Contact

Drexel University
Department of Information Science
College of Computing and Informatics
30 North 33^{rd} Street
Philadelphia, PA 19104

Email: Jake[dot]Williams[at]drexel[dot]edu
Homepage: http://www.pages.drexel.edu/~jw3477/

Research Interests

Data science, scientific programming, computational social science, computational linguistics and natural language processing, mathematics, machine learning, algorithms, and scalability.

Higher education employment

Assistant Professor, Department of Information Science, Drexel University **2016–**

Courses teaching/taught:

Data acquisition and pre-processing (Drexel DSCI 511)	Fall 2018 and 2019; Spring 2019
Methods for analysis and interpretation (Drexel DSCI 521)	Winter 2019 and 2020
Introduction to Data Science (Drexel INFO 103)	Spring 2018
Foundations of Data Science (Drexel INFO 825)	Winter 2018; Spring 2020
Perspectives on Information Systems (Drexel INFO 540)	Fall 2017
Introduction to data science (Drexel INFO 240)	Winter, Spring 2017

Course development:

Methods for analysis and interpretation (Drexel DSCI 521)	Fall 2018
Data acquisition and pre-processing (Drexel DSCI 511)	Summer 2018
Foundations of Data Science (Drexel INFO 825)	Fall 2017
Introduction to Data Science (Drexel INFO 103)	Fall 2016

Postdoctoral Researcher, School of Information, UC Berkeley **2015–2016**

Faculty Instructor, School of Information, UC Berkeley **2015**

Courses taught:

Machine Learning at Scale (UC Berkeley DATASCI W261)	2015

Graduate Research Assistant, Department of Mathematics, University of Vermont (UVM) **2009–2015**

Graduate Teaching Fellow, Department of Mathematics and Statistics, UVM **2009–2014**

Courses taught:

College Algebra (UVM MATH 009)	(seven semesters) 2009–2014
Calculus with Applications (UVM MATH 019)	2012
Applications of Finite Math (UVM MATH 017)	2011
Guest lecturer, Principles of Complex Systems (UVM MATH 300)	2012, 2013
Guest lecturer, Linear Algebra (UVM MATH 124)	2011

Education

Ph.D., Mathematical Sciences, UVM	2015
Certificate of graduate study in complex systems, UVM	2012

M.S., Mathematics, UVM 2011

B.A., Physics, UVM 2007

Software

Python *partitioner*

Website: http://jakerylandwilliams.github.io/partitioner/

Github: https://github.com/jakerylandwilliams/partitioner

PyPi: https://pypi.python.org/pypi/partitioner

Support

Award Title: Moderating Effects of Automation on Information Transmission in Social Forums
Lead Institution PI: J. R. Williams, Drexel University
Prime Sponsor: National Science Foundation

Award Title: LIS Education and Data Science for the National Digital Platform
Lead Institution PI: J. Greenberg, Drexel University
Prime Sponsor: Institute of Museum and Library Sciences

Award Title: Mitigating Deception and Misinformation in Social Forums
Lead Institution PI: J. R. Williams, Drexel University
Prime Sponsor: Drexel University, Bridge Funding

Subaward Title: Clinical Trial Description Simplification Using Patient-Facing Feature Extraction
Subrecipient PI: J. R. Williams, Drexel University
Award Title: Clinical and Translational Science Institute Lead Institution PI: J. Grandis, The Regents of the University of California, San Francisco
Prime Sponsor: NIH National Center for Advancing Translational Sciences

Publications

A general solution to the preferential selection model.
J. R. Williams, D. Solano-Propeza, and J. R. Hunsberger.
Arxiv Preprint (2020).

NewsTweet: A Dataset of Social Media Embedding in Online Journalism.
M. I. Mujib, H. S. Heidenreich, C. J. Murphy, G. C. Santia, A. Zelenkauskaite, and J. R. Williams.
Arxiv Preprint (2020).

Investigating Coordinated ‘Social’ Targeting of High-Profile Twitter Accounts.
H. S. Heidenreich, M. I. Mujib, and J. R. Williams.
Arxiv Preprint (2020).

Investigating Coordinated ’Social’ Targeting of High-Profile Twitter Accounts.
H. S. Heidenreich, M. I. Mujib, and J. R. Williams.
International Conference on Computational Social Science (2020).

Tailorable Autonomous Motivational Interviewing Conversational Agent.
D. Smriti, J. Y. Shin, M. I. Mujib, M. Colosimo, T.-S. Kao, J. R. Williams, and J. Huh-Yoo.
Conference on Human Factors in Computing Systems (2020).

A scalable machine learning approach for measuring violent and peaceful forms of political protest participation with social media data.
L. J. Anastasopoulos and J. R. Williams.
PLoS ONE (2019).

Latent semantic network induction in the context of linked example senses.
H. S. Heidenreich and J. R. Williams.
Proceedings of the 2019 EMNLP Workshop W-NUT: The 5th Workshop on Noisy User-generated Text (2019).

Detecting Social Bots on Facebook in an Information Veracity Context.
G. C. Santia, M. I. Mujib, and J. R. Williams.
Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (2019).

Making Sense of Clinical Trial Descriptions: A Text Analysis Approach.
M. I. Mujib, J. R. Williams, A. Gottsegen, Y. Sharma, A. Chatterjee, O. Gologorskaya.
Text Analysis Across Domains Conference (2019).

BuzzFace: A News Veracity Dataset with Facebook User Commentary and Egos.
G. C. Santia and J. R. Williams.
Proceedings of the Twelfth International AAAI Conference on Web and Social Media (2018).

Expanding Consumer Health Vocabularies with Frequency-Conserving Internal Context Models.
M. I. Mujib, C. C. Yang, M. Zhao, and J. R. Williams.
IEEE International Conference on Healthcare Informatics (2018).

Empowering targeted tenant organizing: geographic forecasting of housing insecurity.
A. Gottsegen and J. R. Williams.
Women in Data Science Conference (2018).

Understanding disciplinary vocabularies using a full-text enabled domain-independent term extraction approach.
E. Yan, J. R. Williams, and Z. Chen.
PloS ONE (2017).

The Lexicocalorimeter: Gauging public health through caloric input and output on social media.
S. E. Alajajian, J. R. Williams, A. J. Reagan, S. C. Alajajian, M. R. Frank, L. Mitchell, J. Lahne, C. M. Danforth, and P. S. Dodds.
PloS ONE (2017).

Benchmarking sentiment analysis methods for large-scale texts: A case for using continuum-scored words and word shift graphs.
A. J. Reagan, B. Tivnan, J. R. Williams, C. M. Danforth, and P. S. Dodds.
EPJ Data Science (2017).

Simon's fundamental rich-gets-richer model entails a dominant first-mover advantage.
P. S. Dodds, D. R. Dewhurst, F. F. Hazlehurst, C. M. Van Oort, L. Mitchell, A. J. Reagan, J. R. Williams, C. M. Danforth.
Physical Review E (2017).

Context-Sensitive Recognition for Emerging and Rare Entities.
J. R. Williams and Giovanni C. Santia.
Proceedings of the 3rd Workshop on Noisy User-generated Text (2017).

Boundary-Based MWE Segmentation With Text Partitioning.
J. R. Williams.
Proceedings of the 3rd Workshop on Noisy User-generated Text (2017).

Is space a word, too?
J. R. Williams and G. S. Santia.
Arxiv Preprint (2017).

Identifying violent protest activity with scalable machine learning.
L. Anastasopoulos and J. R. Williams.
Annual Meeting of the Americal Politial Science Association (2016).

Vaporous Marketing: Uncovering Pervasive Electronic Cigarette Advertisements on Twitter.
E. M. Clark, C. A. Jones, J. R. Williams, A. N. Kurti, M. C. Norotsky, C. M. Danforth, P. S. Dodds.
PLoS ONE (2016).

Sifting robotic from organic text: A natural language approach for detecting automation on Twitter.
E. M. Clark, J. R. Williams, C. A. Jones, R. A. Galbraith, C. M. Danforth, P. S. Dodds.
Journal of Computational Science (2016).

Photographic home styles in Congress: a computer vision approach.
L. J. Anastasopoulos, D. Badani, C. Lee, S. Ginosar, J. R. Williams.
Arxiv Preprint (2016).

Zipf 's law is a consequence of coherent language production
J. R. Williams, J. P. Bagrow, A. J. Reagan, S. E. Alajajian, C. M. Danforth, and P. S. Dodds.
Arxiv Preprint (2016).

Identifying missing dictionary entries with frequency-conserving context models. J. R. Williams, E. M. Clark, J. P. Bagrow, C. M. Danforth, and P. S. Dodds.
Physical Review E (2015).

Zipf's law holds for phrases, not words.
J. R. Williams, P. R. Lessard, S. Desu, E. M. Clark, J. P. Bagrow,C. M. Danforth, P. S. Dodds.
Nature Scientific Reports (2015).

Reply to Garcia et al.: Common mistakes in measuring frequency-dependent word characteristics.
P. S. Dodds, E. M. Clark, S. Desu, M. R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, K. Megerdoomian, M. T. McMahon, B. F. Tivnan, and C. M. Danforth.
PNAS (2015).

Text mixing shapes the anatomy of rank-frequency distributions.
J. R. Williams, J. P. Bagrow, C. M. Danforth, and P. S. Dodds.
Physical Review E (2015).

Human language reveals a universal positivity bias.
P. S. Dodds, E. M. Clark, S. Desu, M. R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, K. Megerdoomian, M. T. McMahon, B. F. Tivnan, and C. M. Danforth.
PNAS (2015).

Constructing a taxonomy of fine-grained human movement and activity motifs through social media.
M. R. Frank, J. R. Williams, L. Mitchell1, J. P. Bagrow, P. S. Dodds, C. M. Danforth.
Arxiv Preprint (2015).

Low-power, phase-preserving 2R Amplitude Regenerator. T. I. Lakoba, J. R. Williams, and M. Vasilyev.
Optics Communications (2011).

NALM–based, phase–preserving 2R regenerator of high–duty–cycle pulses.
T. I. Lakoba, J. R. Williams, and M. Vasilyev.
Optics Express (2011).

Research participation

University of Pennsylvania Computational Linguistics and Lunch Philadelphia, PA
Invited talk: Constructive network architectures for text segmentation 2018

CNHP Stein Family Fellowship Symposium Philadelphia, PA
Invited talk: Clinical Trial Description Simplification Using Patient-Facing Feature Extraction 2018

First Northeast Regional Conference on Complex Systems Binghamton, NY
Extended Abstract and talk: Is space a word, too? 2018

Text Analysis Across Domains Conference Berkeley, CA
Invited Talk: Minimal Semantic Units in Text Analysis, 2017

Linguistic Data Consortium Institute Philadelphia, PA
Invited Talk: Boundary-Based MWE Segmentation and Applications 2017

EMNLP, 3^{nd} Workshop on Noisy User-generated Text Copenhagen, Denmark
Shared task entrant: Novel and Emerging Named Entity Recognition Poster: Context-Sensitive Recognition of Emerging and Rare Entities 2017

COLING, 2^{nd} Workshop on Noisy User-generated Text Osaka, Japan
Shared task entrant: Geolocation Prediction in Twitter (unpresented) 2016

APSA Annual Meeting & Exhibition Philadelphia, PA
Lecture: “Using Scalable Machine Learning to Understand Violent Collective Action” 2016

Berkeley Institute for Data Science (BIDS) Lectures, UC Berkeley, Berkeley, CA
Lecture: “The bag of phrases approach to text analysis” 2016

Wikimedia Research Groups, Wikimedia Foundation, San Francisco, CA
Lecture: “Collocation-based gap analysis for the Wiktionary” 2016

Ph.D.-postdoc research exchange, UC Berkeley, Berkeley, CA
Lecture: “Phrase-based text analysis” 2015

Multiscale modeling of the food system, American Institute of Mathematics workshop, San Jose, California
Invited talk: “Exploring food systems via the lens of social media” 2015

Student research conference, UVM, Burlington, VT
Poster: “Some effects of text-mixing: Testing the core language hypothesis” 2014

Student research conference, UVM, Burlington, VT
Poster: “Event detection and classification via natural language analysis” 2013

Applied mathematics seminar, Department of Mathematics and Statistics, UVM, Burlington, VT
Lecture: “Identifying lexical units in large corpora” 2012

Student scholars poster competition, UVM, Burlington, VT
Poster: “Identifying idioms in natural language: Massive data text partitioning” 2012

Awards

Graduate research assistantship, Department of Mathematics, UVM 2009–2015

John F. Kenney Award, College of Engineering and Mathematical Sciences, UVM 2014

Graduate teaching assistantship, Department of Mathematics, UVM 2009–2013

Nominee for UVM Graduate Teaching Assistant of the Year award, Department of Mathematics, UVM 2011

Presidential Scholarship, UVM 2003–2007

Languages

Fluent: English, Python, Perl, R, Unix Bash, Hadoop, MRJob, Spark, LaTeX, and Matlab

Capable: German, Javascript, HTML, PBS, and Mathematica

Service

Reviewer, Annual Meeting of the Association for Computational Linguistics, 2019–

Reviewer, International Conference on Web and Social Media, 2018–

Reviewer, Conference on Empirical Methods in Natural Language Processing, 2018–

Referee, Cognitive Science, 2018–

Curriculum committee member and program developer, Drexel MSDS program, 2017–

Referee, Physical Letters A 2017–

Curriculum committee member, Drexel BSDS program 2016–

Referee, Journal of Statistical Physics 2016–

Planning committee member, UC Berkeley D-Lab Computational Text Analysis Working Group 2016–2017

Committee member for the development of the UC Berkeley MIDS statistics curriculum	2015
Planning committee member, Solving Stress Macmillan Symposium and Lecture, UVM	2014–2015
Student liaison to the UVM Department of Mathematics Graduate Committee	2011–2013
Graduate Student Senate representative for the UVM Department of Mathematics	2009–2010
Lead tutor overseeing other educators in the Shades of Ebony tutoring program	2008–2009