xinfan Meng 3055

Stacking Heterogeneous Joint Models of Chinese POS Tagging and Dependency Parsing
Meishan Zhang, Wanxiang Che, Ting Liu and Zhenghua Li 3071

A Lazy Learning Model for Entity Linking using Query-Specific Information
Wei Zhang, Jian Su, Chew-Lim Tan, Yunbo Cao and Chin-Yew Lin 3089

The Use of Dependency Relation Graph to Enhance the Term Weighting in Question Retrieval
Weinan Zhang, Zhaoyan Ming, Yu Zhang, Liqiang Nie, Ting Liu and Tat-Seng Chua 3105

Long-Tail Distributions and Unsupervised Learning of Morphology
Qiuye Zhao and Mitch Marcus 3121

User Behaviors Lend a Helping Hand: Learning Paraphrase Query Patterns from Search Log Sessions
Shiqi Zhao, Haifeng Wang and Ting Liu 3137

Exploiting Bilingual Translation for Question Retrieval in Community-Based Question Answering
Guangyou Zhou, Kang Liu and Jun Zhao 3153

Exploiting Lexical Dependencies from Large-Scale Data for Better Shift-Reduce Constituency Parsing
Muhua Zhu, Jingbo Zhu and Huizhen Wang 3171