Short Paper Table of Contents

Automating XML markup of text documents
Shazia Akhtar, Ronan G. Reilly and John Dunnion 1

Factored Language Models and Generalized Parallel Backoff
Jeff A. Bilmes and Katrin Kirchhoff 4

Getting More Mileage from Web Text Sources for Conversational Speech Language Modeling using Class-Dependent Mixtures
Ivan Bulyko, Mari Ostendorf and Andreas Stolcke 7

Exploiting Diversity for Answering Questions
John Burger and John Henderson 10

Story Link Detection and New Event Detection are Asymmetric
Francine Chen, Ayman Farahat and Thorsten Brants 13

Adaptation Using Out-of-Domain Corpus within EBMT
Takao Doi, Eiichiro Sumita and Hirofumi Yamamoto 16

Active Learning for Classifying Phone Sequences from Unsupervised Phonotactic Models
Shona Douglas 19

A Maximum Entropy Approach to FrameNet Tagging
Michael Fleischman and Eduard Hovy 22

Target Word Detection and Semantic Role Chunking using Support Vector Machines
Kadri Hacioglu and Wayne Ward 25

Question Classification with Support Vector Machines and Error Correcting Codes
Kadri Hacioglu and Wayne Ward 28

Rhetorical Parsing with Underspecification and Forests
Thomas Hanneforth, Silvan Heintze and Manfred Stede 31

Detection Of Agreement vs. Disagreement In Meetings: Training With Unlabeled Data
Dustin Hillard, Mari Ostendorf and Elizabeth Shriberg 34

Automatic Expansion of Equivalent Sentence Set Based on Syntactic Substitution
Kenji Imamura, Yasuhiro Akiba and Eiichiro Sumita 37

Identifying and Tracking Entity Mentions in a Maximum Entropy Framework
Abraham Ittycheriah, Lucian Lita, Nanda Kambhatla, Nicolas Nicolov, Salim Roukos and Margo Stys 40

Unsupervised Learning of Morphology for English and Inuktitut
Howard Johnson and Joel Martin 43

Cognates Can Improve Statistical Translation Models
Grzegorz Kondrak, Daniel Marcu and Kevin Knight 46

Word Alignment with Cohesion Constraint
Dekang Lin and Colin Cherry 49

Towards Emotion Prediction in Spoken Tutoring Dialogues
Diane Litman, Kate Forbes and Scott Silliman 52

Inferring Temporal Ordering of Events in News
Inderjeet Mani, Barry Schiffman and Jianping Zhang 55

A Robust Retrieval Engine for Proximal and Structural Search
Katsuya Masuda, Takashi Ninomiya, Yusuke Miyao, Tomoko Ohta and Jun'ichi Tsujii 58

Precision and Recall of Machine Translation
I. Dan Melamed, Ryan Green and Joseph P. Turian 61

Semantic Extraction with Wide-Coverage Lexical Resources
Behrang Mohit and Srini Narayanan 64

Category-based Pseudowords
Preslav I. Nakov and Marti A. Hearst 67

References to Named Entities: a Corpus Study
Ani Nenkova and Kathleen McKeown 70

Bootstrapping for Named Entity Tagging Using Concept-based Seeds
Cheng Niu, Wei Li, Jihong Ding and Rohini K. Srihari 73

Desparately Seeking Cebuano
Douglas W. Oard, David Doermann, Bonnie Dorr, Daqing He, Philip Resnik, Amy Weinberg, William Byrne, Sanjeev Khudanpur, David Yarowsky, Anton Leuski, Philipp Koehn and Kevin Knight 76

Bayesian Nets for Syntactic Categorization of Novel Words
Leonid Peshkin, Avi Pfeffer and Virginia Savova 79

LM Studies on Filled Pauses in Spontaneous Medical Dictation
Jochen Peters 82

Automatic Derivation of Surface Text Patterns for a Maximum Entropy Based Question Answering System
Deepak Ravichandran, Abraham Ittycheriah and Salim Roukos 85

A Hybrid Approach to Content Analysis for Automatic Essay Grading
Carolyn P. Rosé, Antonio Roque, Dumisizwe Bhembe and Kurt VanLehn 88

Auditory-based Acoustic Distinctive Features and Spectral Cues for Robust Automatic Speech Recognition in Low-SNR Car Environments
Sid-Ahmed Selouani, Hesham Tolba and Douglas O'Shaughnessy 91

Latent Semantic Analysis for Dialogue Act Classification
Riccardo Serafin, Barbara Di Eugenio and Michael Glass 94

Automatically Predicting Information Quality in News Documents
Rong Tang, Kwong Bor Ng, Tomek Strzalkowski and Paul B. Kantor 97

Building lexical semantic representations for Natural Language instructions
Elena Terenzi and Barbara Di Eugenio 100

A Context-Sensitive Homograph Disambiguation in Thai Text-to-Speech Synthesis
Virongrong Tesprasit, Paisarn Charoenpornsawat and Virach Sortlertlamvanich 103

A Phrase-based Unigram Model for Statistical Machine Translation
Christoph Tillmann and Fei Xia 106

Evaluating Answers to Definition Questions
Ellen M. Voorhees 109

Implicit Trajectory Modeling through Gaussian Transition Models for Speech Recognition
Hua Yu and Tanja Schultz 112