John Blitzer
Google
Mountain View, CA 94043
john@blitzer.com
http://john.blitzer.com/

INTERESTS	**Natural Language Processing, Machine Learning, Data Mining**
CURRENT POSITION	**Google**, Mountain View, CA. Research Scientist. August 2010 – present.
PREVIOUS POSITION	**University of California, Berkeley**, Berkeley, CA. Department of Computer Science. Postdoctoral Researcher in the natural language processing group. August 2008 – July 2010.
EDUCATION	**University of Pennsylvania**, Philadelphia, PA. Ph.D. in Computer and Information Science. Advisor: Fernando Pereira. Fall 2002 - Fall 2007. GPA: 4.0. **Cornell University**, Ithaca, NY. AB in Computer Science, 2002. GPA: 3.75.
AWARDS	**Intelligence Community Postdoctoral Fellowship**. 2009. **Winner of UCSD Student Data Mining Competition**. 2006. **University of Pennsylvania CIS Department Chair's Award**. 2002. **Lockheed Martin Research Award**. 2000.
PUBLICATION STATISTICS	**Citation count and h-index.** According to Google Scholar, the articles below have been cited a total of 600 times. My h-index (See http://en.wikipedia.org/wiki/Hirsch number) is 10. As a primary author, I have been cited a total of 240 times, and my h-index is 6.
TOP CITED PUBLICATIONS	**Learning Bounds for Domain Adaptation. (23 Citations)** John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, Jenn Wortman. *Neural Information Processing Systems.* NIPS 2008. **Biographies, Bollywood, Boom-boxes, and Blenders: Domain Adaptation for Sentiment Classification. (74 citations)** John Blitzer, Mark Dredze, and Fernando Pereira. *Association for Computational Linguistics.* ACL 2007.

Analysis of Representations for Domain Adaptation. (49 citations)
Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira.
Neural Information Processing Systems. NIPS 2007.

Domain Adaptation with Structural Correspondence Learning. (73 citations)
John Blitzer, Ryan McDonald, and Fernando Pereira.
Empirical Methods in Natural Language Processing. EMNLP 2006.

Distance Metric Learning for Large Margin Nearest Neighbor Classification. (213 citations)
Kilian Weinberger, John Blitzer, and Lawrence Saul.
19th annual conference on Neural Information Processing Systems. NIPS 2006.

Evaluation challenges in large-scale multi-document Summarization. (41 citations)
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.

JOURNAL
PUBLICATIONS

A Theory of Learning from Multiple Domains.
Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jenn Wortman. *Machine Learning.* Special Issue on Learning From Multiple Sources. 2009.

OTHER
REFEREED
PUBLICATIONS

Better Word Alignment with Supervised ITG Models.
Aria Haghighi, John Blitzer, John DeNero, and Dan Klein.
Association for Computational Linguistics. ACL 2009.

Exploiting Bilingual Information to Improve Monolingual Web Search.
Wei Gao, John Blitzer, Ming Zhou, and Kam-Fai Wong.
Association for Computational Linguistics. ACL 2009.

Regularized Learning with Networks of Features.
Ted Sandler, John Blitzer, Partha Talukdar, and Lyle Ungar.
Neural Information Processing Systems. NIPS 2009.

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.

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.

"Sorry, I forgot the attachment." Email Attachment Prediction.
Mark Dredze, John Blitzer, and Fernando Pereira.
Conference on Email and Anti-Spam. CEAS 2006.

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 2005.

Latent Variable Models for Syntactic Co-occurrence Data.
John Blitzer and Fernando Pereira.
NIPS workshop on Syntax, Semantics, and Statistics. 2003.

Summarizing Archived Discussions: A Beginning.
Paula Newman and John Blitzer.
Intelligent User Interfaces. IUI 2002.

INVITED TALKS   **Microsoft Research Asia, Beijing.** (Contact: Ming Zhou). Jul 2009.

Toyota Technological Institute, Chicago. (Contact: Sham Kakade). Nov 2008.

Harbin Institute of Technology. (Contact: Tiejun Zhao). Jul 2008.

Microsoft Research Asia, Beijing. (Contact: Ming Zhou). Apr 2008.

Massachusetts Institute of Technology. (Contact: Michael Collins). Oct 2007.

Microsoft Research, Redmond. (Contact: Chris Burges). Sept 2007.

Google, Mountain View. (Contact: Hiyan Alshawi). Sept 2007.

Yahoo Research, Santa Clara. (Contact: Raghu Ramakrishnan). Sept 2007.

Yahoo Research, New York. (Contact: Tong Zhang). Aug 2007.

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.
`http://john.blitzer.com/harbin/index.html`

Tutorial Organizer

ACL 2008 Tutorial on Semi-supervised Learning, with Jerry Zhu (U.W. Madison). The tutorial covered semi-supervised learning for natural language processing and focused primarily on three topics: Self-training and bootstrapping, graph regularization, and structural learning. `http://ssl-acl08.wikidot.com/`

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.

CSE-313, Numerical 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.

PROFESSIONAL ACTIVITIES

Reviewer / Program Committee Member

Association for Computational Linguistics, ACL. 2008, 2009.
Association for the Advancement of Artificial Intelligence. 2008.
International Joint Conference on Artificial Intelligence. 2009.
International Conference on Machine Learning, ICML. 2006, 2007, 2008, 2009.
North American Association for Computational Linguistics, NAACL. 2007,2009.
Neural Information Processing Systems, NIPS. 2007, 2008.
Conference on Artificial Intelligence and Statistics, AISTATS. 2007,2010.
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.

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 course work.