

HLT-NAACL 2004

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

Proceedings of the Student Research
Workshop

Ani Nenkova, Nicola Stokes, and Karen Livescu
Student Workshop Co-chairs

Lisa Ballesteros, Eric Fosler-Lussier, and Amanda Stent
Faculty Advisors

May 2-7, 2004
Boston, Massachusetts, USA

PREFACE

We are pleased to present the 10 papers accepted for presentation at the HLT/NAACL 2004 Student Workshop to be held in Boston on May 2. The main mission of the workshop is to provide feedback for students' work in progress. Each presentation will be followed by comments on the work from a panel of senior researchers. We thank the panelists in advance for their time and feedback.

This year, we received 12 submissions representing 4 countries. We would like to thank all of the authors who submitted their work. We are very grateful to the program committee for the time and effort they put into reviewing the papers, and for their thoughtful feedback to the authors. We also thank the HLT/NAACL-2004 main conference organizers, especially Christy Doran, Miles Osborne, Katrin Kirchhoff, Pablo Duboue, Priscilla Rasmussen, and Julia Hirschberg. Several previous student co-chairs provided helpful words of wisdom, including Costas Boulis, Eric Breck, Victor Lavrenko, and Gideon Mann. We are extremely grateful to our faculty advisors, Lisa Ballesteros, Eric Fosler-Lussier, and Amanda Stent, for their generous guidance throughout the organizing process. Finally, we gratefully acknowledge the National Science Foundation, whose support will make it possible to provide financial assistance to the presenters.

Karen Livescu, Ani Nenkova, and Nicola Stokes

ORGANIZERS:

Co-Chairs:

Karen Livescu, MIT
Ani Nenkova, Columbia University
Nicola Stokes, University College Dublin

Faculty Advisors:

Lisa Ballesteros, Mount Holyoke College
Eric Fosler-Lussier, Ohio State University
Amanda Stent, Stony Brook University

PROGRAM COMMITTEE:

(S) Mridul Balaraman, Rutgers U.	Jerome Bellegarda, Apple
(S) John Blitzer, U. Pennsylvania	(S) Nishant Chandra, Mississippi State U.
Stanley Chen, IBM T.J. Watson	Michael Collins, MIT
Sorin Dusan, Rutgers U.	(S) Hal Daume III, USC/ISI
Gregory Grefenstette, Clairvoyance Corp.	Mary Harper, Purdue U.
Mark Hasegawa-Johnson, U. Illinois	Timothy J. Hazen, MIT
(S) Svetlana Hensman, University College Dublin	Djoerd Hiemstra, U. Twente
Gareth Jones, Dublin City U.	Vlado Keselj, Dalhousie U.
(S) Shankar Kumar, Johns Hopkins U.	Guy Lapalme, Université de Montréal
(S) Xiaolong Li, U. Missouri	(S) Xiaoyong Liu, U. Massachusetts
(S) Juha Makkonen, U. Helsinki	(S) Sameer Maskey, Columbia U.
Erik McDermott, NTT Corp.	(S) Preslav Nakov, UC Berkeley
(S) Paul Ogilvie, Carnegie Mellon U.	Mari Ostendorf, U. Washington
(S) Alex Park, MIT	Massimo Poesio, U. Essex
Owen Rambow, Columbia U.	(S) Deepak Ravichandran USC/ISI
Carolyn Rose, Carnegie Mellon U.	Tomek Strzalkowski, SUNY Albany
Jana Sukkarieh, U. Oxford	(S) Kristina Toutanova, Stanford U.
(S) Vivian Tsang, U. Toronto	Geoffrey Zweig, IBM T.J. Watson
(S) Student reviewer	

WORKSHOP WEBSITE:

<http://inismor.ucd.ie/~nstokes/hlt-naacl04-student/>

Student Workshop Table of Contents

<i>Multilingual Speech Recognition for Information Retrieval in Indian Context</i> N. Udhaykumar, S.K. Ramakrishnan and R. Swaminathan	1
<i>Identifying Chemical Names in Biomedical Text: an Investigation of Substring Co-occurrence Based Approaches</i> Alexander Vasserman	7
<i>Maximum Entropy Modeling in Sparse Semantic Tagging</i> Jia Cui and David Guthrie	13
<i>A Computational Framework for Non-Lexicalist Semantics</i> Jimmy Lin	19
<i>A Multi-Path Architecture for Machine Translation of English Text into American Sign Language Animation</i> Matt Huenerfauth	25
<i>Automatic Article Restoration</i> John Lee	31
<i>A Preliminary Look into the Use of Named Entity Information for Bioscience Text Tokenization</i> Robert Arens	37
<i>Greek Word Segmentation Using Minimal Information</i> C. Anton Rytting	43
<i>Construction of Conceptual Graph Representation of Texts</i> Svetlana Hensman	49
<i>Speaker Recognition with Mixtures of Gaussians with Sparse Regression Matrices</i> Constantinos Boulis	55

Student Workshop Author Index

Arens, Robert	37
Boulis, Constantinos	55
Cui, Jia	13
Guthrie, David	13
Hensman, Svetlana	49
Huenerfauth, Matt	25
Lee, John	31
Lin, Jimmy	19
Ramakrishnan, S.K.	1
Rytting, C. Anton	43
Swaminathan, R.	1
Udhaykumar, N.	1
Vasserman, Alexander	7