

Catherine Lesley Blake

Last Updated Dec 22,2021

I. Personal History and Professional Experience

A. Educational Background

University of Wollongong, Australia, B.S., Computer Science, 1994.

University of Wollongong, Australia, M.S., Computer Science, 1995.

University of California, Irvine, M.S., Information and Computer Science, 1999.

University of California, Irvine, Ph.D., Information and Computer Science, 2003.

B. List of Academic Positions since Final Degree

Assistant Professor, School of Information and Library Science, University of North Carolina, Chapel Hill, 1/2004-7/2009.

Associate Professor (with tenure), School of Information Sciences, University of Illinois at Urbana-Champaign (UIUC), 8/2009-7/2021.

Program Director, Master of Information Management and Master of Bioinformatics, 2018 – 2021

Associate Director, Center for Informatics Research in Science & Scholarship (CIRSS), School of Information Sciences, University of Illinois at Urbana-Champaign (UIUC), 8/2011–2021.

Full Professor (with tenure), School of Information Sciences, University of Illinois at Urbana-Champaign (UIUC), 8/2021.

Faculty Affiliate, Medical Information Science, UIUC, 2010 – 2020.

Faculty Affiliate, Department of Computer Science, UIUC, 2009 – 2012; 2016 – present

Faculty Affiliate, National Center for Supercomputing Applications, UIUC, 2019 – present.

Faculty Affiliate, Personalized Nutrition Initiative, UIUC, 2021 – present

C. Other Professional Employment

Applications Programmer Analyst, The Broken Hill Proprietary Company Limited Information Technology (BHP Information Technology), Wollongong, New South Wales, Australia, 1990 – 1995.

Research Scientist, The Broken Hill Proprietary Company Limited Research (BHP Research), Victoria, Australia, 1995 – 1998.

D. Honors, Recognitions, and Outstanding Achievements

Faculty Fellowship, Renaissance Computing Institute (RENCI), 2007.

John Wiley & Sons Best Paper Award(s), Journal of the American Society for Information Science and Technology, 2007.

Faculty Fellowship, Environmental Change Institute, University of Illinois at Urbana-Champaign (UIUC), 2010-2011.

Faculty Fellow, Lister Hill Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, 9/2016 – 8/2017.

Centennial Scholar, School of Information Sciences, UIUC, 2017-2018.

E. Invited Lectures and Invited Conference Presentations

“Information Synthesis”, Invited speaker at the School of Information and Library Science, University of North Carolina at Chapel Hill, 2002.

“Information Synthesis”, Invited speaker at the Department of Computer Science, University of Missouri, 2002.

“Information Synthesis”, Invited speaker at the School of Library & Information Science and the College of Nursing, University of Iowa, 2002.

“Information Synthesis”, Invited speaker at the Department of Chemistry, University of North Carolina at Chapel Hill, 2002.

“Information Synthesis”, Invited speaker at the Department of Emergency Medicine, University of North Carolina at Chapel Hill, 2004.

“Text Mining”, Invited speaker at the Department of Pharmacy, University of North Carolina at Chapel Hill, 2004.

“Text Mining”, Invited speaker at the Department of Computer Science, University of North Carolina at Chapel Hill, 2004.

“Text Mining”, Invited speaker at the School of Biology, University of North Carolina at Greensboro, 2006.

“Text Mining”, Invited panelist at the XXVII Annual Charleston Conference, Nov 2007.

“Information Synthesis”, Invited Speaker at Arizona State University, School of Computing and Informatics, March 2008.

“Uncovering the Creative Processes in Chemistry: Implications for Discovery System Design”, Symposium on Computational Approaches to Creativity in Science, Stanford, March 2008.

“Text Mining Scientific Literature”, Guest panelist on Prospects for Science on the Open Science Grid at the OSG All Hands Consortium Meeting, Chapel Hill, March 2008.

“Information Synthesis”, Invited Speaker at the School of Information University of Michigan, April 2008.

“Reinventing Science Librarianship – Education for New Roles”, Invited Speaker at the Reinventing Science Librarianship: Models for the Future forum, co-sponsored by the ARL and the Coalition for Networked Information, October 2008.

“Claim Jumping through Scientific Literature”, Invited Speaker at the Renaissance Computing Institute for Glaxo Smith Cline, October 2008.

“Evidence-Based Discovery”, Invited Speaker at the University of Illinois, Graduate School of Library and Information Science, December 2008.

“Information Synthesis”, Invited Speaker. Accountability through Algorithm: Developing the Field of Computational Journalism organized by James T. Hamilton and Fred Turner, and hosted by the Center for Advanced Study in the Behavioral Sciences, Summer 2009.

“Towards Evidence-Based Discovery”, Environmental Change Institute Symposium, November 2010.

“Towards Evidence-Based Discovery”, Graduate School of Library and Information Science (GSLIS), American Society for Information Science & Technology Student Chapter, Jan 2011.

“Beyond Genes, Proteins, and Abstracts: Identifying Scientific Claims from Full-text Biomedical Articles”, Invited Speaker at Computational Bioscience Program at the University of Colorado Denver, July 2011.

“How Informatics Can Help to Address Grand Challenges in Health and the Environment”, Illinois Informatics Institute Spring Lecture Series, February 2012.

“Informatics Methods to Scale Up Risk Assessments”, Invited Speaker at Environmental Protection Agency (EPA), March 2012.

“How Informatics Can Help to Address Grand Challenges in Health and the Environment”, Bioinformatics and Medical Library IT Interns, Undergraduate Library, June 2012.

“Claim Jumping: Bridging Disciplinary Boundaries Using the Claim Framework”, Making Sense of Biological Systems: Using Knowledge Mining to Improve and Validate Models of Living Systems, Bozeman Montana, August 23-25, 2012.

“Using Secondary Information to Inform Evidence-Based Discovery”, Health Information Technology Center Workshop, University of Illinois, NCSA November 2, 2012.

“Living on the Edge of Science: A Pilot Study of Concept Formation in Toxicology Literature”, Invited Speaker at the Symposium on Cognitive Systems and Discovery Informatics, Carnegie Mellon University's Silicon Valley Campus, San Jose, CA, June 21-22, 2013.

“Representing Biological Responses to Chemical Exposures Using the Claim Framework”, Invited Speaker at University of Wisconsin, Milwaukee, March 2014.

“Collaborative Data Analytics”, Invited speaker, Department of Computer and Systems Sciences, Stockholm University. May 2014.

“Summarizing Findings with the Claim Framework”, Department of Computing and Information Systems, University of Melbourne, July 2015.

“Using the Claim Framework to Support Evidence Synthesis and Discovery”, Brown Bag Lecture Lister Hill National Center for Biomedical Communications, National Library of Medicine, May 2017.

“Mining Literature to Support Evidence-based Medicine”, Share the Vision, University of Illinois, September 2017.

“Research Trends in Digital Scholarship and Digital Humanities”, University of Illinois, June 2018.

“Human Centered Data Science, International Incubator Session #3”, Dirndorfer Anderson, T., Blake, C., Hutchinson, B., and Fraher, R. *82nd ASIS&T Annual Meeting*, Melbourne, Australia, Oct 19-23, 2019.

“International Incubator Round up, International Incubator Session”, Blake, C., and Brown, C. (2019) at *82nd ASIS&T Annual Meeting Information*, Melbourne Australia, Oct 19-23, 2019.

“Data Science at the iSchool, University of Illinois Urbana-Champaign” Blake, C. 2020 Data Science Leadership Summit, The Academic Data Science Alliance, Online, October 12-14th.

F. Offices Held in Professional Societies

Chair Elect 2008-2009, American Society for Information Science and Technology (ASIST), SIG-MED

Chair 2009-2010, American Society for Information Science and Technology (ASIST), SIG-MED (now SIG-HEALTH)

Faculty mentor, ASIS&T Doctoral Colloquium (2015)

Faculty mentor, iConference Doctoral Colloquium (2012, 2015)

Committee Member 2016 & 2017, Chair 2018, ASIS&T Research in Information Science Award Committee.

G. Editorships of Journals or Other Learned Publications

Editorial board of the *Journal of Biomedical Discovery and Collaboration* (2005-2010)

H. Grants Received

Wanda Pratt and Tammy Tengs (PI), Catherine Blake; University of California, Irvine Multi-Investigator Grant; June 2001 – June 2002; *Using Scientific Text to Identify Breast Cancer Risk Factors*.

Catherine Blake (PI); California Breast Cancer Research Program Dissertation Award; 2002-2004 \$29,136; *Using Scientific Text to Identify Breast Cancer Risk-Factors*.

Gary Marchionini, Paul Solomon and Cathy Marshall (PI); Microsoft Research; January 2004–June 2005; \$100,000 (Blake, Summer salary); *Annotation Project*.

John Smith, Paul Jones, John Reuning, and Catherine Blake (PI); IBM SUR Grant; August 2004 – June 2006; \$17,115 (Blake's portion only); *Text Mining Biomedical Literature*.

Catherine Blake (PI); The Center for Environmentally Responsible Solvents and Processes, STC Program of the National Science Foundation under agreement CHE-9876674; August 2004 – June 2006; \$68,652; *Text Mining Chemistry Literature*.

Catherine Blake (PI); UNC Junior Faculty Award; January 2007 – December 2007; \$7,500; *An Online Text Mining Evaluation Environment (OnTeMEE)*.

Catherine Blake (PI), Sue West (Co-PI); Computing Research Association Multidisciplinary Research Opportunity for Women (MRO-W) Program; Sept 2007 – August 2008 \$22,500; *Automated Information Extraction from Electronic Health Records*.

Catherine Blake (PI); Renaissance Computing Institute (RENCI) Faculty Fellows Program; July 2007 – June 2008; \$50,000 + in-kind programmer support; *Claim Jumping through Scientific Literature*.

Etta Pisano (formally Paul Watkins) (PI); National Institutes of Health; June 2008 – May 2018; \$61,000,000 (Blake, Research assistant support 2008-9); *Clinical and Translational Science Award (CTSA)*.

Helen Tibbo (PI); Catherine Blake, Cal Lee, Gary Marchionini, Jeffrey Pomerantz, Jane Greenberg, Javed Mostafa, Diane Kelly (Co-I); Institute of Museum and Library Services; September 2008 – August 2012; \$878,634 (Blake, 2% effort); *DigCCurr II: Extending an International Digital Curation Curriculum to Doctoral Students and Practitioners*.

David Richardson (PI); Gillings Innovation Labs; August 2008 – July 2010; \$595,000 (Blake, 10% effort); *Linking Data Collected by Hospitals and Ambulance Services*.

Jon Gant (PI), Catherine Blake, Michael Twidale (Co-PIs); State Farm Insurance Companies (#2010-06961); 1 August 2010 – 31 January 2011; \$44,295; *The Future of IT for State Farm*.

Catherine Blake (PI); Environmental Change Institute, UIUC (Faculty Fellowship program); April 2010 – May 2011; \$10,000; *Evidence-Based Discovery in Environmental Literature*.

Catherine Blake (PI); National Science Foundation, Division of Information and Intelligent Systems (IIS), Computer & Information Science & Engineering (CISE), #0812522; September 2008 – August 2011 (No cost extension to August 2012); \$449,317; *Towards Evidence-Based Discovery*.

Michael Twidale (PI), Catherine Blake (Co-PI); University of Illinois; January 2014 – July 2014; \$7,000; *Collaborative Information Seeking and Data Use*.

Jon Gant (PI), Illinois Extension (Co-PI), Catherine Blake (Team Member); University of Illinois; September 15, 2014 – August 14, 2016; \$300,000; *Digital Innovation Leadership*.

Catherine Blake (PI); Institute for Museum and Library Services (IMLS) RE-05-12-0054-12; July 1, 2012 – June 30, 2015 (No cost extension to June 30, 2016); \$712,338 (Award: \$498,777; Matching: \$213,561); *Developing a Model for Sociotechnical Data Analytics (SODA) Education*.

Catherine Blake and Maria Souden (PI), Michael Twidale (Co-PI); US Department of Veterans Affairs; August 15, 2014 – June 30, 2018; \$204,324; *Information Forum Requirements to Promote Knowledge Capture, Knowledge Sharing, and Community Interaction on the VHA Data Portal (Phases I and II)*.

Catherine Blake (PI); National Library of Medicine (Lister Hill Center for Biomedical Communications); Sept 19, 2016-August 15, 2017; *Medical Informatics Research Program for Visiting Scientists*.

Catherine Blake and Maria Souden (Veterans Affairs) (PIs); US Department of Veterans Affairs; July 1, 2018 – June 30, 2019 (Extended to Jan 17, 2020); \$164,300; *Information Forum Requirements to Promote Knowledge Capture, Knowledge Sharing, and Community Interaction on the VHA Data Portal (Phase III)*.

Zahra Mohaghegh (PI), Cheri Ostroff and Catherine Blake (Co-PIs); National Science Foundation (SES Division of Social and Economic Sciences and SBE Directorate for Social, Behavioral & Economic Sciences); May 5, 2015 – May 4, 2020 (No cost extension to August 2020); \$899,663; *A Big Data-Theoretic Approach to Quantify Organizational Failure Mechanisms in Probabilistic Risk Assessment (Award #1535167)*.

Catherine Blake (PI, Co-PI June 2019-May 2020), William Gropp (Co-PI, PI June 2019-May 2020), Elif Ertekin (Co-PI); National Science Foundation; June 1, 2019 – May 31, 2023; \$3,963,137 (UIUC \$2,883,274); *BD Hubs: Collaborative Proposal: Midwest: Midwest Big Data Hub: Building Communities to Harness the Data Revolution. (Award #1916613)*

Rachel Magee (PI), Catherine Blake (Co-PI), Jana Diesner (Co-PI) Jennifer McCaffrey (Collaborator) Patricia McGlaughlin (Collaborator), Susan Sloop (Collaborator); UIUC Extension Collaboration Grant (Award #7586); Jan 1, 2021 – Dec31, 2023; \$59,742; *Health Data Literacy Ambassadors*.

Masooda N. Bashir (PI), Catherine Blake (Co-PI), Alfred Chong & Runhuan Feng (Department of Statistics and Mathematics, Co-PI), Faye Jones (College of Law, Co-PI) *An Inclusive Evaluation of Privacy Standards in the Cloud*, CISCO Jan 2021-Dec 2022. Supports 4 doctoral students and 0.5 month for each faculty member's summer salaries.

I. Review Panels

National Science Foundation (NSF) – Intelligent Information Integration (2007)

Agency for Healthcare Research and Quality (AHRQ) Review Panel (2007)

National Institute of Environmental Health Sciences (NIEHS)/ National Institutes of Health (NIH) Review Panel (2008)

National Science Foundation (NSF) – Social Computational Systems (2011)

Portuguese Foundation for Science and Technology (FCT) (2012)

National Institute of Environmental Health Sciences (NIEHS) (2015)

External Reviewer, School of Library and Information Studies, The University of Oklahoma (Promotion) 2018

External Reviewer, Department of Information Science, Drexel University (Promotion) 2018

External Reviewer, School of Information Science, University of Kentucky (Promotion) 2018

National Science Foundation (NSF), Division of Materials Research (Big Data expert) 2021.

II. Publications and Creative Works

@ derived from thesis; *peer reviewed; + invited. Students supervised: +Ph.D or PostDoc. #master's

A. Doctoral thesis title

Information Synthesis: A Mixed-Initiative Meta-Analytic Approach to Facilitate Knowledge Discovery from Scientific Text

B. Books Authored or Co-Authored (in print or accepted)

Not applicable

C. Books Edited or Co-Edited (in print or accepted)

1. Grove, A., Sonnenwald, D.H., Harrison, L., Blake, C., Schlögl, C., Peters, I., Endler-Jobst, B., Cool, C., and Theng, Y. (Eds.) (2016). *Proceedings of the 79th ASIS&T Annual Meeting*, Copenhagen, Denmark. New York: Wiley. ISBN: 0-87715-548-8.
<https://asistdl.onlinelibrary.wiley.com/toc/23739231/2016/53/1>
2. Blake, C., and Brown, C. (Eds.) (2019). *Proceedings of the 82nd ASIS&T Annual Meeting*, Melbourne Australia, Volume 56. Somerset, NJ: Wiley. ISBN: 978-0-578-59118-6.
<https://asistdl.onlinelibrary.wiley.com/toc/23739231/2019/56/1>

D. Chapters in Books (in print or accepted)

1. +Blake, C. (2010). Text Mining. In *Annual Review of Information Science and Technology*, 45(1): 123-155. <http://onlinelibrary.wiley.com/doi/10.1002/aris.2011.1440450110/full>
2. Stanton, J., Palmer, C.L., Blake, C., and Allard, S. (2012) Chapter 6 - Interdisciplinary Data Science Education. In Xiao, N., and McEwen, LR. (Eds.) *Special Issues in Data Management, American Chemical Society*, 97-113. <http://pubs.acs.org/doi/abs/10.1021/bk-2012-1110.ch006>
3. + Blake, C. (2015). Biological Responses to Chemical Exposure: Case studies in how to manage ostensible inconsistencies using the Claim Framework. In *Inconsistency Robustness (Studies in Logic)*. Edited by Carl Hewitt and John Woods, Part 3, Chapter 6, 10 pages. May 20, 2015. <http://www.amazon.com/Inconsistency-Robustness-Studies-Logic-Hewitt/dp/1848901593>

E. Monographs (in print or accepted)

Not applicable

F. Articles in Journals (in print or accepted)

1. * @Blake, C., and Pratt, W. (2006). Collaborative Information Synthesis I: A Model of information behaviors of scientists in Medicine and Public Health. *Journal of the American Society for Information Science and Technology*, 57(13):1740-1749. **JASIST Best Paper Award**
<https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20487>
2. * @Blake, C., and Pratt, W. (2006). Collaborative Information Synthesis II: Recommendations for information systems that support synthesis activities. *Journal of the American Society for Information Science and Technology*, 57(14):1888-1895. **JASIST Best Paper Award**
<https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20486>
3. * West, S.L., Blake, C., Liu, Z., McKoy, J.N, Oertel, M.D., and Carey, T.S. (2009). Reflections on Using Electronic Health Record Data for Clinical Research. *Health Informatics Journal*, 15(2):108-121. <http://jhi.sagepub.com/content/15/2/108.full.pdf+html>
4. * Blake, C. (2010). Beyond Genes, Proteins, and Abstracts: Identifying scientific claims from full-text biomedical articles. *Journal of Biomedical Informatics*, 43(2):173-189.

<http://www.sciencedirect.com/science/article/pii/S1532046409001476>

5. * +Lucic, A., and Blake, C. (2015). A Syntactic Characterization of Authorship Style Surrounding Proper Names. *Digital Scholarship in the Humanities*, 30(1):53-70, <http://dsh.oxfordjournals.org/content/digitalsh/30/1/53.full.pdf>
6. * Blake, C., and +Lucic, A. (2015). An Automated Approach to Identify Endpoints to Support the Systematic Review Process. *Journal of Biomedical Informatics*, 56:42-56, <http://www.sciencedirect.com/science/article/pii/S1532046415000830>
7. * +Zheng, W., and Blake, C. (2015). Using Distant Supervised Learning to Identify Protein Subcellular Localizations from Full-text Scientific Articles. *Journal of Biomedical Informatics*, 57: 134–144. <http://www.sciencedirect.com/science/article/pii/S1532046415001537>
8. * +Gabb, H.A., and Blake, C. (2016). An Informatics Approach to Cumulative Chemical Exposures from Consumer Products: A case study of asthma-associated and potential endocrine disruptors. *Environmental Health Perspectives*, 124(8):1155-1165. <http://dx.doi.org/10.1289/ehp.1510529>
9. * Rindflesch, T.C., Blake, C.L., Fiszman, M., Kilicoglu, H., Rosemblat, G., Schneider, J., and Zeiss, C.J. (2017). Informatics Support for Basic Research in Biomedicine. *The ILAR Journal (Institute for Laboratory Animal Research)*, 58(1):80-89. <https://doi.org/10.1093/ilar/ilx004>
10. * Blake, C., and Rindflesch, T. (2017). Leveraging Syntax to Better Capture the Semantics of Elliptical Coordinated Compound Noun Phrases. *Journal of Biomedical Informatics*, 72:120-131. <https://doi.org/10.1016/j.jbi.2017.07.001>
11. * Rindflesch, T.C., Blake, C., Cairelli, M.J., Fiszman, M., Zeiss, C.J., and Kilicoglu, H. (2018). Investigating the Role of Interleukin-1 Beta and Glutamate in Inflammatory Bowel Disease and Epilepsy using Discovery Browsing. *Journal of Biomedical Semantics*, 9: article 25. 14 pages. <https://doi.org/10.1186/s13326-018-0192-y>
12. * + Guo, G., Blake, C., and +Guan, Y. (2019). Evaluating Automated Entity Extraction with Respect to Drug and Non-Drug Treatment Strategies. *Journal of Biomedical Informatics*, 94. 9 pages. <https://doi.org/10.1016/j.jbi.2019.103177>
13. * Blake, C., and +Kehm, R. (2019). Comparing Breast Cancer Treatments Using Automatically Detected Surrogate and Clinically Relevant Outcomes Entities from Text. *Journal of Biomedical Informatics: X*, 1. 11 pages. <https://doi.org/10.1016/j.yjbinx.2019.100005>
14. * + Pence, J., + Farshadmanesh, P., + Kim, J., Blake, C., and Mohaghegh, Z. (2020). Data-theoretic Approach for Socio-technical Risk Analysis: Text mining licensee event reports of U.S. nuclear power plants. *Safety Science*. Special Issue: *Safety Analytics*. Vol 124. 21 pages. <https://doi.org/10.1016/j.ssci.2019.104574>
15. * Blake C, Flaws JA (2021) Using semantics to scale up evidence-based chemical risk-assessments. *PLoS ONE* 16(12): e0260712. <https://doi.org/10.1371/journal.pone.0260712>

G. Creative Works

Not applicable

H. Patents

Not applicable

I. Bulletins, Reports, or Conference Proceedings (in print or accepted)

[For published conference papers see L. Refereed Conference Papers and Presentations]

J. **Abstracts (in print or accepted)**

1. * Blake, C., Pratt, W., and Tengs, T. (2002). Automated Information Extraction and Analysis for Information Synthesis. Abstract in the *Proceedings of the Annual American Medical Informatics Association Symposium (AMIA)*, San Antonio, TX. 1 page.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2244281/>
2. * Blake, C (2004). A Text Mining Approach to Enable Detection of Candidate Risk Factors. In M. Fieschi and E.Coiera (Eds.), Abstract in *Proceedings of the 11th World Congress on Medical Informatics – Building High Performance Health Care Organizations(Medinfo)*, p1528, San Francisco, CA.
<https://pdfs.semanticscholar.org/667a/855bdd9ea8de3aa91447ebeece78637f3efb.pdf>
3. * Pratt, W., Srinivasan, P., Smalheiser, N., and Blake, C. (2004). Mining the Literature to Promote Biomedical Discoveries, Panel at *The 11th World Congress on Medical Informatics (Medinfo) – Building High Performance Health Care Organizations*, San Francisco, CA.
4. * + Luo, L., # West, D., Marchionini, G., and Blake, C. (2005). A Study of Annotations for a Consumer Health Portal. Abstract in the *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, p388, Denver, CO.
<http://dl.acm.org/citation.cfm?doid=1065385.1065491>
5. * Blake, C. (2005). The Shift from Information Retrieval to Synthesis, Panel at *The First i-Conference of the i-school Community Bringing Disciplines to Confront Grand Challenges*, State College, PA.
6. * Blake, C., and # Anderson, C. (2005). The Shift from Information Retrieval to Synthesis, Panel at the *The First i-Conference of the i-school Community Bringing Disciplines to Confront Grand Challenges*, State College, PA.
7. * Brown, C., Blake, C., # Rogers Brown, E. A., and Tenopir, C. (2006). How Chemists Are Really Finding and Using Information in Our Digital Environment, Panel, *Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology (ASIST)*, p38, Austin, TX.
8. * Wildemuth, B.M., Blake, C.L., + Spurgin, K., + Oh, S., and + Zhang, Y. (2006). Patients' Perspectives on Personal Health Records: An Assessment of Needs and Concerns, Abstract in *Critical Issues in eHealth Research Conference: Toward Quality Patient Centered Care*, Bethesda, MD.
9. * # Kraus, S., Blake, C., and West, S.L. (2007). Information Extraction from Medical Notes, In K.A. Kuhn, J.R. Warren, and T.-Y. Leong (Eds.) Abstract in the *Proceedings of the 12th World Congress on Health (Medical) Informatics– Building Sustainable Health Systems (MedInfo)*, p1662-4, Brisbane, Australia.
10. * Blake, C., Nassar, N., and # Jones, R. (2007). Using Concepts and Entailment for Passage Retrieval from Biomedical Literature, Abstract at the *2007 Microsoft eScience Workshop* at RENCI, Chapel Hill, NC.
11. * + D'Ignazio, J., Blake, C., and Sonnenwald, D. (2009). Bridging between Scientific Disciplines: Educational Strategies to Meet the Challenge of Managing Information across the Sciences, Panel at the *72nd Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, November 6-11, Vancouver, BC, Canada.
<https://asisdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2009.1450460138>
12. * Stanton, J., Palmer, C., Blake, C., Farmer, L., and Allard, S. (2012). Brainstorming Data Science @iSchools, *iConference 2012*, Toronto, Canada.

13. * Qin, J., Crowston, K. G., Lesk, M., Blake, C., and Mostafa, J., (2012). Data Science and Analytics: What Is in It for iSchools? Workshop panel at the *iConference*, Toronto, Canada.
14. * Souden, M., Blake, C., Twidale, M., # Anderson, C. and Stelmack, J. (2015). Making Sense of Big Data: Online question answering practices supporting healthcare data re-use. *8th Annual Conference on the Science of Dissemination and Implementation: Optimizing Personal and Population Health*, December 14-15, Washington D.C.
<https://academyhealth.confex.com/academyhealth/2015di/meetingapp.cgi/Paper/7867>
15. * Blake, C. (2017). Using the Claim Framework in a Mobile Environment, *HCB3 2017 Online Conference, Health Communication: Barriers, Breakthroughs, and Best Practices*, March 1-3, 2017
http://conferences.illinois.edu/HCB3/assets/docs/HCB3_2017_abstracts.pdf
16. * Souden, M., Stelmack, J., Paris, B., Khan, H., Kok, L., # Sethi, S., # Lee, J., # Aggarwal, A., Blake, C., and Twidale, M. (2019). Developing an Online Platform for a Suicide Prevention Community of Practice at the Department of Veterans Affairs. *Informatics Summit, Bioinformatics, Clinical Research, Implementation, Data Science*, March 25-29, San Francisco.
17. Blake, C., and Flaws, J. (2019) Accelerating Chemical Assessments: A case study in automatic evidence extraction from text. *Evidence Integration Workshop, Evidence Integration in Chemical Assessments: Challenges faced in developing and communicating human health effect conclusions*, Board on Environmental Studies and Toxicology, National Academies of Sciences, Washington, DC, June 3-4, 2019.

K. Book Reviews (in print or accepted)

Not applicable

L. Refereed Conference Papers and Presentations

1. * Blake, C., and Pratt, W. (2000). Multiple Categorization of Search Results: An Extension to Dynamic Categorization. In *Proceedings of the Annual American Medical Informatics Association Symposium (AMIA)*, p81-85, Los Angeles, CA.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243812/pdf/procamiasymp00003-0116.pdf>
2. * Blake, C., and Pratt, W. (2001). Better Rules Fewer Features: A Semantic Approach to Selecting Features from Text. In *Proceedings of the Institute of Electrical and Electronics Engineers Data Mining Conference (IEEE-DM)*, p59-66, San Jose, CA.
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=989501&tag=1
3. * Blake, C., and Pratt, W. (2002). A Semantic Approach to Identify Candidate Treatments from Existing Medical Literature. In Sanda M. Harabagiu and Vinay Chaudhri (Eds.) *Mining Answers from Texts and Knowledge Bases: Papers from the 2002 Spring Symposium*. Technical Report SS-02-06. American Association for Artificial Intelligence, p9-13, Menlo Park, CA.
<https://www.aaai.org/Papers/Symposia/Spring/2002/SS-02-06/SS02-06-003.pdf>
4. * Blake, C., and Pratt, W. (2002). Collaborative Information Synthesis. In Grove, A. (Ed.), In *Proceedings of the 65th Annual Conference of the American Society for Information Science and Technology (ASIST)*, p44-56, Philadelphia, PA.
<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.1450390105>
5. * Blake, C. (2002). Information Synthesis: A Process Used by Scientists in Medicine and Public Health to Overcome Information Overload. In Doctoral Forum of the Fourth International Conference on Conceptions of Library and Information Science: Emerging Frameworks and Methods (CoLIS 4), Seattle, WA.

6. * Blake, C. (2003). A Technique to Resolve Contradictory Answers. In M. Maybury (Ed.) *New Directions in Question Answering, Papers from the AAAI Spring Symposium*, Technical Report SS-03-07, American Association for Artificial Intelligence, p49-52, Stanford, CA.
<http://www.aaai.org/Library/Symposia/Spring/2003/ss03-07-008.php>
7. * Blake, C. (2005). Information Synthesis: A New Approach to Explore Secondary Information in Scientific Literature. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, p56-64, Denver, CO. (Paper acceptance rate 31%)
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4118515&tag=1
8. * Blake, C., # West, D., # Luo, L. and Marchionini, G. (2005). Cataloging On-Line Health information: A Content Analysis of the NC Health Info Portal. In *Proceedings of the Annual American Medical Informatics Association Symposium (AMIA)*, p56-60, Washington, DC. (Paper acceptance rate 37%) <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560486/>
9. * Blake, C. (2006) A Comparison of Document, Sentence and Term Event Spaces. In *Proceedings of the Joint 21st International Conference on Computational Linguistics (COLING) and the 44th Annual Meeting of the Association for Computational Linguistics (ACL)*, p601-608, Sydney, Australia. (Paper acceptance rate 22.3%) <http://dl.acm.org/citation.cfm?id=1220251>
10. * Blake, C., and # Randall, M. (2006). Scientific Discovery: A View from the Trenches, In Ljupco Todorovski, Nada Lavrac, Klaus P. Jantke (Eds.). *Lecture Notes in Computer Science, Discovery Science*, 9th International Conference, p41-52, Barcelona, Spain. (Long paper acceptance rate 27%)
http://link.springer.com/chapter/10.1007/11893318_8
11. * Blake, C., # Kampov, J., # Orphanides, A., # West, D., and # Lown, C. (2007). UNC-CH at DUC 2007: Query Expansion, Lexical Simplification, and Sentence Selection Strategies for Multi-Document Summarization, Presentation at Document Understanding Conference (DUC) 2007, 8 pages, Rochester, NY. <https://www-nlpir.nist.gov/projects/duc/pubs/2007papers/unc-ch.blake.final.pdf>
12. * Blake, C. (2007). In Support of e-science - Shifting from Information Retrieval to Information Synthesis, *Microsoft eScience Workshop* at RENCI, Chapel Hill, NC.
13. * Blake, C. (2007) The Role of Sentence Structure in Recognizing Textual Entailment. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, p101-106, Prague, Czech Republic. <http://dl.acm.org/citation.cfm?id=1654557>
14. * Blake, C. (2008). Beyond Genes, Proteins, and Abstracts: A Framework to Capture Scientific Claims, *Microsoft eScience Workshop*, Indianapolis, IN.
15. * Blake, C. (2008). Uncovering the Creative Processes in Chemistry: Implications for Discovery System Design, *Symposium on Computational Approaches to Creativity in Science*, Stanford, March 29-30.
16. * + Zheng, W., and Blake, C. (2010). Automatic Extraction of Location Relations from Text. Poster *iConference*, p573-575, Urbana-Champaign, IL, USA.
<https://www.ideals.illinois.edu/bitstream/handle/2142/15046/wuzheng.pdf?sequence=2>
17. * + Zheng, W., and Blake, C. (2010). Bootstrapping Location Relations from Text. In *Proceedings of the American Society for Information Science and Technology Annual Meeting*, Oct, p1-9. Pittsburgh, PA. <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/meet.14504701114>
18. * Blake, C., + Zheng, W., # Painter, K., and # Weyerhaeuser, W. (2010). The Role of Semantics in Recognizing Textual Entailment. In *Proceedings of the Text Analysis Conference, Recognizing Textual Entailment (RTE6)*, 6 pages, Gaithersburg, Maryland.
<http://www.nist.gov/tac/publications/2010/participant.papers/UIUC.proceedings.pdf>

19. * + Lucic, A., and Blake, C. (2011). Comparing the Similarities and Differences between Two Translations. In *Digital Humanities* (DH 2011), July, p174-176, Palo Alto, CA.
https://ecommons.luc.edu/cgi/viewcontent.cgi?article=1027&context=cs_facpubs#page=195
20. * Blake, C., and + Zheng, W. (2011). In Search of Protein Locations. In *Association for Computational Linguistics/Human Language Technology Conference (ACL/HLT), Proceedings of the Biological Natural Language Processing (BioNLP) Workshop*, p101-102, Portland OR.
<http://www.aclweb.org/anthology/W11-0212>
21. * + Hoon Park, D., and Blake, C. (2012). Identifying Comparative Sentences in Full-text Scientific Articles. In *Association for Computational Linguistics, Proceedings of the Workshop on Detecting Structure in Scholarly Discourse*, July, p1-9, Jeju, South Korea.
<http://www.aclweb.org/anthology/W12-4301>
22. * + Lucic, A., and Blake, C. (2012). Characterizing Authorship Style Using Linguistic Features. In *Digital Humanities*, July 19, 6 pages, Hamburg Germany. <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/characterizing-authorship-style-using-linguistic-features/>
23. * Blake, C. (2012). Using Secondary Information to Inform Public Policy. In *Internet, Politics, Policy, Big Data Big Challenges*, Sept 20-21, 2012 UK. <http://blogs.oii.ox.ac.uk/ipp-conference/2012/programme-2012/track-b-policy/panel-1b-using-big-data-to-inform-public/catherine-blake-using-secondary.html>
24. * + AbdelRahman, S.E., and Blake, C. (2012). UIUC_RuleBased: A Rule-based Human Interpretation System for Semantic Textual Similarity Task. In *Workshop on Semantic Evaluation (SemEval), First Joint Conference on Lexical and Computational Semantics, Collocated with NAACL-HLT* June 7-8, p536-542, Montreal, Canada
<http://dl.acm.org/citation.cfm?id=2387726>
25. * Blake, C., Stanton, J., and Saxenian, A. (2013). Filling the Workforce Gap in Data Science and Data Analytics, Workshop at the *iConference*, p1015-1016, Fort Worth TX.
<https://www.ideals.illinois.edu/bitstream/handle/2142/42501/424.pdf?sequence=4>
26. * Twidale, M., Blake, C., and Gant, J. (2013). Towards a Data Literate Citizenry. In *iConference*, p247-257, Fort Worth, TX.
<https://www.ideals.illinois.edu/bitstream/handle/2142/38385/189.pdf?sequence=4>
27. * + Ahmed, S., Blake, C., Williams, K., + Lenstra, N., and Liu, Q. (2013). Identifying Claims in Social Science Literature. Poster at *iConference*, p942-946, Fort Worth TX. **Honorable mention**
<https://www.ideals.illinois.edu/bitstream/handle/2142/42061/485.pdf?sequence=2>
28. * + Lee, J., and Blake, C. (2014). Information Integration: A case study of air quality in Chicago and St. Louis. In *Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIS&T)* October 31 – November 5, 4 pages, Seattle, WA.
<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2014.14505101154>
29. * Blake, C., and + Gabb, H. (2014). Parameter Tuning: Exposing the gap between data curation and effective data analytics. In *Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, October 31 – November 5, 4 pages, Seattle, WA.
<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2014.14505101138>
30. * # Svetlozara, S., Cohen, D., and Blake, C. (2014). Exploring Cultural Differences in Language Usage: The Case of Negation. In *Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, October 31 – November 5, 4 pages. Seattle, WA.
<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2014.14505101086>

31. * Blake, C., and + Guo, J. (2014). Information Behaviors at the Edge of Reason: The role of uncertainty, science, and culture on environmental policy. In *Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, 4 pages, Seattle, WA. <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2014.14505101164>
32. * Blake, C. (2014). Biological Responses to Chemical Exposure: Case studies in how to manage ostensible inconsistencies using the Claim Framework. In *Inconsistency Robustness'14*, Stanford University July 29-31. 10 pages. http://cirss.lis.illinois.edu/Documents/Publications_docs/Blake_2014c.pdf
33. * + Sherman, G., Blake, C., and + Lee, J. (2015). Identifying Population Characteristics from Tables in Full Text Articles. In *American Medical Informatics Association Symposium*, Nov 14 - 18, San Francisco, CA. <https://knowledge.amia.org/59310-amia-1.2741865/t005-1.2744350/f005-1.2744351/2246244-1.2744574/2249080-1.2744571?qr=1>
34. * + Lucic, A., and Blake, C. (2015). Learning User-Defined, Domain-Specific Relations: A Situated Case Study and Evaluation in Plant Science. In *Proceedings of the Annual Meeting of the Association for Information Science and Technology*, November 6-10, 12 pages, St. Louis, MO. <http://onlinelibrary.wiley.com/doi/10.1002/pr2.2015.145052010033/pdf>
35. * Blake, C., Souden, M., + Anderson, C.L., Twidale, M., and Stelmack, J.E. (2015). Online Question Answering Practices Supporting Healthcare Data Re-use. In *Proceedings of the Annual Meeting of the Association for Information Science and Technology (ASIS&T)*, November 6-10, 4 pages, St Louis, MO. <https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/pr2.2015.1450520100116>
36. * + Guo, J., Lu, Y., Mori, T., and Blake, C. (2015). Expert-Guided Contrastive Opinion Summarization for Controversial Issues. In *3rd International Workshop on Natural Language Processing for Social Media, 24th International World Wide Web Conference*, May 19th, p1105-1110, Florence.
37. * Blake, C. (2015). Evidence-based Discovery, *iConference*, March 23-27, 8 pages, Newport Beach, CA. <https://www.ideals.illinois.edu/handle/2142/73452>
38. * + Gabb, H., + Lucic, A., and Blake, C. (2015). A Method to Automatically Identify the Results from Journal Articles, *iConference*, March 23-27, 10 pages, Newport Beach, CA. <https://www.ideals.illinois.edu/handle/2142/73645>
39. * + Evans, C., Diesner, J., and Blake, C. (2015). Email Data Analysis as an Alternate Lens into Historical Events, *HASTAC Conference*, May, Lansing, MI.
40. * + Lucic, A., and Blake, C. (2016). Preparing a Workforce to Effectively Reuse Data. In *Proceedings of the Annual Meeting of the Association for Information Science and Technology*, October 14-16, p1-10. Copenhagen, Denmark. <http://dl.acm.org/citation.cfm?id=3017522>
41. * + Lucic, A., and Blake, C. (2016). Improving Endpoint Detection to Support Automated Systematic Reviews. In *American Medical Informatics Association Symposium*, Nov 12 - 16, p1900-1909, Chicago, IL. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333237/>
42. * + Gabb, H.A., and Blake, C. (2016), Factoring Near-field Chemical Exposure into Personalized Medicine. In *American Medical Informatics Association Symposium*, Nov 12 - 16, Chicago, IL.
43. * Blake, C. (2019). Tracking Breast Cancer Survivorship over Time using the Claim Framework. *Annual Symposium of the American Medical Informatics Association*, Nov. 16-20, Washington, DC.
44. * + Lee, J., Blake, C., and McInnes, B. (2019). Evaluating Ontology Coverage and Structure to Better Align Clinical Trials with Patients and vice versa. In *Proceedings of the 82nd ASIS&T*

Annual Meeting, Oct 19-23, p157-165, Melbourne Australia.
<https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pr2.14?af=R>

45. * + Kim, J., Blake, C., and + Gabb, H. (2019). Empowering Citizens to Manage Their Chemical Exposures: Step 1 - Identify Ingredients in Consumer Products. In *Proceedings of the 82nd ASIS&T Annual Meeting*, Oct 19-23, p436-440, Melbourne Australia.
<https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pr2.72>
46. * Blake, C., + Kim, J., # Mathur, M., and # Aggarwal, A. (2019). A Gap Analysis of Survivorship Terminology: Knowledge Resources versus Literature Usage. In *Proceedings of the 82nd ASIS&T Annual Meeting*, Oct 19-23, p612-614, Melbourne Australia.
<https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pr2.137>
47. * + Kim, J., Blake, C., and Darch, P. (2019). Toward Computational Reproducibility: A Doctoral Student's Story of Passing the Baton, *Proceedings of the 82nd ASIS&T Annual Meeting*, Oct 19-23, p688-690, Melbourne Australia.
<https://asistdl.onlinelibrary.wiley.com/doi/10.1002/pr2.134>
48. * Twidale, M., Blake, C. Souden, M., Stelmack, J., and + Kim, J. (2019). Repurposing: Another Essential R for Data Curation. In *Proceedings of the 82nd ASIS&T Annual Meeting*, Oct 19-23, p785-787, Melbourne Australia,
<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pr2.176?af=R>
49. + Blake, C (2020) Data Science at the iSchool, University of Illinois Urbana-Champaign. In *Data Science Leadership Summit* October 12-14,2020, Held online due to COVID
<https://academicdatascience.org/adsa-meetings/2020-data-science-leadership-summit/2020dslupdates>
50. + Yang, J., + Farshadmanesh, P., +Tatsuya, S., Pence, J., Reiani, S., Blake, C., and Mohaghegh,Z. (2020). Data-theoretic Equipping Machine Learning with Uncertainty Quantification to Update Probabilistic Risk Assessment of Nuclear Power Plants using NRC Licensee Event Reports 2020 *ANS Virtual Winter Meeting*, November 16-19, 2020
<https://www.ans.org/pubs/transactions/article-48697/>
51. # Fu, Y., Schneider,J. and Blake, C. (2021) Keystone Citations for Constructing Validity Chains among Research Papers. *1st International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment (Sci-K 2021)*, Co-located with The Web Conf 2021, April 13, 2021, pp 451-455. <https://dl.acm.org/doi/10.1145/3442442.3451368>
52. Keefer, D. and Blake, C. (2021) The Reproducible Data Reuse (ReDaR) Framework to Capture and Assess Multiple Data Streams. Submitted to *Proceedings of the 84th ASIS&T Annual Meeting*, Oct 30-Nov 2, Salt Lake City, Utah,
<https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/pr2.451>

M. Other - Research Activities Without Peer Review

1. Blake, C. (1996). Data Access and Data Integrity in a Geographically Distributed Environment- 'Global Information Warehouse'. *BHP TechNote* BHPR/CP/N/029, Melbourne Research Laboratories, Australia.
2. Blake, C. (1996). An Overview of Quinlan's C4.5 Algorithm. *BHP TechNote* BHPR/CP/N/015, Melbourne Research Laboratories, Australia.
3. Blake, C., and Merz, C. (1998). UCI Repository of Machine Learning Databases, Department of Information and Computer Science, University of California at Irvine, CA.

4. Blake, C., Keogh, E., and Merz, C. J. (1999). UCI Repository of Machine Learning Databases. Department of Information and Computer Science, University of California at Irvine, CA.
5. West, S., Liu, Z., McKoy, N., Oertel, M., Blake, C., Schwartz, B., Ochart, F., and Carey, T. (2007). Use of Electronic Medical Records and Administrative Claims Data for Assessing Type 2 Diabetes Care, *UNC DEcIDE project report for Grant AHRQ 290-05-0040-1*.
6. Blake, C., Gasser, L., and Downie, S. (2011). Institutional Organizations that Foster Interdisciplinary Education and Research, *Innovation Summit*, University of Illinois, IL.
7. Surbeck, E. and Blake, C. (2014) The Impact of Ground Ozone on Asthma: A Case Study using Project Indicator in Champaign-Urbana Poster at the *Graduate School of Library and Information Science Research Showcase*, University of Illinois, IL, February 2014.
8. Jenna Kim, Catherine Blake, Harsh Baberwal, Ben Chen, Xinyuan Chen, Xueying Cui, Siyu Duan, Anisha Jauhari, Ankit Katariya, Ziyi Kou, Haesook Ma, Shufan Ming, Baisheng Qiu, Alex Rosenberger, Lanyu Shang, Xiaolei Shao, Nikhil Satish Shenoy, Michael Steffen, Shraddha Tiwari, Qianqian Wang, Yanmeng Xin, Yash Yamsanwar (2021) Using Text Mining to Solve Real-world Problems: Student Projects for Fall 2021, Poster at the *School of Information Sciences Research Showcase*, University of Illinois at Urbana-Champaign, Champaign, IL, USA

III. Resident Instruction

Mentoring and outreach activities

My strategy to mentor students varies greatly on the individual being mentored. For MS students in my text mining course, I require that they first submit a dream statement where they share their ideal project goals and identify 2 text collections that might help them achieve those goals. We then work together to transform their passion into a project that can be addressed using a supervised learning approach. For doctoral students, it is important that students leave the program with the ability to discern between the challenges that can be overcome and the challenges that are indicators that the current path will not be productive. Several of my doctoral students come to the program with 20 years of industry experience and I help them funnel that expertise into a form that is conducive to academic research.

I certainly was not the only woman in my computer science undergraduate program, nor the only female research scientist who was working for a mining and heavy manufacturing company, but those experiences have enabled me to help junior faculty by having them focus on the message on not get distracted from the way in which the message was delivered. This helps to manage interaction styles and feedback from reviewers.

I have published work that advocates for a data literate citizenry[L26] and I am currently collaborating with colleagues and with UIUC extension to create a health data literacy ambassador program. This program reaches students at the point where they are most likely to turn away from STEM careers. We submitted a proposal with 2 junior faculty from UIUC and 2 faculty from North Carolina State University that leveraged libraries to create a network of data literacy ambassadors. Reviewers were highly polarized, but the grant was ultimately not funded. We are currently pursuing alternative strategies to move that work forward.

The Midwest Big Data Hub has also enabled me to reach communities such as the Tribal Nations who are less represented in data science. I see those activities as a long term and sustained commitments, not just one-off activities. From a science perspective we need a diversity of voices to help us understand the context in which data is collected and to engage with us as we set priorities for automated steps within the knowledge discovery process.

A. Summary of Instruction

1. Descriptive Data

Database Design and Prototyping (IS455) (formerly Introduction to Databases (LIS490DB))

Fall 2009, 2010, 2011, 2012, Spring 2010 (online), Spring 2011(online), Fall 2015, Fall 2017 (x2), Spring 2018, Fall 2018, Spring 2019, Fall 2019

List of instructors ranked as excellent Fall 2010, Spring 2019

Text Mining (LIS590TX)

Spring 2010, Fall 2010 (online), Spring 2012, 2013, 2014, 2015 (hybrid), 2016 (hybrid), 2018, 2019, Fall 2020, Fall 2021

List of instructors ranked as excellent Spring 2018

Evidence-based Discovery (LIS590AG)

Core in the SODA specialization Fall 2013 (Inaugural offering), Fall 2014, Fall 2015 (hybrid, both on campus and online)

List of instructors ranked as excellent (online) Fall 2014

Foundations in Socio-technical Data Analytics (LIS590AD)

Core in the SODA specialization, Spring 2013 (Inaugural offering), Spring 2014, Spring 2015 (Co-taught with V. Stodden)

Socio-technical Data Analytics Practicum (LIS591)

Developed practicum requirements for the SODA specialization, coordinated experiences with industry and academic partners

Spring 2013, Fall 2013, Summer 2014, Fall 2014, Summer 2015, Fall 2015

Foundations of Information Processing in LIS (LIS452)

Fall 2011 with Wu Zheng, new course on Java

2. Supervision of Graduate Student Research

Ph.D. Students: Dissertation Research

Qi, Chen, Ph.D. Computer Science, UNC 2009, Research Advisor Fall 2004

Fu, Xin (Robert), Ph.D. Information and Library Science, UNC 2011, *Evaluating Sources of Implicit Feedback for Web Search* Research Advisor 2004

Baker, Nancy, Ph.D. Information and Library Science, UNC 2010, *Methods in Literature-based Drug Discovery* Research Advisor 2004-2005

Cao, Leo, Ph.D. Information and Library Science, UNC 2011, Research Advisor Fall 2005

Lunga, Anita, Ph.D. Computer Science, Duke University 2008, Research Advisor Fall 2005

Winget, Megan, Ph.D. Information and Library Science, UNC 2006, *Annotation of Musical Scores: Interaction and Use Behaviors of Performing Musicians* Committee Member prior to advancing to candidacy. Initial position Faculty Member at University of North Texas

MacMullen, John, Ph.D. Information and Library Science, UNC 2007, *Contextual Analysis of Variation and Quality in Human-curated Gene Ontology Annotations* Committee Member. Initial position Faculty member at University of Illinois.

Phokhwang, Wiriya, Ph.D. Nursing, UNC 2008, Committee Member, *Information Needs and Uses of Nurses in Thailand*.

Zheng, Wu, Ph.D. ABD Primary Advisor 2013-2015, MS GSLIS UIUC 2013, Primary Advisor 2010-2013, *Automatic Detection of Location Relationships*. Initial position software engineer.

Wei, Qin, Ph.D. GSLIS UIUC 2011, Spring 2011, *Information Fusion in Taxonomic Descriptions*. Committee Member

Chee, Brant, Ph.D. GSLIS UIUC 2011, Fall 2011, *Pharmacovigilance: Exploring Alternative Data Sources with Machine Learning and Visualization Techniques*. Committee Member

Vydiswaran, V.G. Vinod, Ph.D. Computer Science UIUC 2012, *Modeling and Predicting Trustworthiness of Online Textual Information*, Committee Member First position post-graduation: University of Michigan.

Park, Dae Hoon, Ph.D. Computer Science UIUC 2016, *Joint Analysis of User-generated Content and Product Information to Enhance User Experience in E-commerce* Advisor for comparison project 2011-2012. Initial position at Yahoo.

Chao, Tiffany C., Ph.D. GSLIS UIUC 2015, *Methods metadata: curating scientific research data for reuse*. Committee Member 2012-2015.

Vauhkonen, Mumtaz, Ph.D Education UIUC current, *Machine Learning and Data Analytics for Multilayer Data in Policy Planning* Proposal Committee Member 2015.

Gabb, Henry, Ph.D. Information Sciences UIUC 2015, Primary Advisor 2012-2015, *An Informatics Approach to Prioritizing Chemical Ingredients in Consumer Products for Risk Assessment*, Currently research scientists at Intel.

Lucic, Ana, Ph.D. Information Sciences UIUC 2017, Primary Advisor 2011 – 2017, *Summarization of Biomedical Texts by Utilizing the Information Extracted from Comparative Sentences*. First position after graduation: De Paul University, now Data Scientist at UIUC.

Park, Hyoungjoo, Ph.D. Information Sciences University of Wisconsin Milwaukee 2018, Committee Member 2017-2018, *The impact of research data sharing and re-use on data citation in STEM fields*. Initial Position Visiting Assistant Professor, University of Wisconsin-Milwaukee,

Guo, Jinlong (Kenney), Ph.D. Information Sciences Primary Advisor 2013-2018 (MS UIUC 2018)

Pence, Justin, Ph.D. Nuclear, Plasma and Radiological Engineering (NPPE) UIUC, *Theorizing and Quantifying Organizational and Social Factors in Probabilistic Risk Assessment of Complex Systems*, Proposal Committee Member Spring 2018, Defended Spring 2020

Lee, Jooho, Ph.D. Information Sciences, UIUC current, *Using grant applications to measure the evolution of collaborative and non-collaborative research* Primary Advisor Fall 2013 –present (Successfully defended proposal Sept 23, 2020 Graduation date: Fall 2020).

Moran, Mark, Ph.D. Informatics, UIUC current, *An Analysis of Text Mining to Detect Impression Management in Chief Executive's Letters with Respect to the Financial Performance of the Company, Industrial Classification (Working Title)*, Primary Advisor Spring 2018 – present.

Keefer, Don, Ph.D. Informatics UIUC current, *Leveraging Representation and Reasoning to Enhance Interpretations and Reduce Knowledge Loss from Multiple Streams of Geological Data (Working Title)* Primary Advisor Fall 2018-present.

Kim, Jenna Ph.D. Information Sciences UIUC, Primary Advisor Fall 2018 - present

Kim, Jinmo Ph.D. Information Sciences UIUC, Primary Advisor Fall 2019 – present

Ryan Wang Ph.D. Information Sciences UIUC, Primary Advisor Spring 2021 – present

Sarol, Janina Ph.D. Informatics UIUC, Primary Co-Advisor (with Halil Kilicoglu) Summer 2021–present

Shufan Ming Ph.D. Information Sciences UIUC, Primary Advisor Summer 2021 – present

George Vazquez Ph.D. Information Sciences UIUC, Primary Advisor Summer 2021– present

Ph.D. Students: Doctoral Research Field Exam

Zheng, Wu, GSLIS 2012

Organisciak, Peter, GSLIS 2012

Ahmed, Shameem, GSLIS 2013

Lucic, Ana, GSLIS 2014

Mishra, Shubhanshu, GSLIS 2015

Kim, Jinsoek, GSLIS 2015

Gabb, Henry, GSLIS 2015

Addawood, Aseel, Informatics Fall 2015
Guo, Jinlong (Kenney), GSLIS 2016
Lee, Jooho, GSLIS 2016
Hoang, Linh Information Sciences Spring 2018
Sarol, Janina, Informatics Fall 2019 & Spring 2020

Ph.D. Students: Annual Review Committee

Choi, Kahyun (2012, Downie Primary Advisor)
Weyerhaeuser, Walker, Certificate of Advanced Study (Spring 2011, Fall 2011)
Johnson, Eric, Certificate of Advanced Study (Fall 2011, Spring 2012)
Thompson, Cheryl (2012-2014, Palmer Primary Advisor)
Kim, Jinseok (2014, Diesner Primary Advisor)
Jiang, Ming (2015, 2016, Torvik Primary Advisor)
Sherman, Garrick (2015, Efron Primary Advisor)
Addawood, Aseel (2015-2016, Bashir Primary Advisor)
Anderson, Caryn (2015, Blake Primary Advisor)
Guan, Yingjun (2018, Torvik Primary Advisor)
Rezapour, Resvaneh (Shadi) (2017- current Diesner Primary Advisor)
Hoang, Linh (2018, Schneider Primary Advisor)
Guo, Jinlong (Kenney) (Ph.D. 2014-2018, Blake Primary Advisor)
Gabb, Henry (Ph.D. 2012-2019, Blake Primary Advisor)
Lucic, Ana (Ph.D. 2011-2017, Blake Primary Advisor)
Lee, Jooho (Ph.D. 2013-current, Blake Primary Advisor)
Moran, Mark (Ph.D. 2018 current, Blake Primary Advisor)
Keefer, Don (Ph.D. 2018-current, Blake Primary Advisor)
Kim, Jenna (Ph.D. 2018-current, Blake Primary Advisor)
Kim, Jinmo (Ph.D. 2018-current, Blake Primary Advisor)

Ph.D. Students: Supervision of Independent Studies:

Churchill, Lewis (Ph.D.) Spring 2004
Armstrong, Chandler (Ph.D., Sociology) Fall 2009
Park, Dae-Hoon (Ph.D., CS) Fall 2011
Gabb, Henry (Ph.D.) Fall 2013
Lucic, Ana (Ph.D.) Spring 2013
Thompson, Cheryl (Ph.D.) Fall 2013
Thomer, Andrea (Ph.D.) Fall 2013

Guo, Jinlong (Kenney) (Ph.D.), Spring 2014

Lee, Joocho (Ph.D.), Spring 2014

Master's Students: MS Practicum

Brown, Ashley (MSLS 2006), Spring 2005

Balke, Kaye (MSIS 2007), Fall 2006

Vasidhyanathan, Vedana (MSLS 2007), Fall 2007

Sackett-Hermann, Erin Griffin (MLS, 2011), Summer 2011

Wang, Zhaoyu (MS/LIS, 2013), Fall 2012, SODA practicum

Chen, Shan (MS/LIS, 2013), Fall 2012, SODA practicum

Lui, Qiyuan (MS/LIS, 2013), Fall 2012, SODA practicum

Fu, Hengyi (MS/LIS, 2013), Spring 2013, SODA practicum

Tieman, Jessica (MS/LIS, 2014), Summer 2014, SODA practicum

Stoytcheva, Svetlozara (MS/LIS, 2015), Summer 2014, SODA practicum

Jiang, Minhao (MS/LIS, 2015), Spring 2015, SODA practicum

Young, Sophie (MS/LIS, 2015) Spring 2015, SODA practicum

Lundberg, Kristen (MS/LIS, 2015), Spring 2015, SODA practicum

Stephens, Andrew (MS/LIS, 2016), Summer 2015, SODA Practicum

Sun, Yidan (Bioinformatics MS/LIS, 2016), Summer 2015, SODA practicum

Bruckner, Lorin (MS/LIS, 2016), Fall 2015, SODA Practicum

Mathur, Milind (MS/IM, 2018), Spring 2018,

Tian, Xinyu (MS/IM, 2018), Spring 2018

Master's Students: *Funded MS Research Assistants*

Maier, Chris (Fall 2004, Spring 2005)

Randall, Meredith (Fall 2005 -Spring 2007)

Burns-Johnson, Toshiba (Fall 2005-Spring 2007)

Purvis, Joshua (Spring 2006)

Bapat, Amol (Fall 2007-Spring 2008)

Jones, Ryan (Fall 2007-Spring 2008)

Chen, Annie (Fall 2008-Spring 2009)

Doty, Christopher (Fall 2008-Spring 2009)

Dunda-Lucca, Carrie (Spring 2009)

Weyerhaeuser, Walker (Summer 2010)

Fleisher, Aaron (Summer 2010)

Painter, Kyle (Summer 2010, Fall 2010, Spring 2011)

Brookshire, Amber (Spring 2011)
Surbeck, Elizabeth (2013-2014), SODA research assistantship
Failing, David (2013-2014), SODA research assistantship
Wang, Zhaoyu (Spring 2013), SODA research assistantship
Chen, Shan (Spring 2013), SODA research assistantship
Stoytcheva, Sveta (Summer 2014), SODA research assistantship
Isenholt, Maxwell (Fall 2014, Spring 2015) SODA research assistantship
Isaacs, Greta (Summer 2015, Fall 2015, Spring 2016) SODA research assistantship

Master's Students: *Supervision of Independent Studies:*

Brown, Ashley (MSLS) Fall 2005
Harrelson, Margaret (MSIS) Fall 2005
Chakravorty, Sayan (MSIS) Spring 2006
Kampov, Julia (MSLS) Spring 2007
Lown, Cory (MSIS) Spring 2007
MacKay, Adrienne (MSIS) Spring 2007
Bapat, Amol (MSIS) Fall 2008
Evans, Monte (MSIS) Fall 2008
Kraus, Scott (MSIS) Fall 2008
Zheng, Wu (PhD) Fall 2009
Dennis, Ellen (MS/LIS) Spring 2011
Johnson, Eric (MS/LIS) Fall 2011
Lynch, Tom (MS/LIS) Fall 2011
Ming, Shufan (MS/IM) Spring 2020
Zhang, Zhongyuan (MS/IM) Spring 2020
Lan, Mengfei (MS/IM) Spring 2020

3. Supervision of Undergraduate Students

Funded Undergraduate Research Assistants

Fowler-McDonald, Lynnelle (Fall 2007)
Mansour, Anna (Fall 2007-Spring 2008)
Dickson, Laura (Fall 2007-Spring 2008)
McKittrick, Max (Fall 2014) SODA research assistantship

4. Other Contributions to Instructional Programs

I wrote the proposal for the Socio-technical Data Analytics (SODA) program in 2010 and served as PI when it was subsequently funded by the IMLS (July 1, 2012 to June 30, 2016). I developed and designed the overall MS and PhD specialization including the core courses below, created the

advertising materials and web site, and reported key outcomes to the funding agency and academic community. We also conducted ongoing interviews and student assessments that are reported in L40. Work on the project informed other information science programs and has been cited by the Committee on Future Career Opportunities and Educational Requirements for Digital Curation.

Foundations of Socio-technical Data Analytics: Socio-technical data analytics combines both the technical (mathematical modeling, databases, social networking, and text mining) and social (economic, ethical, policy, and political) aspects of data analytics. Students work through a series of case studies with real data to develop both the theoretical and hands-on experience necessary to fill leadership roles in e-science, eResearch, and big data. (Inaugural class in Fall 2013). I was included in the UIUC list of instructors ranked as excellent for the (online) Fall 2014 offering of this course.

Evidence-Based Discovery: The evidence-based discovery course introduces students to theoretical models of discovery and decision making and new informatics tools that support discovery and decision making in practice. Students will explore how massive increases in data (and the accompanying analytical methods) are both challenging and reinforcing what we mean by evidence. Students will also develop the rhetorical and statistical methods necessary to combine evidence, which are particularly important in the information intensive world in which we live. (Inaugural class in Spring 2014).

Socio-technical Data Analytics Practicum - Developed practicum requirements for the SODA specialization, coordinated experiences with industry and academic partners.

Material from the two core SODA courses above formed the basis for Sociotechnical Information Systems (IS 504) and Data, Statistical Models and Information (IS 507), that are 2 of the 3 required courses in the new MS in Information Management and doctoral students continue to use the SODA reading list for their field exam. In my role as IM and Bioinformatics Program Director, I continue to shape the MS degree by sharpening the program outcomes, and the campus wide Bioinformatics program where major changes were just approved (Sept 2020) to the iSchool requirements.

IV. Service (Public, Professional/Disciplinary, and University)

A. Summary of Service

1. Public Engagement

Research Reported in the Local and Popular Press

Heavey, A. (2007). Open Science Grid Research Highlights July 2007 – June 2008.

Green, Karen (2008) Feature – Claim jumping through scientific literature, Issue 76, International Science Grid (subsequently mentioned in "American Libraries Direct." June 4, 2008).

<http://renci.org/research/claim-jumping/>

R. Pordes et al (2008), New Science on the Open Science Grid, 125, pg 012070, Journal of Physics: Conference Series

Goldberg, B. (2012) A Snapshot of Librarianship's Never-ending Metamorphosis, What's New in LIS Schools, American Library Association. <http://americanlibrariesmagazine.org/2012/08/21/whats-new-in-lis-schools/>

Impact Illinois 2015 (2015) SODA for Real-world Solutions, p8-9

<https://www.uillinois.edu/common/pages/DisplayFile.aspx?itemId=221303>

GSLIS Magazine (2015) Blake Builds Claim Framework to Analyze and Synthesize Medical Research, https://issuu.com/ischoolui/docs/141152-intersections_web

2. Service to Disciplinary and Professional Societies or Associations

Reviewer for Scholarly Journals and Publications

Journal of the Association for Information Science & Technology (2005-current)

Information Processing and Management (2005-2006, 2011, 2012)

Data and Knowledge Engineering (2005)

Information Retrieval (2007)

Transactions of Information Systems (2007)

Decision Support Systems (2007)

Bioinformatics (2008)

Journal of Biomedical Informatics (2008, 2012, 2013, 2015)

Transactions on Interactive Intelligent Systems (2011)

Journal of American Medical Informatics Association (2015)

Editorial and Advisory Boards

Editorial board of the *Journal of Biomedical Discovery and Collaboration* (2005-2010)

Advisory Board: Curriculum Development: Digital Libraries PIs: E. Fox (Virginia Tech), B. Wildemuth (UNC CH), J. Pomerantz (UNC CH), (2006-2009)

Advisory Board of NEH proposal by Gregory Crane (Tufts University) (June 2004-2006)

Advisory Board of IMLS proposal by Kate Oliver (Medical Library, Johns Hopkins University) and Wendy Scott (North Carolina State University Libraries) "Recruiting Expertise: 21st Century Academic Research Library"

Conference and Chapter Reviewing

American Medical Informatics Association Symposium (2001-2009,2011,2014, 2015-current)

ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (2016, 2017)

Joint Conference on Digital Libraries (2004, 2013)

Biomedical Natural Language Processing (2007, 2008, 2009)

Conference on Information and Knowledge Management (2004-2005, 2007, 2008, 2009)

Conference of the ACM Special Interest Group on Information Retrieval (SIGIR) (2004-2008)

American Society for Information Science & Technology (2005, 2010)

Annual Review of Information Science and Technology (2007)

Program Chair and Committee Activities

Program Committee Member, Conference on Information and Knowledge Management (CIKM) (2006, 2007)

Program Committee Member, Symposium of the American Medical Informatics Association (AMIA) (2006)

Program Committee Member, Association of Computational Linguistics Workshop on Biology Natural Language Processing (ACL BioNLP) (2007, 2008, 2009, 2010)

Program Committee Member, 5th International Conference on e-Social Science (2009), National Centre for e-Social Science (UK) and German Social Science Infrastructure Services (GESIS).

Program Committee Member, 73rd American Society for Information Science & Technology Annual Meeting (2010)

Program Committee Member, 8th International conference on Language Resources and Evaluation (LREC) (2012)

Program Committee Member, Detecting Structure in Scholarly Discourse (DSSD) Workshop at ACL 2012, Jeju Island, Korea on July 8-14, 2012.

Program Committee Member, The 7th Linguistic Annotation Workshop & Interoperability with Discourse, ACL Workshop, Sofia, August 8-9, 2013

Poster Chair, iConference Fort Worth, TX, 2013

Scientific Committee member: 9th International Conference on Language Resources and Evaluation, 2014

Program Committee member, iConference Program Committee, Berlin, Germany, 2014; Irvine, CA 2015; Philadelphia, PA 2016

Program Committee, 2016 ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB 2016)

Paper Co-chair, American Society for Information Science & Technology Annual Meeting, Copenhagen, Denmark, October 14-18, 2016

Paper Co-chair for the Text mining and Classification track of BCB'17 (ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics), Boston, MA, August 20-23, 2017

iConference 2019 Associate Chair of Program Committee, College Park, Maryland.

Conference Co-chair, Association for Information Science & Technology Annual Meeting, Melbourne, Australia, October 19-23, 2019

Symposium on Intelligent Data Analysis 2021 (IDA 2021), Program Committee, Porto, Portugal (moved online due to COVID), April 26 – 28, 2021.

3. University/Campus Service

GSLIS Research Showcase Committee Member (2009-2010)

UIUC Institutional Organizations that Foster Interdisciplinary Education and Research, Illinois Innovation Summit (with Gasser and Downie, 2010-11)

Ad Hoc Tenure case committee (2010/11)

UIUC Health Education Committee (2010-2011)

GSLIS Research Computing Programmer/Manager Search Committee (2011)

UIUC Illinois Informatics Initiative Director Search Committee (Spring 2011)

GSLIS Awards Committee (2011/12)

UIUC Campus Budget Oversight Committee (2010-2012)

University of Illinois Visioning Future Excellence in Health and Wellness (Spring 2013)

Committee on Research and Education at the NCSA (2015)

GSLIS Bioinformatics Coordinator (2012-15)

GSLIS Master's and Certificate of Advanced Study admissions committee (2009/10, 2015/16)

iSchool Doctoral Studies Committee (2014/15, 2017/18)

Chair iSchool Doctoral Studies Committee (2012/13, 2013/14)

Space committee (2017/18, 2018/19)

Chair Ad hoc 3rd Year review committee (Diesner) (2018/19)

Ad hoc ISchool Bioinformatics committee (2018/19)

Co-Chair (with Knox) Ad Hoc committee on Human Computer Interaction Curriculum Development (2019)

Chair iSchool Cloud Computing and or cyberinfrastructure, data visualization Instructor/Teaching Faculty Search Committee (2019/2020)

Chair iSchool Explainable AI Faculty Search Committee (2019/2020)

Committee Member Curriculum Committee (2018-2021)

Committee Member Adjunct Faculty Search Committees (Fall 2019 and Spring 2020)

Elected Committee Member, iSchool Executive Committee (2010/11, 2011/12, 2018/19, 2019/20)

Committee member of Ad Hoc Faculty Grievance sub-committee

Committee member Non-area specific Faculty Search Committee (2019/20)
Committee Member Executive Committee of the Graduate College (2019-2021)
Chair of MSIM/BIO Program Committee (2018-2021)
Program Director MSIM/BIO (2018-2021)
PI Midwest Data Hub (PI June 2020-2023, Co-PI June 2019-May 2020)
NCSA Delta project's External Advisory Committee (April 2021-current)
Committee Member iSchool Explainable AI Faculty Search Committee (2021-2022)
Chair 3rd Year review committee (Chin) (2021/22)