

ANDREW B. GOLDBERG

CONTACT INFORMATION

Department of Computer Sciences
University of Wisconsin–Madison
1210 West Dayton St
Madison, WI 53706-1685 USA

(608) 335-6085
goldberg@cs.wisc.edu
www.cs.wisc.edu/~goldberg

RESEARCH INTERESTS

Statistical machine learning, semi-supervised learning, natural language processing, sentiment analysis, graph-based algorithms, information retrieval

EDUCATION

University of Wisconsin–Madison, Madison, WI USA

Ph.D., Computer Science, Expected Summer 2010

- Area of Study: Artificial Intelligence (Machine Learning)
- Advisor: Assistant Professor Xiaojin “Jerry” Zhu
- Dissertation: *New Directions in Semi-Supervised Learning*
- Ph.D. Minor: Statistics

M.S., Computer Science, December 2006, GPA 4.0

Amherst College, Amherst, MA USA

B.A., Computer Science, May 2003, GPA 4.0 (CS), 3.93 (overall)

- *Magna cum laude*, with departmental distinction
- Advisors: Professor Lyle A. McGeoch and Professor Catherine C. McGeoch
- Honors Thesis: *Highly Constrained Sports Scheduling with Genetic Algorithms*

Universiteit van Amsterdam, Amsterdam, The Netherlands

Study abroad program in Humanities and Social Sciences, Spring 2002, GPA 4.0

ACADEMIC RESEARCH EXPERIENCE

University of Wisconsin–Madison, Madison, WI USA

Research Assistant (Professor Xiaojin Zhu) September 2006 to present

- Current research projects: Semi-Supervised Learning for Non-Experts (NSF); A Cognitive Study of Learning with Labeled and Unlabeled Data (AFOSR). Text-to-Picture Synthesis (NSF).

UW-Madison Fellow September 2005 to August 2006

- Course projects: Graph-based semi-supervised sentiment categorization; gene clustering using microarray data and biomedical literature; Madison bus trip planner using A* search and Google maps.

Amherst College, Amherst, MA USA

Undergraduate Researcher September 2001 to May 2003

- Senior honors thesis examined using genetic algorithms to schedule the NESCAC (New England Small College Athletic Conference) men’s soccer season.
- Other projects included building an expert system in Lisp and developing a snake-like virtual robot that learns locomotion in a simulated environment.

INDUSTRY
EXPERIENCE

Microsoft Research, Mountain View, CA USA

Research Intern May 2008 to August 2008

- Mentored by Ariel Fuxman in the Search Labs research group.
- Developed novel methods for query classification and automatic labeling of log data (see *KDD 2009* paper and patent filing).

Google, Inc., Mountain View, CA USA

Software Engineering Intern (in Research) May 2007 to August 2007

- Worked with Peng Xu in the statistical machine translation research group.
- Designed and implemented n -best list re-ranking algorithms using large-scale, distributed supervised and semi-supervised machine learning.

Deitel & Associates, Inc., Maynard, MA USA

Software Developer / Textbook Author June 2003 to June 2005

- Co-authored and edited CS textbooks and 100-page OOD/UML case study.
- Managed team of interns and held project-control responsibilities.

Microsoft Corporation, Redmond, WA USA

Software Test Engineer Intern June 2002 to August 2002

- Interacted closely with developers and program managers to design and implement test plans for features of Microsoft Office.
- Tested malicious code detection script and SOAP APIs in Office.

Prodigy Communications Corporation, Yorktown Heights, NY USA

Application Developer Intern June 2001 to August 2001

- Created applications for internal portal, including trouble-ticket tracking tool.
- Programmed load and stress tests for Prodigy Internet IM and Web applications.

Prodigy Communications Corporation, White Plains, NY USA

Software Test Analyst Intern June 2000 to August 2000, January 2001

- Tested Web applications (e.g., financial portfolio tracker, Webmail) and dial-up software for national ISP Prodigy Internet.

TEACHING
EXPERIENCE

University of Wisconsin–Madison, Madison, WI USA

Reading Group Leader Fall 2009

- Organized and led reading group in Non-Parametric Bayesian statistics for machine learning.

Guest Lecturer Spring 2007 to present

- Gave guest lectures in UW-Madison's CS 540: Introduction to Artificial Intelligence and CS 769: Advanced Natural Language Processing.

Mentor to Undergraduates Fall 2006 to present

- Oversee NSF REU students in Text-to-Picture synthesis project.
- Managed student software development to support interdisciplinary Psychology research project in Embodied Computer Aided Instruction (see *Instructional Science* journal article).

Amherst College, Amherst, MA USA

Lab Teaching Assistant and Grader

January 2001 to May 2003

- Aided students in introductory Computer Science lab programming assignments, organized review sessions, and graded multivariate calculus homework assignments.

BOOKS

Xiaojin Zhu and Andrew B. Goldberg. *Introduction to Semi-Supervised Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2009.

Harvey M. Deitel, Paul J. Deitel, and Andrew B. Goldberg. *Internet & World Wide Web How to Program*. 3rd ed. Prentice Hall, 2003.

CONFERENCE
& JOURNAL
PUBLICATIONS

Andrew B. Goldberg and Xiaojin Zhu. Keepin' it real: Semi-supervised learning with realistic tuning. *NAACL 2009 Workshop on Semi-supervised Learning for NLP*.

Ariel Fuxman, Anitha Kannan, Andrew B. Goldberg, Rakesh Agrawal, Panayiotis Tsaparas, John Shafer. Improving classification accuracy using automatically extracted training data. *15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Industrial Track (KDD 2009)*.

Arthur Glenberg, Andrew B. Goldberg, and Xiaojin Zhu. Improving early reading comprehension using embodied CAI. *Instructional Science*, 2009.

Andrew B. Goldberg, Nathanael Fillmore, David Andrzejewski, Zhiting Xu, Bryan Gibson and Xiaojin Zhu. May All Your Wishes Come True: A Study of Wishes and How to Recognize Them. *Annual Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2009)*.

Andrew B. Goldberg, Xiaojin Zhu, Aarti Singh, Zhiting Xu, and Robert Nowak. Multi-manifold semi-supervised learning. *Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS 2009)*.

Xiaojin Zhu, Andrew B. Goldberg, and Tushar Khot. Some new directions in graph-based semi-supervised learning. (invited paper) *IEEE International Conference on Multimedia and Expo (ICME), Special Session on Semi-Supervised Learning for Multimedia Analysis, 2009*.

Andrew B. Goldberg, Ming Li, and Xiaojin Zhu. Online Manifold Regularization: A New Learning Setting and Empirical Study. *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2008)*.

Andrew B. Goldberg, Xiaojin Zhu, Charles R. Dyer, Mohamed Eldawy, and Lijie Heng. Easy as ABC? Facilitating pictorial communication via semantically enhanced layout. *12th Conference on Computational Natural Language Learning (CoNLL 2008)*.

Xiaojin Zhu, Andrew B. Goldberg, Michael Rabbat, and Robert Nowak. Learning bigrams from unigrams. *46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL 2008)*.

Xiaojin Zhu and Andrew B. Goldberg. Kernel regression with order preferences. *22nd AAAI Conference on Artificial Intelligence (AAAI 2007)*.

Xiaojin Zhu, Andrew B. Goldberg, Mohamed Eldawy, Charles Dyer, and Bradley Strock. A text-to-picture synthesis system for augmenting communication. *The Integrated Intelligence Track of AAAI 2007*.

Xiaojin Zhu, Andrew B. Goldberg, Jurgen Van Gael, and David Andrzejewski. Improving diversity in ranking using absorbing random walks. *Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2007)*.

Andrew B. Goldberg, Xiaojin Zhu, and Stephen Wright. Dissimilarity in graph-based semi-supervised classification. *11th International Conference on Artificial Intelligence and Statistics (AISTATS 2007)*.

Andrew B. Goldberg, David Andrzejewski, Jurgen Van Gael, Burr Settles, Xiaojin Zhu, and Mark Craven. Ranking biomedical passages for relevance and diversity: UW-Madison at TREC Genomics 2006. *15th Text Retrieval Conference (TREC 2006)*.

Andrew B. Goldberg and Xiaojin Zhu. Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. *HLT-NAACL 2006 TextGraphs Workshop: Graph-based Algorithms for Natural Language Processing*.

PATENTS *Query Classification Using Implicit Labels*. Filed in 2009 (based on work done during internship at Microsoft Research).

INVITED TALKS *Text-to-Picture Synthesis*. HAMLET (Human, Animal, and Machine Learning: Experiment and Theory) interdisciplinary lecture series. October 2009.

Improving diversity in ranking using absorbing random walks. SIGIR 2009 workshop: Redundancy, Diversity, and Interdependent Document Relevance. July 2009.

Invited panelist, NAACL HLT 2009 Workshop on Semi-supervised Learning for Natural Language Processing. June 2009.

Novel Extensions of Graph-Based Semi-Supervised Learning. Communications and Signal Processing research group, UW-Madison Department of Electrical and Computer Engineering. February 2009.

Learning bigrams from unigrams. Machine Learning Group, University of Cambridge, United Kingdom. September 2008.

AWARDS
Yahoo! Key Technical Challenges Grant, 2008–2009
AISTATS and Machine Learning Summer School travel awards 2009
UW-Madison First-Year Graduate School Fellowship, 2005–2006
Phi Beta Kappa, 2003
Sigma Xi Scientific Research Society, 2003
National Merit Scholarship, 1999
Robert C. Byrd Honors Scholarship, 1999
High School Valedictorian, 1999

SERVICE
Program Committee for ICML 2010, EMNLP 2009, MLG-2009, ACL-IJCNLP 2009
Reviewer for Data Mining and Knowledge Discovery journal, June 2009.
Program Committee for TextGraphs-3 workshop at Coling, August 2008

Program Committee for AAAI 2008 Web Track
Program Committee for TextGraphs-2 workshop at HLT-NAACL, April 2007
Reviewer for ACM Transactions on Information Systems, February 2007
Co-founder of UW-Madison CS Graduate Mentoring Committee, Fall 2006 to present
UW-Madison Guide to Graduate School Life contributing editor, Spring 2009.
UW-Madison CS Department Admissions Committee, Winter 2007 and Winter 2008
Webmaster of the Mathematical Programming Society's Optimization-Online.org e-print repository for the optimization research community, June 2005 to present
Web developer for the Amherst College Daily Jolt community portal (amherst.dailyjolt.com), January 2001 to May 2003
Founding layout designer for Circus literary magazine at Amherst College, Spring 2001
Radio DJ on WAMH, Amherst, MA, Spring 2000 to Fall 2002

TECHNICAL
SKILLS

Java, Matlab, Python, C++, Linux shell scripting, Condor High Throughput Computing, MapReduce, C# & .NET, JavaScript, XHTML/CSS, Ajax, PHP, Perl, SQL, OOD/UML

REFERENCES

Available upon request.