Shomir Wilson

EECS Department   shomir.wilson@uc.edu
University of Cincinnati   `http://shomir.net`
PO Box 210030
Cincinnati OH 45221-0030

EDUCATION

2011	**Ph.D.**, Computer Science	University of Maryland
2008	**M.S.**, Computer Science	University of Maryland
2005	**B.S.**, Computer Science, **B.S.**, Mathematics, **B.A.**, Philosophy	Virginia Tech

RESEARCH AND TEACHING EXPERIENCE

Assistant Professor   August 2016 – present
University of Cincinnati - EECS Department   (Cincinnati, OH)
Director of the Human Language Technologies Lab
Member of the University of Cincinnati Institute for Analytics Innovation

Project Scientist   August 2015 – August 2016
Carnegie Mellon University - School of Computer Science   (Pittsburgh, PA)
Supervisor: Norman Sadeh

Lecturer   January 2015 – May 2015
Carnegie Mellon University - School of Computer Science   (Pittsburgh, PA)
Course: Natural Language Processing (11-411 / 11-611)
Co-Instructors: Chris Dyer and Alan W Black

NSF International Research Fellow   August 2014 – July 2015
Carnegie Mellon University - School of Computer Science   (Pittsburgh, PA)
Host: Alan W Black

NSF International Research Fellow   August 2013 – July 2014
University of Edinburgh - School of Informatics   (Edinburgh, United Kingdom)
Host: Jon Oberlander

Postdoctoral Fellow   September 2011 – July 2013
Carnegie Mellon University - School of Computer Science   (Pittsburgh, PA)
Supervisor: Norman Sadeh

NSF EAPSI Visiting Scholar   June – August 2010
National University of Singapore - School of Computing   (Singapore)
Host: Min-Yen Kan

NSF EAPSI Visiting Scholar   June – August 2009
Macquarie University - School of Computing   (Sydney, Australia)
Host: Robert Dale

Graduate Research Assistant   May 2006 – May 2011
University of Maryland - Computer Science Department   (College Park, MD)
Advisor: Donald Perlis

Graduate Teaching Assistant   Fall 2005 – Spring 2006
University of Maryland - Computer Science Department   (College Park, MD)
Course: Introduction to Low-Level Programming Concepts (CMSC 212)
Supervisors: Jan Plane and Peter Keleher (fall), Larry Herman (spring)

GRANTS AND RESEARCH SUPPORT

Carnegie Mellon University $36,625   2016-09-01 – 2018-02-28
Subcontract from the Usable Privacy Policy Project, an NSF SaTC Frontier Project

Ohio Supercomputer Center   (5,000 Resource Units)   2016-2017
#PES0731-1: Text Analysis for Online Privacy

National Science Foundation $98,100   2013-08-01 – 2015-03-31
#1159236 IRFP: Metalanguage Identification for Interactive Language Technologies

National Science Foundation $5,000   2010-06-13 – 2010-08-06
#1015666 EAPSI (Singapore): Parsing Metalanguage and the Use-Mention Distinction

National Science Foundation $5,000   2009-06-22 – 2010-08-16
#0914091 EAPSI (Australia): Distinguishing Use and Mention in Natural Language

AWARDS AND RECOGNITION

Best Paper Finalist at the 25th World Wide Web Conference, 2016.

National Science Foundation SEE Innovation Research Highlight: "Refining a Computer's Understanding of Language", 2012.

University of Maryland International Conference Student Support Award, 2011.

University of Maryland Block Grant Fellowship, 2005–2007.

SERVICE

NSF SaTC proposal review panelist, 2016 and 2017.

Faculty marshal for the College of Engineering and Applied Sciences, 2017.

Lead organizer of the AAAI Fall Symposium on Privacy and Language Technologies, 2016.
(https://sites.google.com/site/fsplt2016/)

Grand awards judge for the Intel International Science and Engineering Fair, 2012 and 2015.

Proposal reviewer for the Portuguese Foundation of Science and Technology (FCT), 2012.

Graduate student representative to the Computer Science Department Council, 2008-2009.

Session chair for Logic Programming and Knowledge Engineering at AAAI 2008.

<h1 style="text-align:center">PUBLICATIONS AND WRITINGS</h1>

Peer-Reviewed Conference Proceedings

1. Automated analysis of privacy requirements for mobile apps. Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven M. Bellovin, and Joel Reidenberg. In *Proceedings of the Network and Distributed System Security Symposium*, San Diego, California, March 2017.

2. The creation and analysis of a website privacy policy corpus. Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N. Cameron Russell, Thomas B. Norton, Eduard Hovy, Joel Reidenberg, and Norman Sadeh. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany, August 2016.

3. Crowdsourcing annotations of websites' privacy policies: Can it really work? Shomir Wilson, Florian Schaub, Rohan Ramanath, Norman Sadeh, Fei Liu, Noah Smith and Frederick Liu. In *Proceedings of the 25th International World Wide Web Conference*, Montreal, Canada, April 2016. **Best Paper Finalist.**

4. This table is different: A WordNet-based approach to identifying references to document entities. Shomir Wilson, Alan W Black, and Jon Oberlander. In *Proceedings of the 8th International Global WordNet Conference*, Bucharest, Romania, January 2016.

5. Identifying relevant text fragments to help crowdsource privacy policy annotations. Rohan Ramanath, Florian Schaub, Shomir Wilson, Fei Liu, Norman Sadeh, and Noah Smith. In *Proceedings of the Second AAAI Conference on Human Computation and Crowdsourcing*, works-in-progress track, Pittsburgh, PA, November 2014.

6. Determiner-established deixis to communicative artifacts in pedagogical text. Shomir Wilson and Jon Oberlander. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, MD, June 2014.

7. Toward automatic processing of English metalanguage. Shomir Wilson. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan, October 2013.

8. Privacy manipulation and acclimation in a location sharing application. Shomir Wilson, Justin Cranshaw, Norman Sadeh, Alessandro Acquisti, Lorrie Cranor, Jay Springfield, Sae Young Jeong, and Arun Balasubramanian. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Zurich, Switzerland, September 2013.

9. Tweets are forever: A large-scale quantitative analysis of deleted tweets. Hazim Almuhimedi, Shomir Wilson, Bin Liu, Norman Sadeh, and Alessandro Acquisti. In *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work*, San Antonio, TX, February 2013.

10. The creation of a corpus of English metalanguage. Shomir Wilson. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, South Korea, July 2012.

11. Application of MCL in a dialog agent. Darsana Josyula, Scott Fults, Michael L. Anderson, Shomir Wilson, and Don Perlis. In *Papers from the Third Language and Technology Conference*, Poznań, Poland, October 2007.

12. Toward domain-neutral human-level metacognition. Michael L. Anderson, Matt Schmill, Tim Oates, Don Perlis, Darsana Josyula, Dean Wright, and Shomir Wilson. In *Papers from the 2007 AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, Palo Alto, CA, March 2007.

Peer-Reviewed Symposium Proceedings

1. Analyzing vocabulary intersections of expert annotations and topic models for data practices in privacy policies. Frederick Liu, Shomir Wilson, Florian Schaub and Norman Sadeh. In *Proceedings of the AAAI Fall Symposium on Privacy and Language Technologies*, Arlington, VA, November 2016.

2. Automatic extraction of opt-out choices from privacy policies. Kanthashree Mysore Sathyendra, Florian Schaub, Shomir Wilson and Norman Sadeh. In *Proceedings of the AAAI Fall Symposium on Privacy and Language Technologies*, Arlington, VA, November 2016.

3. Analyzing and predicting privacy law compliance of mobile apps. Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven M. Bellovin and Joel Reidenberg. In *Proceedings of the AAAI Fall Symposium on Privacy and Language Technologies*, Arlington, VA, November 2016.

4. The Metacognitive Loop: An architecture for building robust intelligent systems. Hamid Haidarian, Wikum Dinalankara, Scott Fults, Shomir Wilson, Don Perlis, Matt Schmill, Tim Oates, Darsana Josyula, and Michael Anderson. In *Proceedings of the AAAI Fall Symposium on Commonsense Knowledge*, Arlington, VA, November 2010.

Peer-Reviewed Journal Articles

1. (Forthcoming) Nudges for privacy and security: Understanding and assisting users' choices online. Pedro Giovanni Leon, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Saranga Komanduri, Florian Schaub, Manya Sleeper, Yang Wang, Shomir Wilson, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. Accepted to appear in *ACM Computing Surveys*.

2. PrivOnto: A semantic framework for the analysis of privacy policies. Alessandro Oltramari, Dhivya Piraviperumal, Florian Schaub, Shomir Wilson, Sushain Cherivirala, Thomas B. Norton, N. Cameron Russell, Peter Story, Joel Reidenberg, and Norman Sadeh. *Semantic Web*, 2017.

3. In search of the use-mention distinction and its impact on language processing tasks. Shomir Wilson. *The International Journal of Computational Linguistics and Applications* 2(1-2), pp. 139-154, 2011.

Book Chapters

1. (Forthcoming) A bridge from the use-mention distinction to natural language processing. Shomir Wilson. In Saka, P., Johnson, M. (Ed.), *The Semantics and Pragmatics of Quotation*. Springer, 2017.

2. The metacognitive loop and reasoning about anomalies. Matthew Schmill, Michael L. Anderson, Scott Fults, Darsana Josyula, Tim Oates, Donald Perlis, Hamid Haidarian Shahri, Shomir Wilson, and Dean Wright. In Cox, M., Raja, A. (Ed.), *Metareasoning: Thinking About Thinking.* MIT Press, MA, 2010.

Magazine Article

- A self-help guide for autonomous systems. Michael L. Anderson, Scott Fults, Darsana P. Josyula, Tim Oates, Don Perlis, Matthew D. Schmill, Shomir Wilson, and Dean Wright. *AI Magazine*, Summer 2008.

Peer-Reviewed Conference Poster Abstracts

1. Visualization and interactive exploration of data practices in privacy policies. Sushain K. Cherivirala, Florian Schaub, Mads Schaarup Andersen, Shomir Wilson, Norman Sadeh, and Joel R. Reidenberg. In *Proceedings of the Twelfth Symposium on Usable Privacy and Security*, Denver, CO, 2016.

2. Towards usable privacy policies: Semi-automatically extracting data practices from websites' privacy policies. Norman Sadeh, Alessandro Acquisti, Travis Breaux, Lorrie Cranor, Aleecia McDonald, Joel Reidenberg, Noah Smith, Fei Liu, N. Cameron Russell, Florian Schaub, Shomir Wilson, James Graves, Pedro Leon, Rohan Ramanath, and Ashwini Rao. In *Proceedings of the Tenth Symposium on Usable Privacy and Security*, Palo Alto, CA, July 2014.

Peer-Reviewed Workshop Proceedings

1. Demystifying Privacy Policies with Language Technologies: Progress and Challenges. Shomir Wilson, Florian Schaub, Aswarth Dara, Sushain K. Cherivirala, Sebastian Zimmeck, Mads Schaarup Andersen, Pedro Giovanni Leon, Eduard Hovy, and Norman Sadeh. In *Proceedings of the Workshop on Text Analytics for Cybersecurity and Online Safety (TA-COS) at LREC*, Portoroz, Solvenia, May 2016.

2. Distinguishing use and mention in natural language. Shomir Wilson. In *Proceedings of the NAACL HLT Student Research Workshop*, Los Angeles, CA, June 2010.

3. The role of metacognition in robust AI systems. Matt Schmill, Tim Oates, Michael L. Anderson, Darsana Josyula, Don Perlis, Shomir Wilson, and Scott Fults. In *Papers from the Workshop on Metareasoning at the 23rd AAAI Conference on Artificial Intelligence*, Chicago, IL, July 2008.

4. Ontologies for reasoning about failures in AI systems. Michael L. Anderson, Scott Fults, Darsana Josyula, Tim Oates, Don Perlis, Matt Schmill, and Shomir Wilson. In *Proceedings of the First International Workshop on Metareasoning in Agent-Based Systems*, Honolulu, HI, 2007.

Dissertation

- *A Computational Theory of the Use-Mention Distinction in Natural Language.* Shomir Wilson. University of Maryland, 2011.

Technical Reports

1. The Usable Privacy Policy Project: Combining crowdsourcing, machine learning and natural language processing to semi-automatically answer those privacy questions users care about. Norman Sadeh, Alessandro Acquisti, Travis Breaux, Lorrie Cranor, Aleecia McDonald, Joel Reidenberg, Noah Smith, Fei Liu, N. Cameron Russell, Florian Schaub, and Shomir Wilson. Technical Report CMU-ISR-13-119, Carnegie Mellon University, 2013.

2. Automatic categorization of privacy policies: A pilot study. Waleed Ammar, Shomir Wilson, Norman Sadeh, and Noah A. Smith. Technical Report CMU-LTI-12-019 / CMU-ISR-12-114, Carnegie Mellon University, 2012.

INVITED TALKS

Northern Kentucky University College of Informatics, 24 March 2017.

North Carolina State University Computer Science Department, 15 March 2017.

"Text Analysis to Support the Privacy of Internet Users", Hutton Lecture Series, Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, 17 Feburary 2017.

University of Dayton Computer Science Department Colloquium Series, 3 February 2017.

Georgetown University Computer Science Colloquium Series. 21 November 2016.

"Crowdsourcing Annotations of Websites' Privacy Policies: Can It Really Work?", Encore Track at the Fourth AAAI Conference on Human Computation and Crowdsourcing, Austin, Texas. 1 November 2016.

"Introspective Users and Introspective Text: Some Recent Results", CHIME Text Seminar, National University of Singapore. 5 January 2016.

"Identifying Deixis to Communicative Artifacts in Text", NLIP Seminar Series, Cambridge University. 9 May 2014.

"An Empirical Approach to Metalanguage", ILCC/HCRC Seminar Series, University of Edinburgh. 6 September 2013.

"A Computational Approach to Metalanguage and the Use-Mention Distinction", CL+NLP Lunch, Carnegie Mellon University. Pittsburgh, PA, 23 April 2013.

"Distinguishing Use and Mention in Natural Language", Centre for Language Technology Seminar Series, Macquarie University. Sydney, Australia, 2009.

MISCELLANY

Safe Zone training, 2008 and 2017.

Bias Busters training, 2015.

Licensed amateur radio operator, technician class.