

John Blitzer

387 Soda Hall Berkeley, CA 94720-1776
blitzer@cs.berkeley.edu
<http://john.blitzer.com/>

INTERESTS **Machine Learning, Natural Language Processing, Data Mining**

EDUCATION **University of Pennsylvania**, Philadelphia, PA
PhD in Computer and Information Science
Fall 2002 - Fall 2007. GPA: 4.0

Cornell University, Ithaca, NY
AB in Computer Science, 2002. GPA: 3.75

REFEREED
PUBLICATIONS

Regularized Learning with Networks of Features.

Ted Sandler, John Blitzer, Partha Talukdar, and Lyle Ungar.
Neural Information Processing Systems - NIPS 2008.

Using English Information in Non-English Web Search.

Wei Gao, John Blitzer, and Ming Zhou.
Improving Non-English Web Search - INEWS 2008.

Multi-View Learning over Structured and Non-Identical Outputs.

Kuzman Ganchev, Joao Graca, John Blitzer, and Ben Taskar.
Uncertainty in Artificial Intelligence - UAI 2008.

DRASO: Declaratively Regularized Alternating Structural Optimization.

Partha Talukdar, John Blitzer, Ted Sandler, Mark Dredze, Koby Crammer, Fernando Pereira.
ICML 2008 Workshop on Prior Knowledge in Text and Language Processing.

Intelligent Email: Reply and Attachment Prediction.

Mark Dredze, Tova Broks, Josh Carroll, Joshua Magarick, John Blitzer, and Fernando Pereira.
Intelligent User Interfaces - IUI 2008.

Learning Bounds for Domain Adaptation.

John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, Jenn Wortman.
Neural Information Processing Systems - NIPS 2007.

Frustratingly Hard Domain Adaptation for Parsing.

Mark Dredze, John Blitzer, Partha Pratim Talukdar, Kuzman Ganchev, Joao Graca, and Fernando Pereira.
Conference on Natural Language Learning - CoNLL 2007.

Biographies, Bollywood, Boom-boxes, and Blenders: Domain Adaptation for Sentiment Classification.

John Blitzer, Mark Dredze, and Fernando Pereira.
Association for Computational Linguistics - ACL 2007.

Analysis of Representations for Domain Adaptation.

Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira.
Neural Information Processing Systems - NIPS 2006.

“Sorry, I forgot the attachment.” Email Attachment Prediction

Mark Dredze, John Blitzer, and Fernando Pereira.
Conference on Email and Anti-Spam - CEAS 2006.

Domain Adaptation with Structural Correspondence Learning.

John Blitzer, Ryan McDonald, and Fernando Pereira.
Empirical Methods in Natural Language Processing - EMNLP 2006.

Distance Metric Learning for Large Margin Nearest Neighbor Classification.

Kilian Weinberger, John Blitzer, and Lawrence Saul.
19th annual conference on Neural Information Processing Systems - NIPS 2005.

Reply Expectation Prediction for Email Management.

Mark Dredze, John Blitzer, and Fernando Pereira.
2nd Conference on Email and Anti-Spam - CEAS 2005.

Distributed Latent Variable Models of Lexical Co-occurrences.

John Blitzer, Amir Globerson, and Fernando Pereira.
10th International Workshop on Artificial Intelligence and Statistics - AISTATS 2005.

Hierarchical Distributed Representations for Statistical Language Modeling.

John Blitzer, Kilian Weinberger, Lawrence Saul, and Fernando Pereira.
18th annual conference on Neural Information Processing Systems - NIPS 2004.

MEAD - A Platform for Multidocument Multilingual Text Summarization.

Dragomir Radev, Timothy Allison, Sasha Blair-Goldensohn, John Blitzer, Arda Celebi, Stanko Dimitrov, Elliott Drabek, Ali Hakim, Wai Lam, Danyu Liu, Jahna Otterbacher, Hong Qi, Horacio Saggion, Simone Teufel, Michael Topper, Adam Winkel, and Zhu Zhang.
Language Resources and Evaluation - LREC 2003.

Latent Variable Models for Syntactic Co-occurrence Data.

John Blitzer and Fernando Pereira.
NIPS workshop on Syntax, Semantics, and Statistics - 2003.

Evaluation challenges in large-scale multi-document Summarization.

Dragomir Radev, Simone Teufel, Wai Lam, Horacio Saggion, John Blitzer, Hong Qi, Arda Celebi, Danyu Liu, and Elliott Drabek.
41st Annual Meeting of the Association for Computational Linguistics - ACL 2003.

Summarizing Archived Discussions: A Beginning.

Paula Newman and John Blitzer.
Intelligent User Interfaces - IUI 2002.

TALKS AND
POSTER
PRESENTATIONS

Conference on Natural Language Learning. June 2007. Prague, Czech Republic.
Neural Information Processing Systems. December 2006. Vancouver, Canada.
Empirical Methods in Natural Language Processing. July 2007. Sydney, Australia.
Snowbird Learning Workshop. April 2007. Utah, United States.
Neural Information Processing Systems. December 2005. Vancouver, Canada.
Conference on Artificial Intelligence and Statistics. January 2005. Barbados.
Neural Information Processing Systems. December 2004. Vancouver, Canada.

PATENTS

PARC Docket A2560Q1. Method And Apparatus For Generating Overview Information For Hierarchically Related Information

PARC Docket A2560Q2. Method And Apparatus For Generating Summary Information For Hierarchically Related Information

PROFESSIONAL
ACTIVITIES

Tutorial Organizer

ACL 2008 Tutorial on Semi-supervised Learning.
Co-taught with Xiaojin Zhu (University of Wisconsin).

Reviewer / Program Committee Member

Association for Computational Linguistics, ACL (2008).
Association for the Advancement of Artificial Intelligence (2008).
International Conference on Machine Learning, ICML (2006, 2007, 2008).
North American Association for Computational Linguistics, NAACL (2007).
Neural Information Processing Systems, NIPS (2007, 2008).
Conference on Artificial Intelligence and Statistics, AISTATS (2007).
Empirical Methods in Natural Language Processing, EMNLP (2007).

Workshop Organizer

NIPS 2006 workshop on Applications of Dimensionality Reduction.
Co-organized with Kilian Weinberger, Rajarshi Das, and Irina Rish.

RESEARCH AND
INDUSTRY
EXPERIENCE

University of California, Berkeley, Department of Computer Science (August 2008): Postdoctoral Researcher in the natural language processing group.

Microsoft Research Asia (February 2008 - August 2008): Visiting Researcher. Developed techniques for using machine learning in cross-lingual ranking. For many Chinese queries, we can find an accurate English (or other foreign) translation. Our system learns a better ranking function using information from both Chinese and foreign results.

University of Pennsylvania, Computer and Information Science Department (Fall 2002 - Fall 2007): PhD Student. My PhD thesis focused on using low-dimensional representations to adapt language tools for use in new domains, like transcribed speech and biomedical literature.

SRI CALO Project, University of Pennsylvania CIS department (Fall 2003 - present): Designing a Cognitive Agent that Learns and Organizes. We developed email reply prediction and contextual document search and ranking systems.
<http://www.ai.sri.com/project/CALO>

Google (Summer 2004): Intern with the Google News team. Developed a story development timeline view for Google News. <http://news.google.com/>

Neural Information Processing Systems (Summer and Fall 2003): Managed submission website including scripting, database design, and website maintenance.

TEACHING

HIT/MSRA Joint Summer School on NLP

Harbin Institute of Technology and Microsoft Research Asia 2008 joint summer school on natural language processing. Attended by students from universities throughout China. I taught a course on supervised and semi-supervised learning with linear models.

CSE-121, Introduction to Cognitive Science. University of Pennsylvania, 2003. Teaching Assistant. CSE-121 is an introductory course which covers topics in computer science, linguistics, and psychology. I taught recitation and prepared homework assignments.

CSE-313, Computational Linear Algebra. University of Pennsylvania, 2004. Teaching Assistant. CSE-313 introduces students to techniques in numerical computing such as matrix factorization, techniques for solving linear equations, and eigenvalue problems. I taught recitation and prepared homework assignments.

NATURAL
LANGUAGES

English. Native Fluency.

Mandarin Chinese. Conversational Fluency. Earned a Level 3 on the National Chinese Proficiency Test (Hanyu Shuiping Kaoshi). Certified to attend college in Chinese.

German. Four years of high school and two years of college coursework.

AWARDS

Winner of UCSD Student Data Mining Competition (2006).

University of Pennsylvania Graduate Student Fellowship (2002-2007).

University of Pennsylvania CIS Department Chair's Award (2002).

Lockheed Martin Research Award (2000).

Golden Key International Honor Society (2000-2002).