11th Conference on Empirical Methods in Natural Language Processing 2006

(EMNLP 2006)

Held in conjunction with COLING/ACL 2006

Sydney, Australia
22-23 July 2006

ISBN: 978-1-61738-500-1

Printed from e-media with permission by:

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

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

Copyright© (2006) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2010)

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

Table of Contents

Preface ix

Organizers xi

Conference Program Outline xiii

Conference Program in Detail xiv

Unsupervised Discovery of a Statistical Verb Lexicon
Trond Grenager and Christopher D. Manning 1

An Empirical Approach to the Interpretation of Superlatives
Johan Bos and Malvina Nissim 9

Paraphrase Recognition via Dissimilarity Significance Classification
Long Qiu, Min-Yen Kan and Tat-Seng Chua 18

Detecting Parser Errors Using Web-based Semantic Filters
Alexander Yates, Stefan Schoenmackers and Oren Etzioni 27

Distributional measures of concept-distance: A task-oriented evaluation
Saif Mohammad and Graeme Hirst 35

SPMT: Statistical Machine Translation with Syntactified Target Language Phrases
Daniel Marcu, Wei Wang, Abdessamad Echihabi and Kevin Knight 44

Phrasetable Smoothing for Statistical Machine Translation
George Foster, Roland Kuhn and Howard Johnson 53

The impact of parse quality on syntactically-informed statistical machine translation
Chris Quirk and Simon Corston-Oliver 62

Statistical Machine Reordering
Marta R. Costa-jussà and José A. R. Fonollosa 70

Re-evaluating Machine Translation Results with Paraphrase Support
Liang Zhou, Chin-Yew Lin and Eduard Hovy 77

Exploiting Discourse Structure for Spoken Dialogue Performance Analysis
Mihai Rotaru and Diane J. Litman 85

Learning Information Status of Discourse Entities
Malvina Nissim 94

Automatic classification of citation function
Simone Teufel, Advaith Siddharthan and Dan Tidhar 103

Is it Really that Difficult to Parse German?
Sandra Kübler, Erhard W. Hinrichs and Wolfgang Maier 111

Domain Adaptation with Structural Correspondence Learning
John Blitzer, Ryan McDonald and Fernando Pereira 120

Incremental Integer Linear Programming for Non-projective Dependency Parsing
Sebastian Riedel and James Clarke 129

Semantic Role Labeling of NomBank: A Maximum Entropy Approach
Zheng Ping Jiang and Hwee Tou Ng 138

Identification of Event Mentions and their Semantic Class
Steven Bethard and James H. Martin 146

Extremely Lexicalized Models for Accurate and Fast HPSG Parsing
Takashi Ninomiya, Takuya Matsuzaki, Yoshimasa Tsuruoka, Yusuke Miyao and Jun'ichi Tsujii 155

Multilingual Deep Lexical Acquisition for HPSGs via Supertagging
Phil Blunsom and Timothy Baldwin 164

Lexical Reference: a Semantic Matching Subtask
Oren Glickman, Eyal Shnarch and Ido Dagan 172

Semantic Role Labeling via Instance-Based Learning
Chi-san Althon Lin and Tony C. Smith 180

Inducing Temporal Graphs
Philip Bramsen, Pawan Deshpande, Yoong Keok Lee and Regina Barzilay 189

A Weakly Supervised Learning Approach for Spoken Language Understanding
Wei-Lin Wu, Ru-Zhan Lu, Jian-Yong Duan, Hui Liu, Feng Gao and Yu-Quan Chen 199

*Humor: Prosody Analysis and Automatic Recognition for F*R*I*E*N*D*S**
Amruta Purandare and Diane Litman 208

Distributed Language Modeling for N-best List Re-ranking
Ying Zhang, Almut Silja Hildebrand and Stephan Vogel 216

Efficient Search for Inversion Transduction Grammar
Hao Zhang and Daniel Gildea 224

A Discriminative Model for Tree-to-Tree Translation
Brooke Cowan, Ivona Kučerová and Michael Collins 232

Modeling Impression in Probabilistic Transliteration into Chinese
LiLi Xu, Atsushi Fujii and Tetsuya Ishikawa 242

Unsupervised Named Entity Transliteration Using Temporal and Phonetic Correlation
Tao Tao, Su-Youn Yoon, Andrew Fister, Richard Sproat and ChengXiang Zhai 250

Capturing Out-of-Vocabulary Words in Arabic Text
Abdusalam F.A. Nwesri, S.M.M. Tahaghoghi and Falk Scholer 258

Using linguistically motivated features for paragraph boundary identification
Katja Filippova and Michael Strube 267

BESTCUT: A Graph Algorithm for Coreference Resolution
Cristina Nicolae and Gabriel Nicolae 275

Automatic Construction of Predicate-argument Structure Patterns for Biomedical Information Extraction
Akane Yakushiji, Yusuke Miyao, Tomoko Ohta, Yuka Tateisi and Jun'ichi Tsujii 284

Protein folding and chart parsing
Julia Hockenmaier, Aravind K. Joshi and Ken A. Dill 293

Learning Phrasal Categories
William P. Headden III, Eugene Charniak and Mark Johnson 301

Priming Effects in Combinatory Categorial Grammar
David Reitter, Julia Hockenmaier and Frank Keller 308

Better Informed Training of Latent Syntactic Features
Markus Dreyer and Jason Eisner 317

Get out the vote: Determining support or opposition from Congressional floor-debate transcripts
Matt Thomas, Bo Pang and Lillian Lee 327

Partially Supervised Coreference Resolution for Opinion Summarization through Structured Rule Learning
Veselin Stoyanov and Claire Cardie 336

Sentiment Retrieval using Generative Models
Koji Eguchi and Victor Lavrenko 345

Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis
Hiroshi Kanayama and Tetsuya Nasukawa 355

A Skip-Chain Conditional Random Field for Ranking Meeting Utterances by Importance
Michel Galley 364

Style & Topic Language Model Adaptation Using HMM-LDA
Bo-June (Paul) Hsu and James Glass 373

Text data acquisition for domain-specific language models
Abhinav Sethy, Panayiotis G. Georgiou and Shrikanth Narayanan 382

Corrective Models for Speech Recognition of Inflected Languages
Izhak Shafran and Keith Hall 390

Lexicon Acquisition for Dialectal Arabic Using Transductive Learning
Kevin Duh and Katrin Kirchhoff 399

Arabic OCR Error Correction Using Character Segment Correction, Language Modeling, and Shallow Morphology
Walid Magdy and Kareem Darwish 408

Partially Supervised Sense Disambiguation by Learning Sense Number from Tagged and Untagged Corpora
Zheng-Yu Niu, Dong-Hong Ji and Chew Lim Tan 415

Automatically Assessing Review Helpfulness
Soo-Min Kim, Patrick Pantel, Tim Chklovski and Marco Pennacchiotti 423

Joint Extraction of Entities and Relations for Opinion Recognition
Yejin Choi, Eric Breck and Claire Cardie 431

Feature Subsumption for Opinion Analysis
Ellen Riloff, Siddharth Patwardhan and Janyce Wiebe 440

Relevance Feedback Models for Recommendation
Masao Utiyama and Mikio Yamamoto 449

Random Indexing using Statistical Weight Functions
James Gorman and James R. Curran 457

A Hybrid Markov/Semi-Markov Conditional Random Field for Sequence Segmentation
Galen Andrew 465

Boosting Unsupervised Relation Extraction by Using NER
Ronen Feldman and Benjamin Rosenfeld 473

Short Text Authorship Attribution via Sequence Kernels, Markov Chains and Author Unmasking: An Investigation
Conrad Sanderson and Simon Guenter 482

Entity Annotation based on Inverse Index Operations
Ganesh Ramakrishnan, Sreeram Balakrishnan and Sachindra Joshi 492

Unsupervised Information Extraction Approach Using Graph Mutual Reinforcement
Hany Hassan, Ahmed Hassan and Ossama Emam 501

Empirical Study on the Performance Stability of Named Entity Recognition Model across Domains
Hong Lei Guo, Li Zhang and Zhong Su 509

Statistical Ranking in Tactical Generation
Erik Velldal and Stephan Oepen 517

Sentence ordering with manifold-based classification in multi-document summarization
Paul D Ji and Stephen Pulman 526

Quality Assessment of Large Scale Knowledge Resources
Montse Cuadros and German Rigau 534

Graph-based Word Clustering using a Web Search Engine
Yutaka Matsuo, Takeshi Sakaki, Kôki Uchiyama and Mitsuru Ishizuka 542

Context-Dependent Term Relations for Information Retrieval
Jing Bai, Jian-Yun Nie and Guihong Cao 551

Loss Minimization in Parse Reranking
Ivan Titov and James Henderson 560

Unsupervised Relation Disambiguation with Order Identification Capabilities
Jinxiu Chen, Donghong Ji, ChewLim Tan and Zhengyu Niu 568

Competitive generative models with structure learning for NLP classification tasks
Kristina Toutanova 576

Two graph-based algorithms for state-of-the-art WSD
Eneko Agirre, David Martínez, Oier López de Lacalle and Aitor Soroa 585

Broad-Coverage Sense Disambiguation and Information Extraction with a Supersense Sequence Tagger
Massimiliano Ciaramita and Yasemin Altun 594

Learning Field Compatibilities to Extract Database Records from Unstructured Text
Michael Wick, Aron Culotta and Andrew McCallum 603

Discriminative Methods for Transliteration
Dmitry Zelenko and Chinatsu Aone 612

Solving the Problem of Cascading Errors: Approximate Bayesian Inference for Linguistic Annotation Pipelines
Jenny Rose Finkel, Christopher D. Manning and Andrew Y. Ng 618

Author Index 627