Michael J. Paul

michael.j.paul@colorado.edu
http://cs.jhu.edu/~mpaul/

Last Updated: September 23, 2015

Research interests: natural language processing (especially for the web and social media), machine learning (especially unsupervised and semi-supervised), (probabilistic) topic modeling, health and medical informatics, computational epidemiology, computational social science

EMPLOYMENT

2015 – Present	**University of Colorado Boulder** *Assistant Professor*, Information Science *Assistant Professor by Courtesy*, Computer Science

EDUCATION

2010 – 2015	**Johns Hopkins University** *Ph.D.*, Computer Science, August 2015 *M.S.E.*, Computer Science, May 2012 • Advisors: Mark Dredze (primary) and Jason Eisner • Committee: Mark Dredze, Jason Eisner, Eric Horvitz, Hanna Wallach
2005 – 2009	**University of Illinois at Urbana-Champaign** *B.S. with Honors*, Computer Science, May 2009
Spring 2008	**Lund University** Study abroad: Lund, Sweden

AWARDS AND HONORS

2013 – 2015	**Microsoft Research Fellowship** • One of 12 recipients across North America, and the only recipient in natural language processing	Microsoft Corporation
2010 – 2012	**Graduate Research Fellowship**	National Science Foundation
2010 – 2015	**Dean's Fellowship**	Johns Hopkins University
2005 – 2008	**Edmund J. James Scholar** • Undergraduate honors program	University of Illinois

TEACHING

12/04/2013	**Guest Lecture** • CS465: Natural Language Processing • Instructor: Dr. Jason Eisner • Title: An introduction to topic models	Johns Hopkins University
10/30/2013	**Guest Lecture** • NETS 213: Crowdsourcing & Human Computation • Instructor: Dr. Chris Callison-Burch • Title: Mining health information from Twitter • My slides were reused for Fall 2014	University of Pennsylvania

Spring 2013	**Teaching Assistant**	Johns Hopkins University

- CS476/676: Machine Learning in Complex Domains
- Instructor: Dr. Suchi Saria
- First time this course was offered

Fall 2012	**Teaching Assistant**	Johns Hopkins University

- CS475: Machine Learning
- Instructor: Dr. Mark Dredze
- Lectured on: 9/26/2012 (Generative vs discriminative learning), 10/8/2012 (Kernel methods)
- Substitute lecturer: 2/10/2014 (Regression)

JOURNAL PUBLICATIONS

2015   David A. Broniatowski, Mark Dredze, Michael J. Paul, Andrea Dugas. **Using social media to perform local influenza surveillance in an inner-city hospital: a retrospective observational study.** *JMIR Public Health and Surveillance* 1(1): e5.

Michael J. Paul and Mark Dredze. Sᴘʀɪᴛᴇ**: Generalizing topic models with structured priors.** *Transactions of the Association for Computational Linguistics (TACL)* 3: 43-57.

Shiliang Wang, Michael J. Paul, Mark Dredze. **Social media as a sensor of air quality and public response in China.** *Journal of Medical Internet Research (JMIR)* 17(3): e22.

2014   David A. Broniatowski, Michael J. Paul, Mark Dredze. **Twitter: Big data opportunities (letter).** *Science* 345(6193): 148.

Michael J. Paul, Mark Dredze, David Broniatowski. **Twitter improves influenza forecasting.** *PLOS Currents Outbreaks.* October 2014. doi: 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117.

Michael J. Paul and Mark Dredze. **Discovering health topics in social media using topic models.** *PLOS ONE* 9(8): e103408.

Ahmed Abbasi, Donald Adjeroh, Mark Dredze, Michael J. Paul, Fatemeh Mariam Zahedi, Huimin Zhao, Nitin Walia, Hemant Jain, Patrick Sanvanson, Reza Shaker, Marco D. Huesch, Richard Beal, Wanhong Zheng, Marie Abate, Arun Ross. **Social media analytics for smart health.** *IEEE Intelligent Systems* 29(2):60-80.

Byron C. Wallace, Michael J. Paul, Urmimala Sarkar, Thomas A. Trikalinos, Mark Dredze. **A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews.** *Journal of the American Medical Informatics Association (JAMIA)* 21(6), 1098-1103.

2013   David A. Broniatowski, Michael J. Paul, Mark Dredze. **National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic.** *PLOS ONE* 8(12): e83672.

2011   Roxana Girju and Michael J. Paul. **Modeling reciprocity in social interactions with probabilistic latent space models.** *Natural Language Engineering 17(1).* Cambridge University Press 2011.

PEER-REVIEWED CONFERENCE PUBLICATIONS

2015	Michael J. Paul, Ryen W. White, Eric Horvitz. **Diagnoses, decisions, and outcomes: Web search as medical decision support for cancer.** *24th International World Wide Web Conference (WWW 2015)*, Florence, Italy. May 2015.
2013	Alex Lamb, Michael J. Paul, Mark Dredze. **Separating fact from fear: Tracking flu infections on Twitter.** *2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013)*, Atlanta. June 2013.
	Michael J. Paul and Mark Dredze. **Drug extraction from the Web: Summarizing drug experiences with multi-dimensional topic models.** *2013 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2013)*, Atlanta. June 2013.
2012	Michael J. Paul and Mark Dredze. **Factorial LDA: Sparse multi-dimensional models of text.** In *Advances in Neural Information Processing Systems (NIPS)*. Lake Tahoe, Nevada. December 2012.
	Michael J. Paul. **Mixed membership Markov models for unsupervised conversation modeling.** In the *2012 Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2012)*, Jeju, Korea. July 2012.
	Michael J. Paul and Jason Eisner. **Implicitly intersecting weighted automata using dual decomposition.** In the *2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2012)*, Montreal, Canada. June 2012.
2011	Michael J. Paul and Mark Dredze. **You are what you Tweet: Analyzing Twitter for public health.** In the *5th International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*. July 2011.
	Delip Rao, Michael Paul, Clayton Fink, David Yarowsky, Timothy Oates, Glen Coppersmith. **Hierarchical Bayesian models for latent attribute detection in social networks.** In the *5th International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*. July 2011.
2010	Michael J. Paul, ChengXiang Zhai and Roxana Girju. **Summarizing contrastive viewpoints in opinionated text.** In the proceedings of the *2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010)*, Cambridge, Massachusetts. October 2010.
	Michael Paul and Roxana Girju. **Comparative scientific research analysis with a language-independent cross-collection model.** In the proceedings of *XXV Congresso de la Sociedad Española para el Procesamiento del Lenguaje Natural* (SEPLN-2010), Valencia, Spain, September 2010.
	Michael Paul and Roxana Girju. **A two-dimensional topic-aspect model for discovering multi-faceted topics.** In the proceedings of the *Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI 2010)*, Atlanta, Georgia, July 2010.
2009	Michael Paul and Roxana Girju. **Topic modeling of research fields: an interdisciplinary perspective.** In the proceedings of *Recent Advances in Natural Language Processing (RANLP 2009).* September 2009.

Michael Paul and Roxana Girju. **Cross-cultural analysis of blogs and forums with mixed-collection topic models.** *Empirical Methods in Natural Language Processing (EMNLP 2009)*, Singapore. August 2009.

Michael Paul, Roxana Girju, Chen Li. **Mining the Web for reciprocal relationships.** In the proceedings of the *13th Conference on Computational Natural Language Learning (CoNLL 2009)*, Boulder, Colorado. June 2009.

PEER-REVIEWED WORKSHOP PUBLICATIONS

2015
Michael J. Paul, Mark Dredze, David A. Broniatowski, Nicholas Generous. **Worldwide influenza surveillance through Twitter.** *AAAI Workshop on the World Wide Web and Public Health Intelligence*, Austin, Texas. January 2015.

Byron Wallace, Michael J. Paul, Noémie Elhadad. **What predicts media coverage of health science articles?** *AAAI Workshop on the World Wide Web and Public Health Intelligence*, Austin, Texas. January 2015.

2014
Shiliang Wang, Michael J. Paul, Mark Dredze. **Exploring health topics in Chinese social media: an analysis of Sina Weibo.** *AAAI Workshop on the World Wide Web and Public Health Intelligence*, Quebec City. July 2014.

Mark Dredze, Renyuan Cheng, Michael Paul, David Broniatowski. **HealthTweets.org: a platform for public health surveillance using Twitter.** *AAAI Workshop on the World Wide Web and Public Health Intelligence*, Quebec City. July 2014.

2013
Michael J. Paul, Byron C. Wallace, Mark Dredze. **What affects patient (dis)satisfaction? Analyzing online doctor ratings with a joint topic-sentiment model.** *AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI)*, Bellevue, WA. July 2013.

Mark Dredze, Michael J. Paul, Shane Bergsma, Hieu Tran. **Carmen: a Twitter geolocation system with applications to public health.** *AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI)*, Bellevue, WA. July 2013.

2012
Michael J. Paul and Mark Dredze. **Experimenting with drugs (and topic models): Multi-dimensional exploration of recreational drug discussions.** In submission to the *AAAI 2012 Fall Symposium on Information Retrieval and Knowledge Discovery in Biomedical Text*, Arlington, VA. November 2012.

Alex Lamb, Michael J. Paul, Mark Dredze. **Investigating Twitter as a source for studying behavioral responses to epidemics.** In submission to the *AAAI 2012 Fall Symposium on Information Retrieval and Knowledge Discovery in Biomedical Text*, Arlington, VA. November 2012.

Atul Nakhasi, Ralph J. Passarella, Sarah G. Bell, Michael J. Paul, Mark Dredze, Peter J. Pronovost. **Malpractice and malcontent: Analyzing medical complaints in Twitter.** In submission to the *AAAI 2012 Fall Symposium on Information Retrieval and Knowledge Discovery in Biomedical Text*, Arlington, VA. November 2012.

Ralph J. Passarella, Atul Nakhasi, Sarah G. Bell, Michael J. Paul, Peter J. Pronovost, Mark Dredze. **Twitter as a source for learning about patient safety events.** To appear in the *AMIA 2012 Annual Symposium (American Medical Informatics Association)*, Chicago, IL. November 2012.

William M. Darling, Michael J. Paul and Fei Song. **Unsupervised part-of-speech tagging in noisy and esoteric domains with a syntactic-semantic Bayesian HMM.** In the *EACL 2012 Workshop on Semantic Analysis in Social Media*, Avignon, France. April 2012.

2008   Michael Paul and Roxana Girju. **AIRTA: an automatic interdisciplinary research topic advisor.** *NSF-sponsored Symposium on Semantic Knowledge Discovery, Organization and Use* - Demo session, New York University. November 2008.

INVITED TALKS

3/02/2015   **University of Maryland, Baltimore County**   Baltimore, MD
- Host: Dr. Tim Finin and Dr. Nilanjan Banerjee
- Title: Topic modeling with structured-priors for text-driven science

2/27/2015   **University of Colorado Boulder**   Boulder, CO
- Host: Dr. Leysia Palen
- Title: Topic modeling with structured-priors for text-driven science

2/23/2015   **Rice University**   Houston, TX
- Host: Dr. Lydia E. Kavraki and Dr. Chris Jermaine
- Title: Topic modeling with structured-priors for text-driven science

2/18/2015   **University of California, Santa Barbara**   Santa Barbara, CA
- Host: Dr. Subhash Suri
- Title: Topic modeling with structured-priors for text-driven science

1/21/2015   **National Cancer Institute (NIH)**   Rockville, MD
- Host: Dr. Wen-ying Chou
- Title: Making sense of the web for public health using NLP

12/03/2014   **University of Maryland at College Park**   College Park, MD
- CLIP Colloquium
- Host: Dr. Jimmy Lin
- Title: Making sense of the web for public health using NLP

5/29/2014   **Cafe Scientifique**   Annapolis, MD
- Monthly science talk series for the general public
- Host: Danielle Lucid
- Title: Using social media to study public health

9/16/2013   **Association of Public Data Users**   Washington, DC
- Annual conference, Big Data session
- Host: Robert Scardamalia
- Title: Data mining social websites for public health

12/08/2010   **University of Maryland at College Park**   College Park, MD
- CLIP Colloquium
- Host: Dr. Jordan Boyd-Graber
- Title: Summarizing contrastive viewpoints in opinionated text

PROFESSIONAL EXPERIENCE

Summer 2013,
Summer 2014

Intern
Microsoft Research   Redmond, WA
Supervisors: Dr. Eric Horvitz, Dr. Ryen White
  • Mined search query and browsing logs to understand web activity
    pertaining to medical information.

October 2011
– Present

Advisory Board
Sickweather LLC   Baltimore, MD
http://www.sickweather.com
  • Provides geographic surveillance of illnesses in the United States by
    monitoring social networking sites.

Summer 2011

Intern, Relevance Team
Twitter Inc.   San Francisco, CA
Supervisor: Dr. Alek Kolcz
  • Worked on modeling user interests.

Fall 2009 –
Summer 2010

Research Programmer
University of Illinois, Dept of Computer Science   Urbana, IL
Supervisor: Dr. Dan Roth
  • Worked on the Learning Based Java modeling language and library.
  • Implemented various machine learning algorithms and helped
    develop a framework for learning with complex input and output.
  • Made efficiency improvements, increasing speed by up to 16x.

Summer 2009

Research Assistant
University of Illinois, Department of Linguistics   Urbana, IL
Advisor: Dr. Roxana Girju
  • Developed probabilistic models of reciprocal social interactions.
  • Worked on a National Endowment for the Humanities (NEH) project
    on a family of endangered Peruvian languages.

Summer 2008

Research Fellow
Univ. of Illinois, Data Sciences Summer Institute   Urbana, IL
  • My group created a search engine of UIUC courses, capable of
    displaying a course's prerequisite to/for map and automatically
    finding courses of similar content.

STUDENT ADVISING

Fall 2013 –
Current

Shiliang Wang
  • Masters student at JHU
  • Project: Studying health topics in Chinese social media

Fall 2013 –
Current

Michael Smith
  • Masters student at JHU
  • Project: Studying influenza awareness and media in social media

Spring 2012 –
Fall 2012

Alex Lamb
  • Undergraduate at JHU (now at Amazon)
  • Project: Building a classification model of influenza in Twitter

Spring 2012 –
Fall 2012

Hieu Tran
  • Undergraduate at JHU (now graduate student at MICA)
  • Project: Building a geolocation system for Twitter

ACADEMIC SERVICE

Co-chair	• PSB 2016 (Pacific Symposium on Biocomputing) session on Social Media Mining for Public Health Monitoring and Surveillance ◦ January 4-8, 2016, Big Island of Hawaii
Organizer	• Third Annual Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL) ◦ 10/12/2013, University of Maryland Baltimore County

Reviewer (Journal)	• Computer science / NLP journals:
	  ◦ Human-Computer Interaction (2015)
	  ◦ Transactions on Information Systems (TOIS) (2014)
	  ◦ Journal of Natural Language Engineering (JNLE) (2012, 2013)
	  ◦ Knowledge and Information Systems (KAIS) (2013)
	• Medical and general science journals:
	  ◦ PLOS ONE (2014, 2015)
	  ◦ Journal of the American Medical Informatics Association (2014)
	  ◦ Health Informatics Journal (2014)
	  ◦ American Journal of Preventative Medicine (2014)
	  ◦ Nicotine & Tobacco Research (2013, 2014)

Reviewer / Program Committee	• Computer science / NLP conferences:
	  ◦ 2016: AAAI
	  ◦ 2015: ACL, NAACL, ICML, KDD, EMNLP
	  ◦ 2014: ACL, EACL, ICWSM
	  ◦ 2013: NAACL, *SEM
	  ◦ 2012: NAACL
	• ACL 2014 Workshop on Entity Attribute Prediction in Social Media
	• NIPS 2013 Workshop on Topic Models

Secondary Reviewer	• Computer Science / NLP conferences: ◦ AAAI 2014, WWW 2014, NIPS 2011, NAACL 2010
Departmental Committees	• Student admissions (JHU) (2015) • Student recruiting (JHU) (2011, 2012)