ACL-08: HLT

The Third Workshop on Issues in Teaching Computational Linguistics (TeachCL-08)

Proceedings of the Workshop

June 19–20, 2008
The Ohio State University
Columbus, Ohio, USA

Production and Manufacturing by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53707
USA

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-932432-14-5

Introduction

Many of us in this field face the daily challenge of trying to teach computer scientists, linguists and/or psychologists together. Following the success of the two previous ACL workshops (2002 and 2005, `http://www.eecs.umich.edu/~radev/TeachingNLP`) on this theme, we held this 2-day workshop associated with ACL-HLT 2008 to carefully examine the advantages and disadvantages of an interdisciplinary approach, and to explore techniques specifically aimed at teaching programming to social scientists and linguistics to computer scientists. As computational linguistics (hopefully) becomes of more and more relevance to industrial applications, we must ensure that our students (both undergraduate and graduate) are given adequate preparation for functioning in a practical industrial environment as well as an academic research environment. We need to exchange views on appropriate curriculum for both undergraduate students and graduate students, and linguists, psychologists and computer scientists. There are many questions to be addressed about the necessary background for linguists, psychologists and computer scientists before they can communicate effectively with each other and learn at the same pace.

How much math is necessary? Is it possible to teach linguists Natural Language Processing techniques without first teaching them how to program? Can undergraduates hold their own in graduate courses? Can linguists and computer scientists make separate but equal contributions to term projects? How can linguistics students get ACL publications? What is the relevance of psycholinguistics?

In addition to fifteen high quality reviewed papers and an invited talk by Lori Levin and Drago Radev on the recent very successful American Computational Linguistics Olympiad, the program includes three panels: a panel on industry expectations for computational linguists organized by Chris Brew; a panel on essential curriculum for computational linguistics organized by Emily Bender and Fei Xia; and a panel on techniques for teaching extremely interdisciplinary classes organized by Gina Levow.

The specific goals of this workshop build upon the goals of past CL teaching workshops:

- To provide a setting for mutual feedback on participants' instructional approaches as well as guidance on future directions.
- To identify and separate from the general teaching aspirations of host departments the key features of an elective undergraduate and graduate curriculum in computational linguistics.
- To determine a curriculum that embraces diversity of background as an opportunity rather than shies from it as a problem.
- To generally promote visibility for the study of CL teaching as a bona fide scholarly activity
- In the case of the industrial panel, to set up a situation in which those responsible for education and training in CL-using industry become more aware of the diversity of backgrounds available in the ACL world.

We are especially grateful to the panel organizers, the presenters who submitted excellent papers and to our hard working program committee. Particular thanks go to Richard Wicentowski for being Publications Chair.

Martha Palmer, Chris Brew, Fei Xia

Organizers:

Martha Palmer, University of Colorado, USA
Chris Brew, The Ohio State University, USA
Fei Xia, University of Washington, USA

Program Committee:

Steven Bird, Melbourne University, Australia
Robert Dale, Macquarie University, Australia
Jason Eisner, Johns Hopkins University, USA
Tomaz Erjavec, Josef Stefan Institute, Slovenia
Mary Harper, University of Maryland, USA
Julia Hirschberg, Columbia University, USA
Graeme Hirst, University of Toronto, Canada
Julia Hockenmaier, University of Illinois - UIUC, USA
Ewan Klein, University of Edinburgh, UK
Lillian Lee, Cornell University, USA
Lori Levin, Carnegie Mellon University, USA
Gina-Anne Levow, University of Chicago, USA
Liz Liddy, Syracuse University, USA
Edward Loper, University of Pennsylvania, USA
Detmar Meurers, Universität Tübingen, Germany
Ani Nenkova, University of Pennsylvania, USA
James Pustejovksy, Brandeis University, USA
Massimo Poesio, University of Trento, Italy / University of Essex, UK
Dragomir Radev, University of Michigan, USA
Anoop Sarkar, Simon Fraser University, Canada,
Harold Somers, University of Manchester, UK
Matthew Stone, Rutgers University, USA
Richard Wicentowski (Publications Chair), Swarthmore College, USA
Dekai Wu, Hong Kong University of Science and Technology, China

Invited Speakers:

Dragomir Radev, University of Michigan, USA
Lori Levin, Carnegie-Mellon University, USA

Panel Organizers:

Emily Bender, University of Washington, USA
Chris Brew, The Ohio State University, USA
Gina-Anne Levow, University of Chicago, USA
Fei Xia, University of Washington, USA

Table of Contents

Teaching Computational Linguistics to a Large, Diverse Student Body: Courses, Tools, and Interdepartmental Interaction
Jason Baldridge and Katrin Erk 1

Building a Flexible, Collaborative, Intensive Master's Program in Computational Linguistics
Emily M. Bender, Fei Xia and Erik Bansleben 10

Freshmen's CL Curriculum: The Benefits of Redundancy
Heike Zinsmeister 19

Defining a Core Body of Knowledge for the Introductory Computational Linguistics Curriculum
Steven Bird 27

Strategies for Teaching "Mixed" Computational Linguistics Classes
Eric Fosler-Lussier 36

The Evolution of a Statistical NLP Course
Fei Xia 45

Exploring Large-Data Issues in the Curriculum: A Case Study with MapReduce
Jimmy Lin 54

Multidisciplinary Instruction with the Natural Language Toolkit
Steven Bird, Ewan Klein, Edward Loper and Jason Baldridge 62

Combining Open-Source with Research to Re-engineer a Hands-on Introductory NLP Course
Nitin Madnani and Bonnie J. Dorr 71

Zero to Spoken Dialogue System in One Quarter: Teaching Computational Linguistics to Linguists Using Regulus
Beth Ann Hockey and Gwen Christian 80

The North American Computational Linguistics Olympiad (NACLO)
Dragomir R. Radev, Lori Levin and Thomas E. Payne 87

Competitive Grammar Writing
Jason Eisner and Noah A. Smith 97

Studying Discourse and Dialogue with SIDGrid
Gina-Anne Levow 106

Teaching NLP to Computer Science Majors via Applications and Experiments
Reva Freedman 114

Psychocomputational Linguistics: A Gateway to the Computational Linguistics Curriculum
William Gregory Sakas 120

Support Collaboration by Teaching Fundamentals
Matthew Stone 129

Workshop Program

Thursday, June 19, 2008

8:55–9:00 Welcome

Paper Presentation I: Curriculum Design

9:00–9:30 *Teaching Computational Linguistics to a Large, Diverse Student Body: Courses, Tools, and Interdepartmental Interaction*
Jason Baldridge and Katrin Erk

9:30–10:00 *Building a Flexible, Collaborative, Intensive Master's Program in Computational Linguistics*
Emily M. Bender, Fei Xia and Erik Bansleben

10:00–10:30 *Freshmen's CL Curriculum: The Benefits of Redundancy*
Heike Zinsmeister

10:30–11:00 Coffee Break

Paper Presentation II and Panel I: Curriculum Design

11:00–11:30 *Defining a Core Body of Knowledge for the Introductory Computational Linguistics Curriculum*
Steven Bird

11:30–12:30 Panel Discussion I: Curriculum Design (organized by Fei Xia and Emily Bender)

12:30–2:00 Lunch Break

Thursday, June 19, 2008 (continued)

Paper Presentation III: Course Design

2:00–2:30 *Strategies for Teaching "Mixed" Computational Linguistics Classes*
Eric Fosler-Lussier

2:30–3:00 *The Evolution of a Statistical NLP Course*
Fei Xia

3:00–3:30 *Exploring Large-Data Issues in the Curriculum: A Case Study with MapReduce*
Jimmy Lin

3:30–4:00 Coffee Break

Paper Presentation IV: Using NLP Tools

4:00–4:30 *Multidisciplinary Instruction with the Natural Language Toolkit*
Steven Bird, Ewan Klein, Edward Loper and Jason Baldridge

4:30–5:00 *Combining Open-Source with Research to Re-engineer a Hands-on Introductory NLP Course*
Nitin Madnani and Bonnie J. Dorr

Panel II: Industry Panel

5:00–6:00 Panel Discussion II: Industry Panel (organized by Chris Brew)

Friday, June 20, 2008

Friday, June 20, 2008 (continued)

Paper Presentation V and Invited Talk

9:00–9:30 *Zero to Spoken Dialogue System in One Quarter: Teaching Computational Linguistics to Linguists Using Regulus*
Beth Ann Hockey and Gwen Christian

9:30–10:30 *The North American Computational Linguistics Olympiad (NACLO)*
Dragomir R. Radev, Lori Levin and Thomas E. Payne

10:30–11:00 Coffee Break

Paper Presentation VI: Course Design

11:00–11:30 *Competitive Grammar Writing*
Jason Eisner and Noah A. Smith

11:30–12:00 *Studying Discourse and Dialogue with SIDGrid*
Gina-Anne Levow

12:00–12:30 *Teaching NLP to Computer Science Majors via Applications and Experiments*
Reva Freedman

12:30–1:30 Lunch Break

Paper Presentation VII: Course Design

1:30–2:00 *Psychocomputational Linguistics: A Gateway to the Computational Linguistics Curriculum*
William Gregory Sakas

2:00–2:30 *Support Collaboration by Teaching Fundamentals*
Matthew Stone

Friday, June 20, 2008 (continued)

Panel III: Course Design

2:30–3:30 Panel Discussion III: Course Design (organized by Gina-Anne Levow)

3:30–4:00 Coffee Break

4:00–5:00 General Discussion and Closing