

David Mimno

35 Olden St.
Princeton, NJ
mimno@cs.princeton.edu

Education

University of Massachusetts, Amherst, 2005-

- Department of Computer Science.
- Advisor: Andrew McCallum
- Research Interests: Machine Learning, NLP, Text Mining, Topic Models
- MSc Jan 2009
- PhD (expected) Winter 2011

Swarthmore College, 1995-1999

- Bachelor of Arts with High Honors (Classics and Computer Science)

Experience

University of Massachusetts, Amherst. Research Assistant, Fall 2005 - Present.

- Designed and implemented topic modeling package for MALLET toolkit, with dramatic gains in time and memory performance.
- Collaborated on publication of theoretical advances in statistical topic modeling.
- Supervised development and support of MALLET machine learning toolkit.
- Co-organized NESCAI 2010 student AI conference.

Perseus Project, Tufts University, Medford, MA. Head Programmer, Fall 2002 - Fall 2005.

- Supervised a high-traffic digital library for historical texts and images.
- Rebuilt production website infrastructure in Java.
- Designed new ways to integrate text with contextual information from databases.
- Trained and managed student workers.

University of Sheffield, Sheffield, UK. Web Programmer, 2002.

- Developed components for General Architecture for Text Engineering (GATE) entry in ACE information extraction competition.

TradeOut, Inc., Cambridge, MA. Lead Web Programmer, Fall 1999 - Summer 2001.

- Designed, built, and supported an e-commerce web application.
- Served as technical lead to a group of seven programmers.
- Interacted with Quality Assurance, Customer Service, and Product Development personnel.

Publications

Peer-reviewed Conference Papers

- Rethinking LDA: Why Priors Matter. Hanna Wallach, David Mimno and Andrew McCallum. NIPS, 2009, Vancouver, BC.
- Polylingual Topic Models. David Mimno, Hanna Wallach, Jason Naradowsky, David Smith and Andrew McCallum. EMNLP, 2009, Singapore.
- Evaluation Methods for Topic Models. Hanna Wallach, Iain Murray, Ruslan Salakhutdinov and David Mimno. ICML, 2009, Montreal, Quebec.
- Efficient Methods for Topic Model Inference on Streaming Document Collections. Limin Yao, David Mimno and Andrew McCallum. KDD, 2009, Paris, France.
- Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression. David Mimno and Andrew McCallum. UAI, 2008, Helsinki, Finland.
- Expertise Modeling for Matching Papers with Reviewers. David Mimno and Andrew McCallum. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) 2007, San Jose, CA.
- Mixtures of Hierarchical Topics with Pachinko Allocation. David Mimno, Wei Li and Andrew McCallum. International Conference on Machine Learning (ICML) 2007, Corvallis, OR.
- Mining a digital library for influential authors. David Mimno and Andrew McCallum. Joint Conference on Digital Libraries (JCDL) 2007, Vancouver, BC, Canada.
- Organizing the OCA: Learning faceted subjects from a library of digital books. David Mimno and Andrew McCallum. Joint Conference on Digital Libraries (JCDL) 2007, Vancouver, BC, Canada.
- Beyond Digital Incunabula: Modeling the Next Generation of Digital Libraries. Gregory Crane, David Bamman, Lisa Cerrato, Alison Jones, David Mimno, Adrian Packel, D. Sculley, and Gabriel Weaver. European Conference on Digital Libraries (ECDL) 2006, Alicante, Spain.
- Bibliometric Impact Measures Leveraging Topic Analysis. Gideon Mann, David Mimno and Andrew McCallum. Joint Conference on Digital Libraries (JCDL) 2006, Chapel Hill, NC.
- Finding a Catalog: Generating Analytical Catalog Records from Well-structured Digital Texts. David Mimno, Alison Jones and Gregory Crane. Joint Conference on Digital Libraries (JCDL) 2005, Denver, CO.
- Services for a Customizable Authority Linking Environment. Mark Patton and David Mimno. demonstration at Joint Conference on Digital Libraries (JCDL) 2004, Tucson, AZ.
- Towards a Cultural Heritage Digital Library. Gregory Crane, Clifford E. Wulfman, Lisa M. Cerrato, Anne Mahoney, Thomas L. Milbank, David Mimno, Jeffrey A. Rydberg-Cox, David A. Smith, and Christopher York. Joint Conference on Digital Libraries (JCDL) 2003, Houston, TX.

Journal Articles

- Computational Historiography. David Mimno (under review).
- Classics in the Million Book Library. Gregory Crane, Alison Babeu, David Bamman, Thomas Breuel, Lisa Cerrato, Daniel Deckers, Anke Luedeling, David Mimno, Rashmi Singhal, David A. Smith, Amir Zeldes. Digital Humanities Quarterly 3(1), Winter 2009.

- Hierarchical Catalog Records: Implementing a FRBR Catalog. David Mimno, Alison Jones and Gregory Crane. DLib. October 2005.

Peer-reviewed Workshop Papers

- Measuring Confidence in Temporal Topic Models with Posterior Predictive Checks. David Mimno and David Blei. Computational Social Science Workshop, NIPS 2010, Whistler, BC.
- Reconstructing Pompeian Households. David Mimno. Applications of Topic Models Workshop, NIPS 2009, Whistler, BC.
- Gibbs Sampling for Logistic Normal Topic Models with Graph-Based Priors. David Mimno, Hanna Wallach and Andrew McCallum. Workshop on Analyzing Graphs, NIPS 2008, Whistler, BC.
- Community-based Link Prediction with Text. David Mimno, Hanna Wallach, and Andrew McCallum. Statistical Network Modeling Workshop, NIPS 2007, Whistler, BC.
- Probabilistic Representations for Integrating Unreliable Data Sources. David Mimno, Andrew McCallum and Jerome Miklau. IIWeb workshop at AAAI 2007, Vancouver, BC, Canada.

Service

Organizer and Co-chair: Fourth Northeast Student Colloquium on Artificial Intelligence (NESCAI) 2010, Amherst, MA

- Organized a three-day, 100+ participant, peer-reviewed conference.

Program committee: JCDL 2008, 2009, 2010; NESCAI 2008

Reviewer: JMLR, IEEE Transactions on the Web, NIPS, ICML, ACL, HLT-NAACL

Chief maintainer: MACHine Learning for Language Toolkit (MALLET).

- Determined development priorities
- Produced documentation and sample code
- Coordinated releases
- Answered user questions

Coordinator, UMass Machine Learning and Friends Lunch. Fall 2008-2010

- Recruited weekly speakers from NY/NE region
- Coordinated corporate sponsorship

Awards and Grants

Google Digital Humanities Research grant. The Open Encyclopedia of Classical Sites. 2010.

Mellon Foundation grant. Cybereditions. 2008.

NSF Fellowship, honorable mention, 2005.

Invited Talks

- Rutgers, January 2011 (scheduled).
- NEH Institute: Networks and Network Analysis for the Humanities. UCLA, August 2010.
- Fast Gibbs sampling for topic models. ICML workshop on topic modeling, Haifa, Israel, June 2010.
- Tutorials: An Introduction to MALLET. University of Maryland, College Park / Johns Hopkins University, Feb 2010.
- Regression-based Topic Models. Google Pittsburgh, October 2009.
- Tutorial: An Introduction to Logistic Regression. Research Experience for Undergraduates. UMass, Amherst, July 2009.
- Text Mining for Humanistic Scholarship. New Directions in Text Analysis Workshop. Harvard University, May 2009.
- Classics in the Million Book Digital Library. Imperial College London/Humboldt University, Berlin, Jan 2008.

Teaching

- Foundations of Probabilistic Modeling, Princeton, Fall 2010. Planned and conducted weekly recitation section.
- NEH Institute: Networks and Network Analysis for the Humanities. UCLA, August 2010. Teaching fellow.
- Natural Language Processing, UMass, 2008. Teaching assistant.
- Extensive informal mentoring of graduate students.