NAACL HLT 2007

Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics

Companion Volume: Short Papers

Candace Sidner, General Chair
Tanja Schultz, Matthew Stone, and ChengXiang Zhai
Program Committee Chairs

22–27 April 2007
Rochester, New York, USA
Table of Contents

Comparing User Simulation Models For Dialog Strategy Learning  
Hua Ai, Joel Tetreault and Diane Litman ........................................... 1

Automatic Acquisition of Grammatical Types for Nouns  
Nuria Bel, Sergio Espeja and Montserrat Marimon .............................. 5

Conquest—An Open-Source Dialog System for Conferences  
Dan Bohus, Sergio Grau Puerto, David Huggins-Daines, Venkatesh Keri, Gopala Krishna, Rohit Kumar, Antoine Raux and Stefanie Tomko ........................................... 9

Joint Versus Independent Phonological Feature Models within CRF Phone Recognition  
Ilana Bromberg, Jeremy Morris and Eric Fosler-Lussier .......................... 13

K-Best Suffix Arrays  
Kenneth Church, Bo Thiesson and Robert Ragno ................................. 17

Translation Model Pruning via Usage Statistics for Statistical Machine Translation  
Matthias Eck, Stephan Vogel and Alex Waibel ..................................... 21

Combination of Statistical Word Alignments Based on Multiple Preprocessing Schemes  
Jakob Elming and Nizar Habash ...................................................... 25

A Fast Method for Parallel Document Identification  
Jessica Enright and Grzegorz Kondrak ............................................ 29

Generalized Graphical Abstractions for Statistical Machine Translation  
Karim Filali and Jeff Bilmes ......................................................... 33

Situated Models of Meaning for Sports Video Retrieval  
Michael Fleischman and Deb Roy ................................................... 37

Exploring Affect-Context Dependencies for Adaptive System Development  
Kate Forbes-Riley, Mihai Rotaru, Diane Litman and Joel Tetreault ............ 41

A Filter-Based Approach to Detect End-of-Utterances from Prosody in Dialog Systems  
Olac Fuentes, David Vera and Thamar Solorio .................................... 45

Document Similarity Measures to Distinguish Native vs. Non-Native Essay Writers  
Olga Gurevich and Paul Deane ....................................................... 49

Arabic Diacritization through Full Morphological Tagging  
Nizar Habash and Owen Rambow .................................................... 53

Are Very Large N-Best Lists Useful for SMT?  
Saša Hasan, Richard Zens and Hermann Ney ...................................... 57
Relationship between Non-Projective Edges, Their Level Types, and Well-Nestedness
  Jiří Havelka ................................................................. 61

iROVER: Improving System Combination with Classification
  Dustin Hillard, Bjoern Hoffmeister, Mari Ostendorf, Ralf Schlueter and Hermann Ney ...... 65

Clustered Sub-Matrix Singular Value Decomposition
  Fang Huang and Yorick Wilks ........................................... 69

Implicitly Supervised Language Model Adaptation for Meeting Transcription
  David Huggins-Daines and Alexander I. Rudnicky ............................ 73

ILR-Based MT Comprehension Test with Multi-Level Questions
  Douglas Jones, Martha Herzog, Hussny Ibrahim, Arvind Jairam, Wade Shen, Edward Gibson and Michael Emonts .......................................................... 77

Semi-Supervised Learning for Semantic Parsing using Support Vector Machines
  Rohit Kate and Raymond Mooney ........................................... 81

Discriminative Alignment Training without Annotated Data for Machine Translation
  Patrik Lambert, Rafael E. Banchs and Josep M. Crego ........................ 85

A Geometric Interpretation of Non-Target-Normalized Maximum Cross-Channel Correlation for Vocal Activity Detection in Meetings
  Kornel Laskowski and Tanja Schultz ...................................... 89

Detection of Non-Native Sentences Using Machine-Translated Training Data
  John Lee, Ming Zhou and Xiaohua Liu ..................................... 93

Exploiting Rich Syntactic Information for Relationship Extraction from Biomedical Articles
  Yudong Liu, Zhongmin Shi and Anoop Sarkar ............................... 97

Look Who is Talking: Soundbite Speaker Name Recognition in Broadcast News Speech
  Feifan Liu and Yang Liu .................................................... 101

Tagging Icelandic Text using a Linguistic and a Statistical Tagger
  Hrafn Loftsson ............................................................... 105

Efficient Computation of Entropy Gradient for Semi-Supervised Conditional Random Fields
  Gideon Mann and Andrew McCallum ....................................... 109

Hybrid Document Indexing with Spectral Embedding
  Irina Matveeva and Gina-Anne Levow ..................................... 113

On using Articulatory Features for Discriminative Speaker Adaptation
  Florian Metze ............................................................... 117

RH: A Retro-Hybrid Parser
  Paula Newman .............................................................. 121
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtree Mining for Relation Extraction from Wikipedia</td>
<td>Dat P.T. Nguyen, Yutaka Matsuo and Mitsuru Ishizuka</td>
<td>125</td>
</tr>
<tr>
<td>Advances in the CMU/Interact Arabic GALE Transcription System</td>
<td>Mohamed Noamany, Thomas Schaaf and Tanja Schultz</td>
<td>129</td>
</tr>
<tr>
<td>Analysis and System Combination of Phrase- and N-Gram-Based Statistical Machine Translation Systems</td>
<td>Marta R. Costa-jussà, Josep M. Crego, David Vilar, José A. R. Fonollosa, José B. Mariño and Hermann Ney</td>
<td>137</td>
</tr>
<tr>
<td>Stating with Certainty or Stating with Doubt: Intercoder Reliability Results for Manual Annotation of Epistemically Modalized Statements</td>
<td>Victoria L. Rubin</td>
<td>141</td>
</tr>
<tr>
<td>Joint Morphological-Lexical Language Modeling for Machine Translation</td>
<td>Ruhi Sarikaya and Yonggang Deng</td>
<td>145</td>
</tr>
<tr>
<td>Agenda-Based User Simulation for Bootstrapping a POMDP Dialogue System</td>
<td>Jost Schatzmann, Blaise Thomson, Karl Weilhammer, Hui Ye and Steve Young</td>
<td>149</td>
</tr>
<tr>
<td>Reversible Sound-to-Letter/Letter-to-Sound Modeling Based on Syllable Structure</td>
<td>Stephanie Seneff</td>
<td>153</td>
</tr>
<tr>
<td>Are Some Speech Recognition Errors Easier to Detect than Others?</td>
<td>Yongmei Shi and Lina Zhou</td>
<td>157</td>
</tr>
<tr>
<td>Simultaneous Identification of Biomedical Named-Entity and Functional Relation Using Statistical Parsing Techniques</td>
<td>Zhongmin Shi, Anoop Sarkar and Fred Popowich</td>
<td>161</td>
</tr>
<tr>
<td>Virtual Evidence for Training Speech Recognizers Using Partially Labeled Data</td>
<td>Amarnag Subramanya and Jeff Bilmes</td>
<td>165</td>
</tr>
<tr>
<td>A High Accuracy Method for Semi-Supervised Information Extraction</td>
<td>Stephen Tratz and Antonio Sanfilippo</td>
<td>169</td>
</tr>
<tr>
<td>The Effects of Word Prediction on Communication Rate for AAC</td>
<td>Keith Trnka, Debra Yarrington, John McCaw, Kathleen F. McCoy and Christopher Pennington</td>
<td>173</td>
</tr>
<tr>
<td>Language Modeling for Determiner Selection</td>
<td>Jenine Turner and Eugene Charniak</td>
<td>177</td>
</tr>
<tr>
<td>Entity Extraction is a Boring Solved Problem—Or is it?</td>
<td>Marc Vilain, Jennifer Su and Suzi Lubar</td>
<td>181</td>
</tr>
</tbody>
</table>
Kernel Regression Based Machine Translation
Zhuoran Wang, John Shawe-Taylor and Sándor Szedmák ............................................................. 185

Modifying SO-PMI for Japanese Weblog Opinion Mining by Using a Balancing Factor and Detecting Neutral Expressions
Guangwei Wang and Kenji Araki ................................................................. 189

Combined Use of Speaker- and Tone-Normalized Pitch Reset with Pause Duration for Automatic Story Segmentation in Mandarin Broadcast News
Lei Xie, Chuan Liu and Helen Meng ............................................................. 193

Chinese Named Entity Recognition with Cascaded Hybrid Model
Xiaofeng Yu .............................................................. 197

A Three-Step Deterministic Parser for Chinese Dependency Parsing
Kun Yu, Sadao Kurohashi and Hao Liu ..................................................... 201

Comparing Wikipedia and German WordNet by Evaluating Semantic Relatedness on Multiple Datasets
Torsten Zesch, Iryna Gurevych and Max Mühlhäusser ........................................... 205

Selective Phrase Pair Extraction for Improved Statistical Machine Translation
Luke Zettlemoyer and Robert Moore ......................................................... 209

Speech Summarization Without Lexical Features for Mandarin Broadcast News
Jian Zhang and Pascale Fung ................................................................. 213

A Semi-Automatic Evaluation Scheme: Automated Nuggetization for Manual Annotation
Liang Zhou, Namhee Kwon and Eduard Hovy ................................................. 217