HLT-NAACL 2004

Human Language Technology
Conference of the
North American Chapter of the
Association for Computational Linguistics

Companion Volume:
Short Papers, Student Research
Workshop, Demonstrations,
Tutorials Abstracts

May 2-7, 2004
Boston, Massachusetts, USA

ISBN 1-932432-24-8

HLT-NAACL 2004

Human Language Technology
Conference of the
North American Chapter of the
Association for Computational Linguistics

Short Papers

Julia Hirschberg, General Chair

Susan Dumais, Daniel Marcu, and Salim Roukos
Program Co-chairs

May 2-7, 2004
Boston, Massachusetts, USA

HLT-NAACL 2004: Companion Volume, May 2–7, 2004

Short Papers Table of Contents

Using N-Grams To Understand the Nature of Summaries
Michele Banko and Lucy Vanderwende 1

MMR-based Feature Selection for Text Categorization
Changki Lee and Gary Geunbae Lee 5

Example-based Rescoring of Statistical Machine Translation Output
Michael Pau, Eiichiro Sumita and Seiichi Yamamoto 9

An Empirical Study on Multiple LVCSR Model Combination by Machine Learning
Takehito Utsuro, Yasuhiro Kodama, Tomohiro Watanabe, Hiromitsu Nishizaki and Seiichi Nakagawa 13

Enhancing Linguistically Oriented Automatic Keyword Extraction
Anette Hulth 17

Language Model Adaptation with MAP Estimation and the Perceptron Algorithm
Michiel Bacchiani, Brian Roark and Murat Saraclar 21

Advances in Children's Speech Recognition within an Interactive Literacy Tutor
Andreas Hagen, Bryan Pellom, Sarel van Vuuren and Ronald Cole 25

Automatic Construction of an English-Chinese Bilingual FrameNet
Benfeng Chen and Pascale Fung 29

Competitive Self-Trained Pronoun Interpretation
Andrew Kehler, Douglas Appelt, Lara Taylor and Aleksandr Simma 33

Using N-best lists for Named Entity Recognition from Chinese Speech
Lufeng Zhai, Pascale Fung, Richard Schwartz, Marine Carpuat and Dekai Wu 37

Performance Evaluation and Error Analysis for Multimodal Reference Resolution in a Conversation System
Joyce Y. Chai, Zahar Prasov and Pengyu Hong 41

UI on the Fly: Generating a Multimodal User Interface
David Reitter, Erin Panttaja and Fred Cummins 45

Web Search Intent Induction via Automatic Query Reformulation
Hal Daumé III and Eric Brill 49

HITIQA: A Data Driven Approach to Interactive Analytical Question Answering
Sharon Small and Tomek Strzalkowski 53

Morphological Analysis for Statistical Machine Translation
Young-Suk Lee 57

Correction Grammars for Error Handling in a Speech Dialog System
Hirohiko Sagawa, Teruko Mitamura and Eric Nyberg 61

A Comparison of Rule–Based and Statistical Methods for Semantic Language Modeling and Confidence Measurement
Ruhi Srikaya, Yuqing Gao and Michael Picheny 65

Improving Automatic Sentence Boundary Detection with Confusion Networks
D. Hillard, M. Ostendorf, A. Stolcke, Y. Liu and E. Shriberg 69

Speech Graffiti vs. Natural Language: Assessing the User Experience
Stefanie Tomko and Roni Rosenfeld 73

Augmenting the Kappa Statistic to Determine Interannotator Reliability for Multiply Labeled Data Points
Andrew Rosenberg and Ed Binkowski 77

Feature-based Pronunciation Modeling for Speech Recognition
Karen Livescu and James Glass 81

Context-based Speech Recognition Error Detection and Correction
Arup Sarma and David D. Palmer 85

Feature Selection for Trainable Multilingual Broadcast News Segmentation
David D. Palmer, Marc Reichman and Elyes Yaich 89

Direct Maximization of Average Precision by Hill-Climbing, with a Comparison to a Maximum Entropy Approach
William Morgan, Warren Greiff and John Henderson 93

Automated Team Discourse Annotation and Performance Prediction Using LSA
Melanie J. Martin and Peter W. Foltz 97

A Unigram Orientation Model for Statistical Machine Translation
Christoph Tillman 101

Summarizing Email Threads
Owen Rambow, Lokesh Shrestha, John Chen and Christy Laurdisen 105

Confidence Estimation for Information Extraction
Aron Culotta and Andrew McCallum 109

A Speech-in List-out Approach to Spoken User Interfaces
Vijay Divi, C. Forlines, Jan Van Gemert, Bhiksha Raj, B. Schmidt-Nielsen, Kent Wittenburg, Joseph Woelfel and Fang-Fang Zhang 113

Nearly-Automated Metadata Hierarchy Creation
Emilia Stoica and Marti A. Hearst 117

Computational Linkuistics: Word Triggers across Hyperlinks
Dragomir Radev, Hong Qi, Adam Winkel and Daniel Tam 121

Parsing Conversational Speech Using Enhanced Segmentation
Jeremy G. Kahn, Mari Ostendorf and Ciprian Chelba 125

Polarity Sensitivity and Evaluation Order in Type-Logical Grammar
Chung-chieh Shan 129

Multi-Speaker Language Modeling
Gang Ji and Jeffrey Bilmes 133

Prosody-based Topic Segmentation for Mandarin Broadcast News
Gina-Anne Levow 137

Parsing Arguments of Nominalizations in English and Chinese
Sameer Pradhan, Honglin Sun, Wayne Ward, James H. Martin and Daniel Jurafsky . 141

A Lightweight Semantic Chunker Based on Tagging
Kadri Hacioglu 145

Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks
Mona Diab, Kadri Hacioglu and Daniel Jurafsky 149

Converting Text into Agent Animations: Assigning Gestures to Text
Yukiko I. Nakano, Masashi Okamoto, Daisuke Kawahara, Qing Li
and Toyoaki Nishida 153

A Lexically-Driven Algorithm for Disfluency Detection
Matthew Snover, Bonnie Dorr and Richard Schwartz 157

Short Papers Author Index

Appelt, Douglas 33
Bacchiani, Michiel 21
Banko, Michele 1
Bilmes, Jeffrey 133
Binkowski, Ed 77
Brill, Eric 49
Carpuat, Marine 37
Chai, Joyce Y. 41
Chelba, Ciprian 125
Chen, Benfeng 29
Chen, John 105
Cole, Ronald 25
Culotta, Aron 109
Cummins, Fred 45
Daumé III, Hal 49
Diab, Mona 149
Divi, Vijay 113
Dorr, Bonnie 157
Foltz, Peter W. 97
Forlines, C. 113
Fung, Pascale 29, 37
Gao, Yuqing 65
Glass, James 81
Greiff, Warren 93
Hacioglu, Kadri 145, 149
Hagen, Andreas 25
Hearst, Marti A. 117
Henderson, John 93
Hillard, D. 69
Hong, Pengyu 41
Hulth, Anette 17
Ji, Gang 133
Jurafsky, Daniel 141, 149
Kahn, Jeremy G. 125
Kawahara, Daisuke 153
Kehler, Andrew 33
Kodama, Yasuhiro 13
Laurdisen, Christy 105
Lee, Changki 5
Lee, Gary Geunbae 5
Lee, Young-Suk 57
Levow, Gina-Anne 137
Li, Qing 153
Liu, Y. 69
Livescu, Karen 81
Martin, James H. 141
Martin, Melanie J. 97

McCallum, Andrew . . . 109
Mitamura, Teruko . . . 61
Morgan, William . . . 93
Nakagawa, Seiichi . . . 13
Nakano, Yukiko I. . . . 153
Nishida, Toyoaki . . . 153
Nishizaki, Hiromitsu . . . 13
Nyberg, Eric . . . 61
Okamoto, Masashi . . . 153
Ostendorf, M. . . . 69
Ostendorf, Mari . . . 125
Palmer, David D. . . . 85, 89
Panttaja, Erin . . . 45
Pau, Michael . . . 9
Pellom, Bryan . . . 25
Picheny, Michael . . . 65
Pradhan, Sameer . . . 141
Prasov, Zahar . . . 41
Qi, Hong . . . 121
Radev, Dragomir . . . 121
Raj, Bhiksha . . . 113
Rambow, Owen . . . 105
Reichman, Marc . . . 89
Reitter, David . . . 45
Roark, Brian . . . 21
Rosenberg, Andrew . . . 77
Rosenfeld, Roni . . . 73
Sagawa, Hirohiko . . . 61
Saraclar, Murat . . . 21
Sarma, Arup . . . 85
Schmidt-Nielsen, B. . . . 113
Schwartz, Richard . . . 37, 157
Shan, Chung-chieh . . . 129
Shrestha, Lokesh . . . 105
Shriberg, E. . . . 69
Simma, Aleksandr . . . 33
Small, Sharon . . . 53
Snover, Matthew . . . 157
Srikaya, Ruhi . . . 65
Stoica, Emilia . . . 117
Stolcke, A. . . . 69
Strzalkowski, Tomek . . . 53
Sumita, Eiichiro . . . 9
Sun, Honglin . . . 141
Tam, Daniel . . . 121
Taylor, Lara . . . 33
Tillman, Christoph . . . 101
Tomko, Stefanie . . . 73
Utsuro, Takehito . . . 13

Van Gemert, Jan 113
van Vuuren, Sarel 25
Vanderwende, Lucy 1
Ward, Wayne 141
Watanabe, Tomohiro 13
Winkel, Adam 121
Wittenburg, Kent 113
Woelfel, Joseph 113
Wu, Dekai 37
Yaich, Elyes 89
Yamamoto, Seiichi 9
Zhai, Lufeng 37
Zhang, Fang-Fang 113