Kyle Williams
Curriculum Vitae

kyle@kmw.ai · kylewilliams.ai · 814-852-8787

January 11, 2023

Education

2016	**The Pennsylvania State University** – State College, PA Ph.D. in Information Sciences and Technology Advisor: C. Lee Giles
2012	**University of Cape Town** – Cape Town, South Africa Master of Science in Computer Science
2009	**University of Cape Town** – Cape Town, South Africa Bachelor of Business Science in Management Studies & Computer Science

Industry experience

2016 – 2022 **Microsoft** – Redmond, WA
Senior Applied Scientist

- **NLP for healthcare**, including abstractive summarization of doctor-patient conversations, social determinants of health, **named entity recognition** and linking.
- **Conversational NLP** including dialog systems.
- Search engine and intelligent assistant **analysis and evaluation** including **A/B testing** with **millions of data points** to measure product impact.
- Managing **machine learning models in production setting**.
- **Mentoring and managing** PhD interns and junior colleagues.
- Research and **advancing state of the art** in NLP and Information Retrieval.

2015 **Microsoft Research** – Redmond, WA
Research Intern
Research and prototypes on good abandonment in mobile search.

2014 **Oracle Labs** – Burlington, MA
Research Intern
Research on ranking and search result diversification for eCommerce.

Teaching experience

2015   **Teaching assistant, IST 441: Information Retrieval and Search Engines (The Pennsylvania State University)**
Office hours; grading; running labs; setting student projects; guest lectures.

2011   **Teaching assistant, CS2 (University of Cape Town)**
Office hours; grading; managing tutors; running labs; setting test cases for algorithms.

2010   **Teaching assistant, CS1 (University of Cape Town)**
Office hours; grading; managing tutors; running labs.

2005-2010   **High School Music Teacher (Cape Town, South Africa)**

Awards and Honors

2014   Highest precision and F_1-score among all participants in the PAN 2014 Source Retrieval Task for Plagiarism Detection

2014   3rd place winner in the Engineering category at the 2014 Penn State Graduate Exhibition

2014   SIGWEB DocEng 2014 Student Travel Award

2014   Best deployed application award at 2014 International Conference on Innovative Applications of Artificial Intelligence for *CiteSeerX: AI in a Digital Library Search Engine*

2014   Best paper award nomination at 2013 International Conference on Cloud Engineering for *Migrating a Digital Library to a Private Cloud*

2013   Best Paper Award at 2013 South African Institute for Computer Scientists and Information Technologists Conference (SAICSIT) for *A Comparison of Machine Learning Techniques for Handwritten |Xam Word Recognition*

2013   Highest F_1-score among all participants in the PAN 2013 Source Retrieval Task for Plagiarism Detection

2013   2nd place winner in the Engineering category at the 2013 Penn State Graduate Exhibition

2012   Jordan H. Rednor Graduate Fellowship

2012   Entelect University of Cape Town Computer Science Best Publication Award by an M.S. (Computer Science) Student

2012   National Research Foundation (South Africa) Free-standing Scholarship

2011   Honourable mention at 13th International Conference on Asia-Pacific Digital Libraries (ICADL) for *Creating a Handwriting Recognition Corpus for Bushman Languages*

2011   National Research Foundation (South Africa) Free-Standing Scholarship

2011   University of Cape Town Travel Grant

2011   University of Cape Town Equity Scholarship

2010	University of Cape Town Equity Scholarship
2010	National Research Foundation (South Africa) Prestigious/Equity Scholarship
2010	University of Cape Town University Research Scholarship
2010	SIGWEB JCDL 2010 Student Travel Award
2010	University of Cape Town Michaelis Fine Art Research Scholarship
2008	Certificate of achievement for achieving top mark for Project and Operations Management in the Academic Development Bachelor of Business Science Class
2007	Certificate of achievement for third overall position in the 3rd Year Academic Development Bachelor of Business Science Class

Peer Reviewed Publications

2022   George Michalopoulos, **Kyle Williams**, Thomas Lin, and Gagandeep Singh. "MedicalSum: A Guided Clinical Abstractive Summarization Model for Generating Medical Reports from Patient-Doctor Conversations". In: *Findings of the Association for Computational Linguistics: EMNLP 2022*

2019   **Kyle Williams**. "Neural Lexicons for Slot Tagging in Spoken Language Understanding". In: *2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL '19)*

Kyle Williams, Seyyed Hadi Hashemi, and Imed Zitouni. "Automatic Task Completion Flows from Web APIs". In: *International Conference on Research and Development in Information Retrieval (SIGIR '19)*. ACM

Kyle Williams. "Zero Shot Intent Classification Using Long-Short Term Memory Networks". In: *INTERSPEECH 2019*

Kunho Kim, Rahul Jha, **Kyle Williams**, Alex Marin, and Imed Zitouni. "Slot Tagging for Task Oriented Spoken Language Understanding in Human-to-Human Conversation Scenarios". In: *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*. Association for Computational Linguistics

2018   Seyyed Hadi Hashemi, **Kyle Williams**, Ahmed El Kholy, Imed Zitouni, and Paul A. Crook. "Impact of Domain and User's Learning Phase on Task and Session Identification in Smart Speaker Intelligent Assistants". In: *International Conference on Information and Knowledge Management (CIKM 2018)*. ACM

Seyyed Hadi Hashemi, **Kyle Williams**, Ahmed El Kholy, Imed Zitouni, and Paul A. Crook. "Measuring User Satisfaction on Smart Speaker Intelligent Assistants Using Intent Sensitive Query Embeddings". In: *International Conference on Information and Knowledge Management (CIKM 2018)*. ACM

2017   **Kyle Williams** and Imed Zitouni. "Does That Mean You're Happy? RNN-based Modeling of User Interaction Sequences to Detect Good Abandonment". In: *Proceedings of the International Conference on Information and Knowledge Management (CIKM '17)*

2016   Shuting Wang, Alexander G. Ororbia, Zhaohui Wu, **Kyle Williams**, Chen Liang, Bart Pursel, et al. "Using Prerequisites to Extract Concept Maps from Textbooks". In: *Interactional Conference on Information and Knowledge Management.* ACM

Madian Khabsa, Aidan C. Crook, Ahmed Hassan Awadallah, Imed Zitouni, Tasos Anastasakos, and **Kyle Williams**. "Learning to Account for Good Abandonment in Search Success Metrics". In: *Interactional Conference on Information and Knowledge Management (CIKM '16).* ACM

Julia Kiselava, **Kyle Williams**, Jiepu Jiang, Ahmed Hassan Awadallah, Imed Zitouni, Aidan C. Crook, et al. "Predicting User Satisfaction with Intelligent Assistants". In: *International Conference on Research and Development in Information Retrieval (SIGIR '16).* ACM

Kyle Williams, Julia Kiselava, Imed Zitouni, Aidan C. Crook, Ahmed Hassan Awadallah, and Madian Khabsa. "Is This Your Final Answer? Evaluating the Effect of Answers on Good Abandonment in Mobile Search". In: *International Conference on Research and Development in Information Retrieval (SIGIR '16).* ACM

Kyle Williams, Jian Wu, Zhaohui Wu, and C. Lee Giles. "Information Extraction for Scholarly Digital Libraries". In: *International Joint Conference on Digital Libraries (JCDL '16).* ACM

Kyle Williams and C. Lee Giles. "Improving Similar Document Retrieval Using a Recursive Pseudo Relevance Feedback Strategy". In: *International Joint Conference on Digital Libraries (JCDL '16).* ACM

Kyle Williams, Julia Kiselava, Aidan C. Crook, Imed Zitouni, Ahmed Hassan Awadallah, and Madian Khabsa. "Detecting Good Abandonment in Mobile Search". In: *International World Wide Web Conference (WWW '16).* ACM

Bart Pursel, Chen Liang, Shuting Wang, Zhaohui Wu, **Kyle Williams**, Benjamin Brautigam, et al. "BBookX: Design of an Automated Web-based Recommender System for the Creation of Open Learning Content". In: *Web Education Workshop @ WWW 2016.* ACM

Julia Kiselava, **Kyle Williams**, Jiepu Jiang, Ahmed Hassan Awadallah, Aidan C. Crook, Imed Zitouni, et al. "Understanding User Satisfaction with Intelligent Assistants". In: *Conference on Human Information Interaction and Retrieval (CHIIR '16).* ACM

Chen Liang, Shuting Wang, Zhaohui Wu, **Kyle Williams**, Bart Pursel, Benjamin Brautigam, et al. "BBookX: Building Online Open Books for Personalized Learning". In: *AAAI Conference on Artificial Intelligence.* Association for the Advancement of Artificial Intelligence

2015   Jian Wu, Jason Killian, Huaiyu Yang, **Kyle Williams**, Sagnik Ray Choudhury, Suppawong Tuarob, et al. "PDFMEF: A Multi-Entity Knowledge Extraction Framework for Scholarly Documents and Semantic Search". In: *Proceedings of the 8th International Conference on Knowledge Capture.* ACM

Kyle Williams and C. Lee Giles. "On the Use of Similarity Search to Detect Fake Scientific Papers". In: *Proceedings of the 2015 International Conference on Similarity Search and Applications.* Springer

Shuting Wang, Chen Liang, Zhaohui Wu, **Kyle Williams**, Bart Pursel, Benjamin Brautigam, et al. "Concept Hierarchy Extraction from Textbooks". In: *Proceedings of the 2015 ACM Symposium on Document Engineering.* ACM

Alexander G. Ororbia, Jian Wu, Madian Khabsa, **Kyle Williams**, and C. Lee Giles. "Big Scholarly Data in CiteSeerX: Information Extraction from the Web". In: *Proceedings of BigScholar at WWW.* ACM

Jian Wu, **Kyle Williams**, Hung-Hsuan Chen, Cornelia Caragea, Madian Khabsa, Suppawong Tuarob, et al. "CiteSeerX: AI in a Digital Library Search Engine". In: *AI Magazine* 36.3 ()

2014   Jian Wu, **Kyle Williams**, Madian Khabsa, and C. Lee Giles. "The Impact of User Corrections To Crawl-Based Digital Libraries: A CiteSeerX Perspective". In: *2014 International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom).* IEEE

Zhaohui Wu, Jian Wu, Madian Khabsa, **Kyle Williams**, Hung-Hsuan Chen, Wenyi Huang, et al. "Towards Building a Scholarly Big Data Platform: Challenges, Lessons and Opportunities". In: *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries.* ACM

Kyle Williams, Jian Wu, and C. Lee Giles. "SimSeerX: A Similar Document Search Engine". In: *Proceedings of the 2014 ACM symposium on Document engineering.* ACM

Kyle Williams, Hung-Hsuan Chen, and C. Lee Giles. "Classifying and Ranking Search Engine Results as Potential Sources of Plagiarism". In: *Proceedings of the 2014 ACM symposium on Document engineering.* ACM

Kyle Williams, Lichi Li, Madian Khabsa, Jian Wu, Patrick C. Shih, and C. Lee Giles. "A Web Service for Scholarly Big Data Information Extraction". In: *Proceedings of the IEEE International Conference on Web Services.* IEEE

Jian Wu, Alexander G. Ororbia, **Kyle Williams**, Madian Khabsa, Zhaohui Wu, and C. Lee Giles. "Utility-Based Control Feedback in a Digital Library Search Engine: Cases in CiteSeerX". in: *9th International Workshop on Feedback Computing.* USENIX

Jian Wu, **Kyle Williams**, Hung-Hsuan Chen, Madian Khabsa, Cornelia Caragea, Alexander G. Ororbia, et al. "CiteSeerX: AI in a Digital Library Search Engine". In: *Twenty sixth Annual Conference on Innovative Applications of Artificial Intelligence.* AAAI

Kyle Williams, Jian Wu, Sagnik Ray Choudhury, Madian Khabsa, and C. Lee Giles. "Scholarly big data information extraction and integration in the CiteSeerX digital library". In: *Proceedings of the 10th International Workshop on Information Integration on the Web.* IEEE

Jian Wu, Pradeep Teregowda, **Kyle Williams**, Madian Khabsa, Douglas Jordan, Eric Treece, et al. "Migrating a Digital Library to a Private Cloud". In: *2014 IEEE International Conference on Cloud Engineering (IC2E)*. IEEE

Kyle Williams, Hung-Hsuan Chen, and C. Lee Giles. "Supervised Ranking for Plagiarism Source Retrieval". In: *CLEF 2014 Evaluation Labs and Workshop Working Notes Papers*

Cornelia Caragea, Jian Wu, Alina Ciobanu, **Kyle Williams**, Juan Fernández-Ramírez, Hung-Hsuan Chen, et al. "CiteSeerX: A Scholarly Big Dataset". In: *Proceedings of the European Conference on Information Retrieval*. Springer

Cornelia Caragea, Jian Wu, **Kyle Williams**, Sujatha G. Das, Madian Khabsa, Pradeep Teregowda, et al. "Automatic Identification of Research Articles from Crawled Documents". In: *Web-Scale Classification: Classifying Big Data from the Web*

Chen Liang, Shuting Wang, Zhaohui Wu, **Kyle Williams**, Bart Pursel, and C. Lee Giles. "BBookX: An Automatic Book Creation Framework". In: *Proceedings of the 2014 ACM Symposium on Document Engineering*. ACM

2013   **Kyle Williams**, Hung-Hsuan Chen, Sagnik Ray Choudhury, and C. Lee Giles. "Unsupervised Ranking for Plagiarism Source Retrieval". In: *PAN AT CLEF 2013*

Kyle Williams, Jorgina K. do R. Paihama, and Hussein Suleman. "A comparison of machine learning techniques for handwritten Xam word recognition". In: *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*. Best Paper Award. ACM

Kyle Williams and C. Lee Giles. "Near duplicate detection in an academic digital library". In: *Proceedings of the 2013 ACM symposium on Document engineering*. ACM

2012   Tresor Mvumbi, Flora Kundaeli, Zafika Mazi, **Kyle Williams**, and Hussein Suleman. "An Online Meeting Tool for Low Bandwidth Environments". In: *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*. ACM

Jorgina K. do R. Paihama, **Kyle Williams**, and Hussein Suleman. "Assessing the Design of Web Interoperability Protocols". In: *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*. ACM

Marius Nel, **Kyle Williams**, and Hussein Suleman. "Simple Large Image Support in DSpace". In: *Proceedings on the International Conference on Asia-Pacific Digital Libraries*. Springer

Marwan Mohamed Nour, **Kyle Williams**, and Hussein Suleman. "ORchiD: Evaluating Simple Repository Deposit for Open Educational Resources". In: *Proceedings on the International Conference on Asia-Pacific Digital Libraries*. Springer

Michelle Havenga, **Kyle Williams**, and Hussein Suleman. "Motivating Users to Build Heritage Collections Using Games on Social Networks". In: *Proceedings on the International Conference on Asia-Pacific Digital Libraries.* Springer

Lighton Phiri, **Kyle Williams**, Miles Robinson, Stuart Hammar, and Hussein Suleman. "Bonolo: A General Digital Library System for File-based Collections". In: *Proceedings on the International Conference on Asia-Pacific Digital Libraries.* Springer

2011   **Kyle Williams** and Hussein Suleman. "Creating a Handwriting Recognition Corpus for Bushman Languages". In: *Proceedings on the International Conference on Asia-Pacific Digital Libraries.* Honorable Mention. Springer

Kyle Williams and Hussein Suleman. "Using a Hidden Markov Model to Transcribe Handwritten Bushman Texts". In: *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries.* ACM

2010   **Kyle Williams**, Sanvir Manilal, Lebogang Molwantoa, and Hussein Suleman. "A Visual Dictionary for an Extinct Language". In: *Proceedings on the International Conference on Asia-Pacific Digital Libraries.* Springer

Kyle Williams and Hussein Suleman. "Translating Handwritten Bushman Texts". In: *Proceedings of the 10th annual joint conference on Digital libraries.* ACM

Rizmari Versfeld, Spencer Lee, Edward Fox, Hussein Suleman, and **Kyle Williams**. "Digital Library in a 3D Virtual World: The Digital Bleek and Lloyd Collection in Second Life". In: *Proceedings of the International Conference on Theory and Practice of Digital Libraries.* Springer

Patents

2020   **Kyle Williams** and Imed Zitouni. *Automatically cross-linking application programming interfaces.* US Patent 10,628,527

2019   Imed Zitouni, Ahmed Hassan Awadallah, Aidan Crook, Bradley Wethington, and **Kyle Williams**. *Identifying Abandonment Using Gesture Movement.* US Patent 10,229,212

Talks

2017   Evaluating Search Engines and Intelligent Assistants. *Guest presentation for IST 441: Information Retrieval and Search Engines*, Pennsylvania State University, State College, PA, USA.

2016   Information Extraction for Scholarly Digital Libraries. *Tutorial at The International Joint Conference on Digital Libraries*, Newark, NJ, USA.

2014   The Impact of User Corrections to Crawl-based Digital Libraries. *Collaborative Computing*, Miami, FL, USA.

Scholarly Big Data: Information Extraction, Data Mining and Semantics *Keynote at Semantic Analysis of Documents Workshop*, Fort Collins, CO, USA.

Classifying and Ranking Search Engine Results as Potential Sources of Plagiarism *ACM Symposium on Document Engineering*, Fort Collins, CO, USA.

SimSeerX: A Similar Document Search Engine *ACM Symposium on Document Engineering*, Fort Collins, CO, USA.

Scholarly Big Data Information Extraction and Integration in the CiteSeerX Digital Library *10th International Workshop on Information Integration on the Web*, Chicago, IL, USA.

Generic Similar Document Search *Guest presentation for IST 441: Information Retrieval and Search Engines*, Pennsylvania State University, State College, PA, USA.

2013   Unsupervised Ranking for Plagiarism Source Retrieval *SIGComp Seminar at Penn State IST*, Pennsylvania State University, State College, PA, USA.

The Bleek and Lloyd Collection: Clicks, Computation and the Digital Humanities *Guest presentation at Penn State African Studies Program, Brown Bag Series*, Pennsylvania State University, State College, PA, USA.

Information and Communication Technology for Development in South Africa *Guest presentation for IST 402: Information and Communication Technologies for Development*, Pennsylvania State University, State College, PA, USA.

2012   Learning to Read Bushman: Automatic Handwriting Recognition for Bushman Texts *University of Cape Town-University of Stellenbosch Industry Seminar*, Centre for Broadband Networks, University of Cape Town, Cape Town, South Africa.

2011   Creating a Handwriting Recognition Corpus for Bushman Languages *13th International Conference on Asia-Pacific Digital Libraries*, Beijing, China.

The Digital Bleek and Lloyd Collection *Courage of ||Kabbo Conference*, University of Cape Town, Cape Town, South Africa.

Handwriting Recognition for Bushman Languages: Clicks, Problems and Preliminary Results *University of Cape Town-University of Stellenbosch Industry Seminar*, Centre for Broadband Networks, University of Cape Town, Cape Town, South Africa.

2010   Learning to Read Bushman *South African Institute for Computer Scientists and Information Technologists Conference Postgraduate Symposium*, Bela-Bela, South Africa.

Translating Handwritten Bushman Text *10th Annual International ACM/IEEE Joint Conference on Digital Libraries*, Gold Coast, Australia

Professional Service

Reviewing   World Wide Web Conference (WWW)
Special Interest Group on Information Retrieval (SIGIR)
Association for the Advancement of Artificial Intelligence (AAAI)
Web Search and Data Mining (WSDM)
North American Association of the ACL (NAACL)
Conference on Information and Knowledge Management (CIKM)
Joint Conference on Digital Libraries (JCDL)
Digital Libraries
Information Retrieval Journal
IEEE Computer

References

C. Lee Giles
David Reese Professor at the College of Information Sciences and Technology, Pennsylvania State University

College of Information Sciences and Technology
The Pennsylvania State University
E350 Westgate Building
University Park, PA 16802, USA

giles@ist.psu.edu
+1 814 865 7884

Hussein Suleman
Associate Professor in the Department of Computer Science, University of Cape Town

Department of Computer Science
University of Cape Town
Private Bag X3
Rondebosch, 7701, Cape Town, South Africa

hussein@cs.uct.ac.za
+27 21 650 5106

Imed Zitouni
Director of Engineering, Google
izitouni@google.com