Conference Program

Wednesday, October 5, 2005
6:00–8:00 pm Opening Reception (partially sponsored by Google)

Thursday, October 6, 2005
8:45–9:00 Opening Session
9:00–10:00 Invited Talk: Larry Hunter
10:00–10:30 Break

Session 1A: Anaphora and Information Structure
10:30–10:55 Improving LSA-based Summarization with Anaphora Resolution
Josef Steinberger, Mijail Kabadjov, Massimo Poesio and Olivia Sanchez-Graillet
10:55–11:20 Data-driven Approaches for Information Structure Identification
Oana Postolache, Ivana Kruijff-Korbayova and Geert-Jan Kruijff
11:20–11:45 Using Semantic Relations to Refine Coreference Decisions
Heng Ji, David Westbrook and Ralph Grishman
11:45–12:10 On Coreference Resolution Performance Metrics
Xiaqiang Luo

Session 1B: Error Detection and Correction
10:30–10:55 Improving Multilingual Summarization: Using Redundancy in the Input to Correct MT errors
Advaith Siddharthan and Kathleen McKeown
10:55–11:20 Error Detection Using Linguistic Features
Yongmei Shi and Lina Zhou
11:20–11:45 Semantic Similarity for Detecting Recognition Errors in Automatic Speech Transcripts
Diana Inkpen and Alain Désilets
11:45–12:10 Redundancy-based Correction of Automatically Extracted Facts
Roman Yangarber and Lauri Jokipii
Thursday, October 6, 2005 (continued)

Session 1C: Word Alignment

10:30–10:55  NeurAlign: Combining Word Alignments Using Neural Networks
             Necip Fazil Ayan, Bonnie J. Dorr and Christof Monz

10:55–11:20  A Discriminative Matching Approach to Word Alignment
             Ben Taskar, Lacoste-Julien Simon and Klein Dan

11:20–11:45  A Discriminative Framework for Bilingual Word Alignment
             Robert C. Moore

             Abraham Ittycheriah and Salim Roukos

12:10–2:00  Lunch

Session 2A: Topic Tracking and Entity Detection

2:00–2:25  A Large-Scale Exploration of Effective Global Features for a Joint Entity Detection and Tracking Model
           Hal Daumé III and Daniel Marcu

2:25–2:50  Novelty Detection: The TREC Experience
           Ian Soboroff and Donna Harman

           Elena Filatova and John Prager

3:15–3:40  Using Names and Topics for New Event Detection
           Giridhar Kumaran and James Allan
Thursday, October 6, 2005 (continued)

Session 2B: Text Categorization and Learning

2:00–2:25  Investigating Unsupervised Learning for Text Categorization Bootstrapping  
          Alfio Gliozzo, Carlo Strapparava and Ido Dagan

2:25–2:50  Speeding up Training with Tree Kernels for Node Relation Labeling  
          Jun’ichi Kazama and Kentaro Torisawa

2:50–3:15  Kernel-based Approach for Automatic Evaluation of Natural Language Generation Technologies:  
          Application to Automatic Summarization  
          Tsutomu Hirao, Manabu Okumura and Hideki Isozaki

3:15–3:40  Discretization Based Learning for Information Retrieval  
          Dmitri Roussinov and Weiguo Fan

Session 2C: Machine Translation

2:00–2:25  Local Phrase Reordering Models for Statistical Machine Translation  
          Shankar Kumar and William Byrne

2:25–2:50  HMM Word and Phrase Alignment for Statistical Machine Translation  
          Yonggang Deng and William Byrne

2:50–3:15  Inner-Outer Bracket Models for Word Alignment using Hidden Blocks  
          Bing Zhao, Niyu Ge and Kishore Papineni

3:15–3:40  Alignment Link Projection Using Transformation-Based Learning  
          Necip Fazil Ayan, Bonnie J. Dorr and Christof Monz

3:40-4:15  Break
Thursday, October 6, 2005 (continued)

Session 3A: Language Modeling

4:15–4:40  Predicting Sentences using N-Gram Language Models
Steffen Bickel, Peter Haider and Tobias Scheffer

4:40–5:05  Training Neural Network Language Models on Very Large Corpora
Holger Schwenk and Jean-Luc Gauvain

5:05–5:30  Minimum Sample Risk Methods for Language Modeling
Jianfeng Gao, Hao Yu, Wei Yuan and Peng Xu

Session 3B: Robust Dialogue Processing

4:15–4:40  A Salience Driven Approach to Robust Input Interpretation in Multimodal Conversational Systems
Joyce Y. Chai and Shaolin Qu

4:40–5:05  Error Handling in the RavenClaw Dialog Management Architecture
Dan Bohus and Alexander Rudnicky

5:05–5:30  Effective Use of Prosody in Parsing Conversational Speech
Jeremy G. Kahn, Matthew Lease, Eugene Charniak, Mark Johnson and Mari Ostendorf

Session 3C: Summarization

4:15–4:40  Automatically Learning Cognitive Status for Multi-Document Summarization of Newswire
Ani Nenkova, Advaith Siddharthan and Kathleen McKeown

4:40–5:05  Bayesian Learning in Text Summarization
Tadashi Nomoto

5:05–5:30  Discourse Chunking and its Application to Sentence Compression
Caroline Sporleder and Mirella Lapata
Friday, October 7, 2005

9:00–10:00 Invited Talk: Sanjeev Khudanpur

10:00-10:30 Break

Session 4A: Training and Adaptation

Hisami Suzuki and Jianfeng Gao

10:55–11:20 PP-attachment Disambiguation using Large Context
Marian Olteanu and Dan Moldovan

11:20–11:45 Compiling Comp Ling: Weighted Dynamic Programming and the Dyna Language
Jason Eisner, Eric Goldlust and Noah A. Smith

11:45–12:10 Learning What to Talk about in Descriptive Games
Hugo Zaragoza and Chi-Ho Li

Session 4B: Question Answering

10:30–10:55 Using Question Series to Evaluate Question Answering System Effectiveness
Ellen Voorhees

Fuchun Peng, Ralph Weischedel, Ana Licuanan and Jinxi Xu

11:20–11:45 Enhanced Answer Type Inference from Questions using Sequential Models
Vijay Krishnan, Sujatha Das and Soumen Chakrabarti

11:45–12:10 A Practically Unsupervised Learning Method to Identify Single-Snippet Answers to Definition Questions on the Web
Ion Androutsopoulos and Dimitrios Galanis
Friday, October 7, 2005 (continued)

Session 4C: Opinion and Sentiment Analysis

10:30–10:55  Collective Content Selection for Concept-to-Text Generation
Regina Barzilay and Mirella Lapata

10:55–11:20  Extracting Product Features and Opinions from Reviews
Ana-Maria Popescu and Oren Etzioni

11:20–11:45  Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis
Theresa Wilson, Janyce Wiebe and Paul Hoffmann

11:45–12:10  Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns
Yejin Choi, Claire Cardie, Ellen Riloff and Siddharth Patwardhan

12:10–2:00  Lunch

Session 5A: Textual Entailment and Inference

2:00–2:25  Disambiguating Toponyms in News
Eric Garbin and Inderjeet Mani

2:25–2:50  A Semantic Approach to Recognizing Textual Entailment
Marta Tatu and Dan Moldovan

2:50–3:15  Detection of Entity Mentions Occuring in English and Chinese Text
Kadri Hacioglu, Benjamin Douglas and Ying Chen

3:15–3:40  Robust Textual Inference via Graph Matching
Aria Haghighi, Andrew Ng and Christopher Manning
Session 5B: Word Sense and Polysemy

2:00–2:25 Bootstrapping Without the Boot
Jason Eisner and Damianos Karakos

2:25–2:50 Differentiating Homonymy and Polysemy in Information Retrieval
Christopher Stokoe

2:50–3:15 Unsupervised Large-Vocabulary Word Sense Disambiguation with Graph-based Algorithms for Sequence Data Labeling
Rada Mihalcea

3:15–3:40 Domain-Specific Sense Distributions and Predominant Sense Acquisition
Rob Koeling, Diana McCarthy and John Carroll

Session 5C: Named Entity Recognition and Transliteration

2:00–2:25 Chinese Named Entity Recognition with Multiple Features
Youzheng Wu, Jun Zhao, Bo Xu and Hao Yu

2:25–2:50 Cluster-specific Named Entity Transliteration
Fei Huang

2:50–3:15 Extracting Personal Names from Email: Applying Named Entity Recognition to Informal Text
Einat Minkov, Richard C. Wang and William W. Cohen

3:15–3:40 Matching Inconsistently Spelled Names in Automatic Speech Recognizer Output for Information Retrieval
Hema Raghavan and James Allan

3:40–4:15 Break
Session 6A: Sequence Models

4:15–4:40  *Part-of-Speech Tagging using Virtual Evidence and Negative Training*
Sheila M. Reynolds and Jeff A. Bilmes

4:40–5:05  *Bidirectional Inference with the Easiest-First Strategy for Tagging Sequence Data*
Yoshimasa Tsuruoka and Jun'ichi Tsujii

5:05–5:30  *Context-Based Morphological Disambiguation with Random Fields*
Noah A. Smith, David A. Smith and Roy W. Tromble

Session 6B: Large-Scale Systems

4:15–4:40  *Mining Key Phrase Translations from Web Corpora*
Fei Huang, Ying Zhang and Stephan Vogel

4:40–5:05  *Robust Named Entity Extraction from Large Spoken Archives*
Benoit Favre, Frédéric Bechet and Pascal Nocéra

5:05–5:30  *Mining Context Specific Similarity Relationships Using The World Wide Web*
Dmitri Roussinov, Leon J. Zhao and Weiguo Fan

Session 6C: Statistical Parsing

4:15–4:40  *Hidden-Variable Models for Discriminative Reranking*
Terry Koo and Michael Collins

4:40–5:05  *Disambiguation of Morphological Structure using a PCFG*
Helmut Schmid

5:05–5:30  *Non-Projective Dependency Parsing using Spanning Tree Algorithms*
Ryan McDonald, Fernando Pereira, Kiril Ribarov and Jan Hajic
Demo and Poster Reception (6:00-9:00)

Making Computers Laugh: Investigations in Automatic Humor Recognition
Rada Mihalcea and Carlo Strapparava

Optimizing to Arbitrary NLP Metrics using Ensemble Selection
Art Munson, Claire Cardie and Rich Caruana

Word Sense Disambiguation Using Sense Examples Automatically Acquired from a Second Language
Xinglong Wang and John Carroll

Using MONA for Querying Linguistic Treebanks
Stephan Kepser

KnowItNow: Fast, Scalable Information Extraction from the Web
Michael J. Cafarella, Doug Downey, Stephen Soderland and Oren Etzioni

A Cost-Benefit Analysis of Hybrid Phone-Manner Representations for ASR
Eric Fosler-Lussier and C. Anton Rytting

Emotions from Text: Machine Learning for Text-based Emotion Prediction
Cecilia Ovesdotter Alm, Dan Roth and Richard Sproat

Combining Multiple Forms of Evidence While Filtering
Yi Zhang and Jamie Callan

Handling Biographical Questions with Implicature
Donghui Feng and Eduard Hovy

The Use of Metadata, Web-derived Answer Patterns and Passage Context to Improve Reading Comprehension Performance
Yongping Du, Helen Meng, Xuanjing Huang and Lide Wu

Identifying Semantic Relations and Functional Properties of Human Verb Associations
Sabine Schulte im Walde and Alissa Melinger

Accurate Function Parsing
Paola Merlo and Gabriele Musillo
Recognising Textual Entailment with Logical Inference
Johan Bos and Katja Markert

A Self-Learning Context-Aware Lemmatizer for German
Praharsana Perera and René Witte

A Robust Combination Strategy for Semantic Role Labeling
Lluís Márquez, Mihai Surdeanu, Pere Comas and Jordi Turmo

A Methodology for Extrinsic Evaluation of Information Extraction Performance
Michael Crystal, Alex Baron, Katherine Godfrey, Linnea Micciulla, Yvette Tenney and Ralph Weischedel

Multi-Lingual Coreference Resolution With Syntactic Features
Xiaoqiang Luo and Imed Zitouni

Analyzing Models for Semantic Role Assignment using Confusability
Katrin Erk and Sebastian Pado

Improving Statistical MT through Morphological Analysis
Sharon Goldwater and David McClosky

A Translation Model for Sentence Retrieval
Vanessa Murdock and W. Bruce Croft

Maximum Expected F-Measure Training of Logistic Regression Models
Martin Jansche

Evita: A Robust Event Recognizer For QA Systems
Roser Saurí, Robert Knippen, Marc Verhagen and James Pustejovsky

Using Sketches to Estimate Associations
Ping Li and Kenneth W. Church

Context and Learning in Novelty Detection
Barry Schiffman and Kathleen McKeown
Friday, October 7, 2005 (continued)

A Shortest Path Dependency Kernel for Relation Extraction
Razvan Bunescu and Raymond Mooney

Multi-way Relation Classification: Application to Protein-Protein Interactions
Barbara Rosario and Marti Hearst

BLANC: Learning Evaluation Metrics for MT
Lucian Lita, Monica Rogati and Alon Lavie

Composition of Conditional Random Fields for Transfer Learning
Charles Sutton and Andrew McCallum

Saturday, October 8, 2005

9:00–10:00 Invited Talk: Ellen Voorhees

10:00-10:30 Break

Session 7A: Machine Translation

10:30–10:55 Translating with Non-contiguous Phrases
Michel Simard, Nicola Cancedda, Bruno Cavestro, Marc Dymetman, Eric Gaussier, Cyril Goutte, Kenji Yamada, Philippe Langlais and Arne Mauser

10:55–11:20 Word-Level Confidence Estimation for Machine Translation using Phrase-Based Translation Models
Nicola Ueffing and Hermann Ney

11:20–11:45 Word-Sense Disambiguation for Machine Translation
David Vickrey, Luke Biewald, Marc Teyssier and Daphne Koller

11:45–12:10 The Hiero Machine Translation System: Extensions, Evaluation, and Analysis
David Chiang, Adam Lopez, Nitin Madnani, Christof Monz, Philip Resnik and Michael Subotin
Saturday, October 8, 2005 (continued)

Session 7B: Grammar and Parsing

10:30–10:55  Comparing and Combining Finite-State and Context-Free Parsers
             Kristy Hollingshead, Seeger Fisher and Brian Roark

10:55–11:20  Morphology and Reranking for the Statistical Parsing of Spanish
             Brooke Cowan and Michael Collins

11:20–11:45  Some Computational Complexity Results for Synchronous Context-Free Grammars
             Giorgio Satta and Enoch Peserico

11:45–12:10  Incremental LTAG Parsing
             Libin Shen and Aravind Joshi

Session 7C: Computational Psycholinguistics and Biomedical Language Processing

10:30–10:55  Automatic Question Generation for Vocabulary Assessment
             Jonathan Brown, Gwen Frishkoff and Maxine Eskenazi

10:55–11:20  Parallelism in Coordination as an Instance of Syntactic Priming: Evidence from Corpus-based Modeling
             Amit Dubey, Patrick Sturt and Frank Keller

11:20–11:45  Using the Web as an Implicit Training Set: Application to Structural Ambiguity Resolution
             Preslav Nakov and Marti Hearst

11:45–12:10  Paradigmatic Modifiability Statistics for the Extraction of Complex Multi-Word Terms
             Joachim Wermter and Udo Hahn

12:10–2:00  Lunch
Saturday, October 8, 2005 (continued)

Session 8A: Multilingual Approaches

2:00–2:25  *A Backoff Model for Bootstrapping Resources for Non-English Languages*  
Chenhai Xi and Rebecca Hwa

2:25–2:50  *Cross-linguistic Projection of Role-Semantic Information*  
Sebastian Pado and Mirella Lapata

2:50–3:15  *OCR Post-Processing for Low Density Languages*  
Okan Kolak and Philip Resnik

3:15–3:40  *Inducing a Multilingual Dictionary from a Parallel Multitext in Related Languages*  
Dmitriy Genzel

Session 8B: Semantic Roles and Relations

2:00–2:25  *Exploiting a Verb Lexicon in Automatic Semantic Role Labelling*  
Robert Swier and Suzanne Stevenson

2:25–2:50  *A Semantic Scattering Model for the Automatic Interpretation of Genitives*  
Dan Moldovan and Adriana Badulescu

2:50–3:15  *Measuring the Relative Compositionality of Verb-Noun (V-N) Collocations by Integrating Features*  
Sriram Venkatapathy and Aravind Joshi

3:15–3:40  *A Semi-Supervised Feature Clustering Algorithm with Application to Word Sense Disambiguation*  
Zheng-Yu Niu, Dong-Hong Ji and Chew Lim Tan
Saturday, October 8, 2005 (continued)

**Session 8C: Question Answering**

2:00–2:25  *Using Random Walks for Question-focused Sentence Retrieval*  
Jahna Otterbacher, Gunes Erkan and Dragomir Radev

2:25–2:50  *Multi-Perspective Question Answering Using the OpQA Corpus*  
Veselin Stoyanov, Claire Cardie and Janyce Wiebe

2:50–3:15  *Automatically Evaluating Answers to Definition Questions*  
Jimmy Lin and Dina Demner-Fushman

3:15–3:40  *Integrating Linguistic Knowledge in Passage Retrieval for Question Answering*  
Jörg Tiedemann

3:40–4:15  Break

**Session 9A: Search**

4:15–4:40  *Searching the Audio Notebook: Keyword Search in Recorded Conversation*  
Peng Yu, Kaijiang Chen, Lie Lu and Frank Seide

4:40–5:05  *Learning a Spelling Error Model from Search Query Logs*  
Farooq Ahmad and Grzegorz Kondrak

5:05–5:30  *Query Expansion with the Minimum User Feedback by Transductive Learning*  
Masayuki Okabe, Kyoji Umemura and Seiji Yamada
Saturday, October 8, 2005 (continued)

**Session 9B: Text Segmentation**

4:15–4:40  *An Orthonormal Basis for Topic Segmentation in Tutorial Dialogue*  
Andrew Olney and Zhiqiang Cai

4:40–5:05  *A Generalized Framework for Revealing Analogous Themes across Related Topics*  
Zvika Marx, Ido Dagan and Eli Shamir

5:05–5:30  *Flexible Text Segmentation with Structured Multilabel Classification*  
Ryan McDonald, Koby Crammer and Fernando Pereira

**Session 9C: Dialog Systems**

4:15–4:40  *The Vocal Joystick: A Voice-Based Human-Computer Interface for Individuals with Motor Impairments*  
Jeff A. Bilmes, Xiao Li, Jonathan Malkin, Kelley Kilanski, Richard Wright, Katrin Kirchhoff,  
Amar Subramanya, Susumu Harada, James Landay, Patricia Dowden and Howard Chizeck

4:40–5:05  *Speech-based Information Retrieval System with Clarification Dialogue Strategy*  
Misu Teruhisa and Kawahara Tatsuya

5:05–5:30  *Learning Mixed Initiative Dialog Strategies By Using Reinforcement Learning On Both Conversants*  
Michael English and Peter Heeman