JIAN WU

CONTACT INFORMATION

Office Address: 3202, ECSB, Old Dominion University, Norfolk, VA, 23529, USA
Office Phone: (757)683-7753
Email: jwu@cs.odu.edu
Homepage: https://www.cs.odu.edu/~jwu/

PROFESSIONAL EXPERIENCE

Assistant Professor   July 2018 – present
· Department of Computer Science, Old Dominion University, Norfolk, Virginia, United States

Assistant Teaching Professor   July 2017 – July 2018
· Information Sciences & Technology, Pennsylvania State University, University Park, Pennsylvania, United States

Post-doctoral Scholar   June 2011 – June 2017
· Information Sciences & Technology, Pennsylvania State University, University Park, Pennsylvania, United States

Data Scientist, Tech Director & DBA of CiteSeerX   June 2013 – July 2018
· Information Sciences & Technology, Pennsylvania State University, University Park, Pennsylvania, United States

Research Assistant   June 2006 – May 2011
· Astronomy & Astrophysics, Pennsylvania State University, University Park, Pennsylvania, United States

Teaching Assistant   August 2004 – May 2006
· Astronomy & Astrophysics, Pennsylvania State University, University Park, Pennsylvania, United States

EDUCATION

The Pennsylvania State University   *August, 2004 – August, 2011*
Ph.D. in Astronomy and Astrophysics. Thesis Advisors: D. P. Schneider & D. E. Vanden Berk
Ph.D. Minor in Computational Science
Summer School in Statistics for Astronomers V

The University of Science and Technology of China   *September, 2000 – July 2004*
B.S. in Physics and Astronomy

RESEARCH INTERESTS

Text mining, scholarly big data, natural language processing, natural language understanding, information retrieval, digital library search engines

AWARDS AND CERTIFICATES

Best short paper in the 2023 ACM/IEEE Joint Conference on Digital Libraries	*2023*
The Old Dominion University Alumni Association New Faculty Award	*2022*
British Computer Society Award 2021 for the Best Open Source Project	*2021*
ODU College of Science Outstanding Undergraduate Research Mentor Award	*2020*
JCDL 2018 Best Reviewer Award	*2018*
Online Teaching Certificate	*2017*
Best paper nomination in the 8th International Conference on Knowledge Capture	*2015*
Best application paper in the 26th Annual Conference on Innovative Applications of Artificial Intelligence	*2014*
Best paper nomination in the IEEE International Conference on Cloud Engineering	*2014*
Zaccheus Daniel Fund	*2009*
Zaccheus Daniel Fund	*2007*
Stephen B. Brumbach Fellowship	*2006*
USTC Excellent Graduate Student Award	*2004*

AWARDED EXTERNAL RESEARCH GRANTS

Conference: Doctoral Consortium Support for JCDL 2023 $15,893
April 19, 2023 – October 18, 2023 *NSF*

· **Jian Wu**(PI), Sampath Jayarathna

REU Site: Research Experiences for Undergraduates in Disinformation Detection and Analytics $324,000
March 1, 2022 — February 28, 2025 *NSF*

· Sampath Jayarathna, **Jian Wu** (Co-PI)

GPU Powered IoT Testbed to Expand CCI Lab Environment $148,740
February 1, 2022 — December 31, 2022 *Coastal Virginia Center of Cyber Innovation (COVA CCI)*

· Jiang Peng, John Pratt, **Jian Wu** (Co-PI), Rashid Khan

Web Science and Web Security $10,000
January 1, 2022 — August 30, 2022 *Coastal Virginia Center of Cyber Innovation (COVA CCI)*

· Michael Nelson, Michele Weigle, **Jian Wu** (Co-PI)

The Acceptance and Effectiveness of Explainable AI-Powered Credibility Labeler for Scientific Misinformation and Disinformation (#H-4Q21-014) $63,790
June 1, 2020 — May 31, 2022 *Virginia Commonwealth Cyber Initiative*

· **Jian Wu (PI)**, Jeremiah Still, Jiang Li

Topological Relations-Based Image Retrieval (#432672) $150,000
August 1, 2020 — July 31, 2023 *DoE (via LANL)*

· **Jian Wu (PI)**

NSF REU Supplement to Proposal #1823288: $16,000
August 1, 2020 — July 31, 2021 *NSF*

· Dr. Wu's budget: **$8000**
· C. Lee Giles (PI), **Jian Wu (Co-PI)**

Synthetic Prediction Markets with Algorithm Traders for Determining Experimental Reproducibility (#W911NF-19-2-0272) $2.9M
October 1, 2019 — March 31, 2022 *DARPA*

· Dr. Wu's budget: **$170,598**
· C. Lee Giles (PI), Christopher Griffin (Co-PI), Anna Squicciarini (Co-PI), Sarah Rajtmajer (Co-PI), Anthony Kwasnica (Co-PI), David M. Pennock (Senior Personnel), James Caverlee (Co-PI), Xia "Ben" Hu (Co-PI), Frank Shipman (Co-PI), **Jian Wu (Co-PI)**

Opening Books and the National Corpus of Graduate Research (LG-37-19-0078-19) $505,214
September 1, 2019 – August 31, 2022 *IMLS*

· Dr. Wu's budget: **$177,884**
· William Ingram (PI; Virginia Tech), Edward A. Fox (PI; Virginia Tech), **Jian Wu (Co-PI, ODU)**

CRI: CI-SUSTAIN: Collaborative Research: CiteSeerX: Toward Sustainable Support of Scholarly Big Data (#1823288) $770,000
August 1, 2018 – July 31, 2022 *NSF*

· Dr. Wu's budget: **$67,000**
· C. Lee Giles (PI, PSU), **Jian Wu (Co-PI, PSU)**

The Swift View of AGN: Insights from Multi-wavelength Variability $244,000
2012 *NASA*

· Jonathan Gelbord (PI), Caryl Gronwall, Daniel Vanden Berk, **Jian Wu (Co-PI)**

Exploring the X-ray Properties of the Highest-Luminosity Double-Peaked Emitters $45,000
2012 *NASA*

· Bin Luo (PI), W. Nielsen Brandt, **Jian Wu (Co-PI)**, Patrick Hall, Alireza Rafiee

IN-KIND GRANTS

PoC: OCR Feature Extraction for ETD Segmentation (#PC1S22HB1QRA4I7)   **$3105.00**
June 30, 2021 — September 30, 2021   *Amazon*
· **Jian Wu (PI)**

XSEDE Educational Resource Allocation (#CIS200001)   **$3125.25**
August 5, 2020 — August 4, 2021   *XSEDE*
· **Jian Wu (PI)**

INTRAMURAL GRANTS

Spring 2023 Undergraduate Research & Creativity Scholarship Program   **$2000**
2023   *ODU*
· **Dominik Soós (student advisee)**

Research Story-Line: Profiling, Organizing, and Tracking Scientific Publications   **$4000**
2022   *ODU Faculty Proposal Preparation Program (FP3)*
· **Jian Wu (PI)**

Toward Knowledge Extraction: Finding Datasets, Methods, and Tools Adopted In Research Papers
$10,000
2020   *ODU Program for Undergraduate Research and Scholarship (PURS)*
· **Jian Wu (PI)**

Knowledge Extraction From Research Papers and Electronic Medical Records   **$3000**
Spring, 2020   *ODU Junior Faculty Research Mentoring Program*
· **Jian Wu (Co-PI)**, Brenda Bradshaw (Co-PI), Michele Weigle (mentor), Holly Gaff (mentor)

Towards Detection of Fake Scientific News   **$3000**
June – August, 2019   *ODU Graduate Summer Award Program*
· Md Reshad Ul Hoque (graduate student in ECE), **Jian Wu (supervisor)**

Large Scale Keyphrase Evaluation via Crowdsourcing   **$3500**
May – August, 2018   *PSU Erickson Discovery Grant*
· Seth Wilcox (freshman in Education), **Jian Wu (supervisor)**

PUBLICATIONS AND RESEARCH REPORTS

Papers in Computer and Information Sciences and Technology:

1. William A. Ingram, **Jian Wu**, Sampanna Yashwant Kahu, Javaid Akbar Manzoor, Bipasha Banerjee, Aman Ahuja, Muntabir Hasan Choudhury, Lamia Salsabil, Winston Shields, and Edward A. Fox. "Building Datasets to Support Information Extraction and Structure Parsing from Electronic Theses and Dissertations." In International Journal on Digital Libraries (**IJDL**). 2024. [**Accepted**]

2. Sai Koneru, Xin Wei, **Jian Wu**, and Sarah M. Rajtmajer. "Can machine learning algorithms predict publication outcomes? A case study of COVID-19 preprints." In Proceedings of the 1st International Workshop on Artificial Intelligence for Science of Science (**AI4SciSci@ICDM 2023**), December 1, 2023. Shanghai, China.

3. Muntabir Hasan Choudhury, Lamia Salsabil, **Jian Wu**, William A. Ingram, Edward A. Fox. "ETDPC: A Multimodality Framework for Classifying Pages in Electronic Theses and Dissertations." In Proceedings of the 36th Innovative Applications of Artificial Intelligence Conference (**IAAI-2024**), February 20-27, 2024, Vancouver, Canada. [**Accepted**]

4. Kehinde Ajayi, Xin Wei, Diane Oyen, **Jian Wu**, Martin Gryder, Winston Shields, Shawn Jones, and Michal Kucer. "DeepPatent2: A Large-Scale Benchmarking Corpus for Technical Drawing Understanding." In (**Scientific Data**), 2023. `https://doi.org/10.1038/s41597-023-02653-7`

5. Heng Lian, Di Wu, Bo-Jian Hou, **Jian Wu**, Yi He. "Online Learning from Evolving Feature Spaces with Deep Variational Models." In Transactions on Knowledge and Data Engineering (**TKDE**), November, 2023. `https://doi.org/10.1109/TKDE.2023.3326365`

6. Michael Evans, Dominik Soós, Ethan Landers, **Jian Wu**. "MSVEC: A Multidomain Testing Dataset for Scientific Claim Verification." In *The Proceedings of the 8th National Workshop for REU Research in Networking and Systems* (**2023 REUNS**), October 23, 2023. Washington DC. USA. `https://doi.org/10.1145/3565287.3617630`

7. Escamilla, E., Salsabil, L., Klein, M., **Wu, J.**, Weigle, M.C., Nelson, M.L. (2023). "It's Not Just GitHub: Identifying Data and Software Sources Included in Publications." In: Linking Theory and Practice of Digital Libraries. (**TPDL 2023**). Lecture Notes in Computer Science, vol 14241. Springer, Cham. `https://:doi.org/10.1007/978-3-031-43849-3_17`

8. Xin Wei, Md Reshad Ul Hoque, **Jian Wu**, Jiang Li. "ClaimDistiller: Scientific Claim Extraction with Supervised Contrastive Learning." In *The Proceedings of the 4th Extraction and Evaluation of Knowledge Entities from Scientific Documents* (**EEKE 2023**) and *the 3rd AI + Informetrics* (**AII 2023**), June 26, 2023. Santa Fe, NM, USA. `https://ceur-ws.org/Vol-3451/paper11.pdf`

9. Muntabir Choudhury, Lamia Salsabil, Himarsha R. Jayanetti, **Jian Wu**, William A. Ingram, Edward A. Fox. "MetaEnhance: Metadata Quality Improvement for Electronic Theses and Dissertations of University Libraries." In *The Proceedings of the 2023 ACM/IEEE Joint Conference on Digital Libraries.* (**JCDL 2023**), June 26-30, 2023. Santa Fe, NM, USA. [**Best short paper**]. `https://doi.org/10.1109/JCDL57899.2023.00019.`

10. Kehinde Ajayi, Muntabir Choudhury, Sarah Rajtmajer, **Jian Wu**. "A Study on Reproducibility and Replicability of Table Structure Recognition Methods." In *The Proceedings of the 17th International Conference on Document Analysis and Recognition.* (**ICDAR 2023**), August 21-26, 2023. San José, CA. `https://doi.org/10.1007/978-3-031-41679-8_1`

11. Zeba Karishma, Shaurya Rohatgi, Kavya Shrinivas Puranik, **Jian Wu**, C. Lee Giles. "ACL-Fig: A Dataset for Scientific Figure Classification." In *The Proceedings of the 2023 AAAI Workshop on Scientific Document Understanding.* (**SDU 2023 @AAAI**), February 13-14, 2023. Washington DC.

12. Bipasha Banerjee, William A. Ingram, **Jian Wu**, and Edward A. Fox. "Applications of Data Analysis on Scholarly Long Documents. " In *The Proceedings of the 7th Computational Archival Science Workshop colocated with the 2022 IEEE Big Data Conference.* (**CAS 2022**), December 19, 2022. Virtual. DOI:10.1109/BigData55660.2022.10020935.

13. Md Reshad Ul Hoque, **Jian Wu**, Chiman Kwan, Krzysztof Koperski, Jiang Li. "ArithFusion: An Arithmetic Deep Model for Temporal Remote Sensing Image Fusion." In *Remote Sensing.* (**Remote Sens.**).

14. **Jian Wu**, Shaurya Rohatgi, Manoj K. Angadi, Kavya S. Puranik, C. Lee Giles. "Design Considerations for a Sustainable Scholarly Big Data Service." In *2022 Forum for Information Retrieaval Evaluation.* (**FIRE 2022**), December 9-13, 2022. Indian Statistical Institute, Kolkata, India.

15. **Jian Wu**, Dominik Soós, Ryan Hiltabrand, and Lee Giles. "Scholarly Big Data Quality Assessment: A Case Study of Document Linking and Conflation with S2ORC." In *2022 ACM Symposium on Document Engineering.* (**DocEng 2022**), September 20-23, 2022. Virtual. DOI:10.1145/3558100.3563850

16. Xin Wei, Lamia Salsabil, and **Jian Wu**. "Theory Entity Extraction for Social and Behavioral Sciences Papers using Distant Supervision." In *2022 ACM Symposium on Document Engineering.* (**DocEng 2022**), September 20-23, 2022. Virtual. DOI:10.1145/3558100.3563855

17. Heng Lian, John Scovil Atwood, Bojian Hou, **Jian Wu**, Yi He. "Online Deep Learning from Doubly-Streaming Data." In *ACM Multimedia Conference.* (**ACMMM 2022**), October 10-14, 2022. Lisbon, Portugal. DOI:10.1145/3503161.3548355

18. Xin Wei, **Jian Wu**, Kehinde Ajayi, Diane Oyen. "Visual Descriptor Extraction from Patent Figure Captions: A Case Study of Data Efficiency Between BiLSTM and Transformer." In *The Proceedings of the 2022 ACM/IEEE Joint Conference on Digital Libraries.* (**JCDL 2022**), June 20-24, 2022. Cologne, Germany. DOI:10.1145/3529372.3533299

19. Lamia Salsabil, **Jian Wu**, Muntabir Hasan Choudhury, William A. Ingram, Edward A. Fox, Sarah M. Rajtmajer, C. Lee Giles. "A Study of Computational Reproducibility using URLs Linking to Open Access Datasets and Software." In *The Proceedings of the 2nd International Workshop on Scientific Knowledge.* (**Sci-K 2021@WebConf**), April 26th, 2022. Online. DOI:10.1145/3487553.3524658

20. Md Reshad Ul Hoque, Xin Wei, Muntabir Choudhury, Kehinde Ajayi, Martin Gryder, **Jian Wu** and Diane Oyen. "Segmenting Technical Drawing Figures in US Patents." In *The Proceedings of the 2022 AAAI Workshop on Scientific Document Understanding.* (**SDU 2022 @AAAI**), February 28 or March 1, 2022, Virtual Event. http://ceur-ws.org/Vol-3164/paper28.pdf

21. Sami Uddin, Bipasha Banerjee, **Jian Wu**, William A. Ingram, Edward A. Fox. "Building A Large Collection of Multi-domain Electronic Theses and Dissertations." In *The Proceedings of the 2021 IEEE Big Data Conference.* (**IEEE BigData 2021**), December 15–18, Virtual Event. https://doi.org/10.1109/BigData52589.2021.9672058

22. William A. Ingram, **Jian Wu**, Edward A. Fox, Bipasha Banerjee. "Applications of Mining ETDs." In *The Proceedings of the 2021 Conferences on Electronic Theses and Dissertations.* (**ETD2021**), November 16–17, 2021, Virtual Event. DOI:10.26226/morressier.614c9b8c87a68d83cb5d59b2

23. Sarah Rajtmajer, Christopher Griffin, **Jian Wu**, Robert Fraleigh, Laxmaan Balaji, Anna Squicciarini, Anthony Kwasnica, David Pennock, Michael McLaughlin, Timothy Fritton, Nishanth Nakshatri, Arjun Menon, Sai Ajay Modukuri, Rajal Nivargi, Xin Wei and C. Lee Giles. "A synthetic prediction market for estimating confidence in published work." In *Proceedings of the 36th AAAI Conference on Artificial Intelligence.* (**AAAI 2022 Demo**), February 21–28, 2022, Vancouver, Canada. DOI:10.1609/aaai.v36i11.21733

24. **Jian Wu**, Shaurya Rohatgi, Sai Raghav Reddy Keesara, Jason Chhay, Kevin Kuo, Arjun Manoj Menon, Sean Parsons, Bhuvan Urgaonkar, C. Lee Giles. "Building an Accessible, Usable, Scalable, and Sustainable Service for Scholarly Big Data." In *Proceedings of the 2021 IEEE BigData Conference*, (**IEEE BigData 2021**), December 15–18. Virtual event. https://doi.org/10.1109/BigData52589.2021.9671612

25. Michal Kucer, Diane Oyen, **Jian Wu**, Juan Castorena. "DeepPatent: Large scale patent drawing recognition and retrieval." In *Proceedings of the 2022 Winter Conference on Applications of Computer Vision*, (**WACV 2022**), January 4–8, 2022. Waikoloa, HI. DOI: 10.1109/WACV51458.2022.00063

26. Shaurya Rohatgi, **Jian Wu**, C Lee Giles. "Ranked List Fusion and Re-ranking with Pre-trainedTransformers for ARQMath-2 Lab." In *Proceedings of the 12th International Conference of the CLEF Association*, (**CLEF 2021**), September 21–24, 2021. Virtual Event.

27. Athar Sefid, **Jian Wu**, Prasenjit Mitra, C. Lee Giles. "Extractive Research Slides Generation Using Windowed Labeling Ranking." In *the 2nd International Workshop on Scientific Document Processing*, (**SDP 2021 @NAACL**). Virtual Event.

28. Chinmayee Rane, Seshasayee Mahadevan Subramanya, Devi Sandeep Endluri, C. Lee Giles, **Jian Wu**. "ChartReader: Automatic Parsing of Bar-Plots." In *the IEEE 22nd International Conference on Information Reuse and Integration for Data Science*, (**IRI 2021**). Virtual Event.

29. Shaurya Rohatgi, **Jian Wu**, and C. Lee Giles. "What Were People Searching For? A Query Log Analysis of An Academic Search Engine." In *the 2021 ACM/IEEE Joint Conference on Digital Libraries*, (**JCDL 2021**). September 27–30, 2021. Virtual Event.

30. Muntabir Hasan Choudhury, Himarsha R. Jayanetti, **Jian Wu**, William A. Ingram, Edward A. Fox. "Automatic Metadata Extraction Incorporating Visual Features from Scanned Electronic Theses and Dissertations." In *the 2021 ACM/IEEE Joint Conference on Digital Libraries*, (**JCDL 2021**). September 27–30, 2021. Virtual Event.

31. Sampanna Yashwant Kahu, William A. Ingram, Edward A. Fox, **Jian Wu**. "ScanBank: A Benchmark Dataset for Figure Extraction from Scanned Electronic Theses and Dissertations." In: *the 2021 ACM/IEEE Joint Conference on Digital Libraries*, (**JCDL 2021**). September 27–30, 2021. Virtual Event.

32. Meng Ling, Jian Chen, Torsten Moller, Petra Isenberg, Tobias Isenberg, Michael Sedlmair, Robert Laramee, Han-Wei Shen, **Jian Wu**, C. Lee Giles. "Document Domain Randomization for Deep Learning Document Layout Extraction." In: *the 2021 International Conference on Document Analysis and Recognition*, (**ICDAR 2021**). Lausanne, Switzerland.

33. Sree Sai Teja Lanka, Sarah Michele Rajtmajer, **Jian Wu**, C. Lee Giles. "Extraction and Evaluation of Statistical Information from Social and Behavioral Science Papers." In: *the 1st International Workshop on Scientific Knowledge*, (**Sci-K 2021**). April 13, 2021. Virtual Event.

34. Wei Wang, Feng Xia, **Jian Wu**, Zhiguo Gong, Hanghang Tong, Brian D. Davison. "Scholar2vec: Vector Representation of Scholars for Lifetime Collaborator Prediction." In: *Transactions on Knowledge Discovery from Data*, (**TKDE 2021**). DOI:10.1145/3442199

35. Sai Ajay Modukuri, Sarah Rajtmajer, Anna Cinzia Squicciarini, **Jian Wu**, C. Lee Giles. "Understanding and Predicting Retractions of Published Work." In: *The 1st AAAI Workshop on Scientific Document Understanding @ AAAI 2021*, (**SDU 2021**), Virtual Event.

36. Ming Gong, Xin Wei, Diane Oyen, **Jian Wu**, Martin Gryder, Liping Yang. "Recognizing Figure Labels in Patents." In: *The 1st AAAI Workshop on Scientific Document Understanding @ AAAI 2021*, (**SDU 2021**), Virtual Event.

37. Yasith Jayawardana, Alexander C. Nwala, Gavindya Jayawardena, **Jian Wu**, Sampath Jayarathna, Michael L. Nelson, C. Lee Giles. "Modeling Updates of Scholarly Webpages Using Archived Data." In: *Proceedings of the Computational Archival Science Workshop @ BigData 2020*, (**CAS 2020**) 2020, Virtual Event.

38. Bharath K. Kandimalla, Shaurya Rohatgi, **Jian Wu**, and C. Lee Giles. "Large Scale Subject Category Classification of Scholarly Papers with Deep Attentive Neural Networks." In: *Frontiers in Research Metrics and Analytics, section Text-mining and Literature-based Discovery*, (**Front. Res. Metr. Anal.**), 2020, https://doi.org/10.3389/frma.2020.600382

39. **Jian Wu**, Pei Wang, Xin Wei, Sarah Rajtmajer, C Lee Giles, and Christopher Griffin. "Acknowledgement Entity Recognition in CORD-19 Papers." In: *Proceedings of the 1st Workshop on Scholarly Document Processing @ EMNLP 2020*, (**SDP 2020**), November 19, 2020. Virtual Event.

40. Shaurya Rohatgi, **Jian Wu**, and C. Lee Giles. "PSU at CLEF-2020 ARQMath Track: Unsupervised Re-ranking using Pretraining." In: *Proceedings of the 10th International Conference of the CLEF Association*, (**CLEF 2020**), September 22–25, 2020. Virtual Event.

41. Shaurya Rohatgi, Zeba Karishma, Jason Chhay, Sai Raghav Reddy Keesara, **Jian Wu**, Cornelia Caragea, and C Lee Giles. "COVIDSeer : Extending the CORD-19 Dataset." In: *Proceedings of the 20th ACM Symposium on Document Engineering*, (**DocEng 2020**), September 29–October 2, 2020. Virtual Event.

42. Chenrui Guo, Haoran Cui, Li Zhang, Jiamin Wang, Wei Lu, **Jian Wu**. "SmartCiteCon: Implicit Citation Context Extraction from Academic Literature Using Supervised Learning." In: *Proceedings of the 2020 International Workshop on Mining Scientific Publications*, (**WOSP 2020**), August 5, 2020. Virtual Event.

43. Krutarth Patel, Cornelia Caragea, **Jian Wu**, C. Lee Giles. "Keyphrase Extraction in Scholarly Digital Library Search Engines." In: *Proceedings of 2020 International Conference on Web Services*, (**ICWS 2020**), August 12–14, 2020. Honolulu, Hawaii, USA. https://doi.org/10.1007/978-3-030-59618-7_12

44. **Jian Wu**, Md Reshad Ul Hoque, Gunnar W. Reiske, Michele C. Weigle, Brenda T. Bradshaw, Holly D. Gaff, Jiang Li, Chiman Kwan. "A Comparative Study of Sequence Tagging Methods for Domain Knowledge Entity Recognition in Biomedical Papers." In: *Proceedings of 2020 Joint Conference on Digital Libraries* (**JCDL 2020**), August 1–5, 2020. Virtual Event.

45. Rui Wang, Xiaoling Zhou, **Jian Wu**, Qing Yin, Ou Wu. "Inter-subdiscipline Analysis based on Mathematical Statements." Poster in: *Proceedings of the 2020 International Conference on Digital Libraries* (**JCDL 2020**), August 1–5, 2020. Virtual Event.

46. Gavindya Jayawardena, Sampath Jayarathna, **Jian Wu**. "Analyzing the Effect of Reading Patterns using Eye Tracking Measures." Poster in: *Proceedings of the 2020 International Conference on Digital Libraries* (**JCDL 2020**), August 1–5, 2020. Virtual Event. [**Best Poster**]

47. Muntabir Hasan Choudhury, **Jian Wu**, William A. Ingram, Edward A. Fox. "A Heuristic Baseline Method for Metadata Extraction from Scanned Electronic Theses and Dissertations." Poster in: *Proceedings of the 2020 International Conference on Digital Libraries* (**JCDL 2020**), August 1–5, 2020. Virtual Event. [**Best Poster Honorable Mention**]

48. **Jian Wu** and C. Lee Giles. "Scholarly Very Large Data: Challenges for Digital Libraries." White paper in: *The 2020 Large Scale Networking Workshop on Huge Data*, April 12–13, 2020, Virtual Event.

49. Wei Zhong, Shaurya Rohatgi, **Jian Wu**, C. Lee Giles, Richard Zanibbi. "Accelerating Substructure Similarity Search for Formula Retrieval." In: *the 42nd European Conference on Information Retrieval* (**ECIR 2020**), April 14–17, 2020, Lisbon, Portugal. `https://doi.org/10.1007/978-3-030-45439-5_47`

50. Ruijing Yao, Linlin Hou, Yingchun Ye, Ou Wu, Ji Zhang, and **Jian Wu**. "Method and Dataset Mining in Scientific Papers." Poster in: *2019 IEEE International Conference on Big Data* (**Bigdata 2019**), December 9–12, 2019, Los Angeles, CA, USA.

51. Athar Sefid, **Jian Wu**, Prasenjit Mitra, and C. Lee Giles. "Automatic Slide Generation for Scientific Papers." In: *The 3rd International Workshop on Capturing Scientific Knowledge* (**SciKnow 2019**), November 21, 2019, Marina del Rey, CA, USA.

52. Md Reshad Ul Hoque, **Jian Wu**, Jiang Li, Chiman Kwan, Agnese Chiatti, and Dash Bradley. "Searching for Evidence of Scientific News in Scholarly Big Data." In: *Proceedings of the 10th International Conference on Knowledge Capture* (**K-CAP 2019**), November 19-21, 2019, Marina del Rey, CA, USA.

53. **Jian Wu**, Kunho Kim, and C. Lee Giles. "CiteSeerX: 20-year Service for Scholarly Big Data." In: *Proceedings of the 2019 Artificial Intelligence for Data Discovery and Reuse* (**AIDR 2019**), May 13-15, 2019, Pittsburgh, PA, USA.

54. Nir Nissim, Aviad Cohen, **Jian Wu**, Andrea Lanzi, Lior Rokach, Yuval Elovici, and C. Lee Giles. "Sec-Lib: Protecting Scholarly Libraries from Infected Papers Using Active Machine Learning Framework." In: **IEEE Access**. 2019.

55. Behrooz Mansouri, Shaurya Rohatgi, Douglas Oard, **Jian Wu**, C. Lee Giles, and Richard Zanibbi. "Tangent-CFT: an Embedding model for Mathematical Formulas." In: *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval* (**ICTIR 2019**), October 2-5, 2019, Santa Clara, CA, USA.

56. Alexander G. Ororbia II, Ankur Mali, Jian Wu, Scott O'Connell, William Dreese, David Miller, and C. Lee Giles. "Learned Neural Iterative Decoding for Lossy Image Compression Systems." In: *Proceedings of the 2019 Data Compression Conference* (**DCC 2019**), March 26-29, 2019, Snowbird, UT, USA.

57. Athar Sefid, **Jian Wu**, Allen C. Ge, Jing Zhao, Lu Liu, Cornelia Caragea, Prasenjit Mitra, C. Lee Giles. "Cleaning Noisy and Heterogeneous Metadata for Record Linking Across Scholarly Big Datasets." In: *Proceedings of the 31st Innovative Applications of Artificial Intelligence Conference* (**IAAI 2019**), January 29-31, 2019, Honolulu, Hawaii, USA.

58. **Jian Wu**, Bharath Kandimalla, Shaurya Rohatgi, Athar Sefid, Jianyu Mao, C. Lee Giles. "CiteSeerX-2018: A Cleansed Multidisciplinary Scholarly Big Dataset." Poster Exhibition in: *the 2018 IEEE International Conference on Big Data* (**Big Data 2018**), Seattle, WA, USA, 2018.

59. **Jian Wu**, Athar Sefid, Allen C. Ge, and C. Lee Giles. "A Supervised Learning Approach To Entity Matching Between Scholarly Big Datasets." In: *Proceedings of the 9th International Conference on Knowledge Capture* (**K-CAP 2017**), December 4-6, 2017, Austin, Texas, USA.

60. Hung-Hsuan Chen, **Jian Wu**, C. Lee Giles. "Compiling Keyphrase Candidates for Scientific Literature Based on Wikipedia. " In: *(meta)-data quality workshop part of the 21st International Conference on Theory and Practice of Digital Libraries* (**MDQual 2017**), September 18-21, 2017, Thessaloniki, Greece

61. Yufeng Ma, Tingting Jiang, Chandani Shrestha, Edward A. Fox, **Jian Wu**, and C. Lee Giles. "Scenarios for Advanced Services in an ETD Digital Library. " *In press for ETD2017, the 20th international symposium on electronic theses and dissertations*, August 7-9, 2017, Washington, DC, USA.

62. **Jian Wu**, Sagnik Ray Choudhury, Agnese Chiatti, Chen Liang, and C. Lee Giles. "HESDK: A Hybrid Approach to Extracting Scientific Domain Knowledge Entities." In: *Proceedings of ACM/IEEE-CS Joint Conference on Digital Libraries* (**JCDL 2017**), June 19-23, 2017, Toronto, Canada.

63. Nir Nissim, Aviad Cohen, **Jian Wu**, Andrea Lanzi, Lior Rokach, Yuval Elovici, and Lee Giles. "Scholarly Digital Libraries as a Platform for Malware Distribution." In: *Proceedings of the 2nd Singapore Cyber Security R&D Conference* (**SG-CRC 2017**), Singapore.

64. **Jian Wu**, Chen Liang, Huaiyu Yang, and C. Lee Giles. "CiteSeerX data: semanticizing scholarly papers." In: *Proceedings of the International Workshop on Semantic Big Data* (**SIGMOD-SBD 2016**), June 26-30, 2016, San Francisco, CA, USA.

65. Kyle Williams, **Jian Wu**, Zhaohui Wu, and C. Lee Giles. "Information Extraction for Scholarly Digital Libraries." In: *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries* (**JCDL 2016**), Newark, NJ, USA.

66. Cornelia Caragea, **Jian Wu**, Sujatha Das G., and C. Lee Giles. "Document Type Classification in Online Digital Libraries." In: *Proceedings of the 2016 AAAI Conference on Artificial Intelligence* (**AAAI 2016**), February 12-17, 2016, Phoenix, AZ, USA.

67. **Jian Wu**, Kyle Williams, Hung-Hsuan Chen, Madian Khabsa, Cornelia Caragea, Suppawong Tuarob, Alexander Ororbia, Douglas Jordan, Prasenjit Mitra, and C. Lee Giles. "CiteSeerX: AI in a Digital Library Search Engine." *Artificial Intelligence Magazine (AI Magazine)*, 2015.

68. **Jian Wu** and C. Lee Giles. "Information Extraction for Scholarly Document Big Data." In: *The 1st International Workshop on Capturing scientific knowledge*, co-located with K-CAP 2015, (**SciKnow 2015**), October 7-10, 2015, Palisades, NY.

69. **Jian Wu**, Jason Killian, Huaiyu Yang, Kyle Williams, Sagnik Ray Choudhury, Suppawong Tuarob, Cornelia Caragea, and C. Lee Giles. "PDFMEF: A Multi-Entity Knowledge Extraction Framework for Scholarly Documents and Semantic Search." In: *Proceedings of the 8th International Conference on Knowledge Capture* (**K-CAP 2015**), October 7-10, 2015, Palisades, NY, USA. [**Best Paper Nomination**]

70. Alexander G. Ororbia II, David Reitter, **Jian Wu**, and C. Lee Giles. "Online Learning of Deep Hybrid Architectures for Semi-Supervised Categorization." In: *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (**ECML PKDD 2015**), Porto, Portugal.

71. Alexander G. Ororbia II, **Jian Wu**, and C. Lee Giles. "Big Scholarly Data in CiteSeerX: Information Extraction from the Web." In: *The 2nd WWW Workshop on Big Scholarly Data: Towards the Web of Scholars* (**BigScholar 2015**), Florence, Italy, 2015.

72. Alexander G. Ororbia II, **Jian Wu**, and C. Lee Giles. "CiteSeerX: Intelligent Information Extraction and Knowledge Creation from Web-Based Data." In: *The 4th Workshop on Automated Knowledge Base Construction at NIPS 2014*, (**AKBC 2014**), Montréal, Canada, 2014.

73. **Jian Wu**, Alexander 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: *The 9th USENIX International Workshop on Feedback Computing* (**Feedback 2014**), Philadelphia, PA, USA, 2014.

74. **Jian Wu**, Kyle Williams, Madian Khabsa and C. Lee Giles. "The Impact of User Corrections on A Crawl-Based Digital Library: A CiteSeerX Perspective." In: *The 10th International Conference on Collaborative Computing: Networking, Applications and Worksharing* (**CollaborateCom 2014**), Miami, Florida, USA. [Invited Paper]

75. Kyle Williams, **Jian Wu**, and C. Lee Giles. "SimSeerX: A Similar Document Search Engine". In: *The 14th ACM Symposium on Document Engineering* (**DocEng 2014**), Fort Collins, CO, USA.

76. **Jian Wu**, Kyle Williams, Hung-Hsuan Chen, Madian Khabsa, Cornelia Caragea, Alexander Ororbia, Douglas Jordan and C. Lee Giles. "CiteSeerX: AI in a Digital Library Search Engine". In: *The 26th Annual Conference on Innovative Applications of Artificial Intelligence* (**IAAI 2014**), Québec City, Québec, Canada. [**Best Application Paper**]

77. Kyle Williams, **Jian Wu**, Sagnik Choudhury, Madian Khabsa, and C. Lee Giles. "Scholarly Big Data Information Extraction and Integration in the CiteSeerX Digital Library." In: *The 10th International Workshop on Information Integration on the Web* (**IIWeb 2014**), Chicago, IL, USA, 2014.

78. Zhaohui Wu, **Jian Wu**, Madian Khabsa, Kyle Williams, Hung-Hsuan Chen, Wenyi Huang, Suppawong Tuarob, Sagnik Ray Choudhury, Alexander Ororbia, Prasenjit Mitra, and C. Lee Giles. "Towards Building a Scholarly Big Data Platform: Challenges, Lessons and Opportunities". In: *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries* (**DL 2014**), London, UK.

79. Kyle Williams, Lichi Li, Madian Khabsa, **Jian Wu**, Patrick Shih, and C. Lee Giles. "A Web Service for Scholarly Big Data Information Extraction". In: *The 21th IEEE International Conference on Web Services* (**ICWS 2014**), Anchorage, AK, USA.

80. **Jian Wu**, Pradeep Teregowda, Kyle Williams, Madian Khabsa, Douglas Jordan, Eric Treece, Zhaohui Wu

and C. Lee Giles. "Migrating A Digital Library to A Private Cloud". In: *Proceedings of the IEEE International Conference on Cloud Engineering 2014* (**IC2E 2014**), Boston, MA, USA. [**Best Paper Nomination**]

81. Cornelia Caragea, **Jian Wu**, Alina Ciobanu, Kyle Williams, Juan Fernández-Ramírez, Hung-Hsuan Chen, Zhaohui Wu and C. Lee Giles. "CiteSeerX: A Scholarly Big Dataset". In: *Proceedings of the 36th European Conference on Information Retrieval* (**ECIR 2014**), Amsterdam, Netherlands.

82. Cornelia Caragea, **Jian Wu**, Kyle Williams, Sujatha Das G., Madian Khabsa, Pradeep Teregowda and C. Lee Giles. "Automatic Identification of Research Articles from Crawled Documents." In: *WSDM 2014 Workshop on Web-scale Classification: Classifying Big Data from the Web* (**WSCBD 2014**), New York City, NY, USA, 2014.

83. **Jian Wu**, Pradeep Teregowda, Eric Treece, Madian Khabsa, Douglas Jordan, Stephen Carman, Prasenjit Mitra and C. Lee Giles. 2013. "Scalability Bottlenecks of the CiteSeerX Digital Library Search Engine." In: *the 10th International Workshop on Large-Scale and Distributed System for Information Retrieval* co-located with WSDM 2013 (**LSDS-IR 2013**), Rome, Italy, 2013.

84. **Jian Wu**, Pradeep Teregowda, Madian Khabsa, Stephen Carman, Douglas Jordan, Jose San Pedro Wandelmer, Xin Lu, Prasenjit Mitra, and C. Lee Giles. "Web crawler middleware for search engine digital libraries: a case study for citeseerX." In: *Proceedings of the 12th international workshop on Web information and data management*, collocated with CIKM 2012 (**WIDM 2012**), Maui, HI, USA, 2012.

85. **Jian Wu**, Pradeep Teregowda, Juan Pablo Fernández Ramírez, Prasenjit Mitra, Shuyi Zheng, and C. Lee Giles. "The evolution of a crawling strategy for an academic document search engine: whitelists and blacklists." In: *Proceedings of the 3rd Annual ACM Web Science Conference* (**WebSci 2012**), Evanston, IL, USA.

86. Sumit Bhatia, Cornelia Caragea, Hung-Hsuan Chen, **Jian Wu**, Pucktada Treeratpituk, Zhaohui Wu, Madian Khabsa, Prasenjit Mitra and C. Lee Giles. "Specialized Research Datasets in the CiteSeerX Digital Library." *D-Lib Magazine*, Vol. 18, Number 7/8, 2012.

Papers in Astronomy & Astrophysics:

87. B. Luo, W.N. Brandt, M. Eracleous, **Jian Wu**, P.B. Hall, A. Rafiee, D.P. Schneider, Jianfeng Wu. "X-ray and multiwavelength insights into the inner structure of high-luminosity disc-like emitters." In: *Monthly Notices of the Royal Astronomical Society (MNRAS)*, 2013, 429, 1479.

88. **Jian Wu**, Daniel Vanden Berk, Dirk Grupe, Scott Koch, Jonathan Gelbord, Donald P. Schneider, Caryl Gronwall, Sarah Wesolowski, and Blair L. Porterfield. "A Quasar Catalog with Simultaneous UV, Optical and X-ray Observations by *Swift*." In: *Astrophysical Journal Supplement (ApJS)*, 2012, 201:10.

89. **Jian Wu**, Jane C. Charlton, Toru Misawa, Michael Eracleous, and Rajib Ganguly. "The Physical Conditions of the Intrinsic NV Narrow Aborption Line Systems of Three Quasars." In: *The Astrophysical Journal (ApJ)*, 2010, 722:997.

90. **Jian Wu**, D.E. Vanden Berk, W.N. Brandt, D.P. Schneider, R.R. Gibon, and Jianfeng Wu. "Probing Origins of the C IV λ1549 and Fe Kα Baldwin Effect." In: *The Astrophysical Journal (ApJ)*, 2009, 702:767.

91. Donald P. Schneider, Patrick B. Hall, Gordon T. Richards, Michael A. Strauss, Daniel E. Vanden Berk, Scott F. Anderson, W. N. Brandt, Xiaohui Fan, Sebastian Jester, Jim Gray, James E. Gunn, Mark U. SubbaRao, Anirudda R. Thakar, Chris Stoughton, Alexander S. Szalay, Brian Yanny, Donald G. York, Neta A. Bahcall, J. Barentine, Michael R. Blanton, Howard Brewington, J. Brinkmann, Robert J. Brunner, Francisco J. Castander, István Csabai, Joshua A. Frieman, Masataka Fukugita, Michael Harvanek, David W. Hogg, Zeljko Ivezic, Stephen M. Kent, S. J. Kleinman, G. R. Knapp, Richard G. Kron, Jurek Krzesinski, Daniel C. Long, Robert H. Lupton, Atsuko Nitta, Jeffrey R. Pier, David H. Saxe, Yue Shen, Stephanie A. Snedden, David H. Weinberg, and **Jian Wu**. "The Sloan Digital Sky Survey Quasar Catalog. IV. Fifth Data Release." In: *The Astronomical Journal (AJ)*, 2007, 134:102.

92. Yu Lu, Tinggui Wang, Hongyan Zhou, **Jian Wu**. "On the Selection Effect of Radio Quasars in the Sloan Digital Sky Survey." In: *The Astronomical Journal (AJ)*, 2007, 133:1615.

Abstracts and Posters:

1. Dominik Soós, Ethan Landers, and **Jian Wu**. "SciNVe: A Generalizability Test Dataset for Scientific News Verification." In: *the 2023 CCI Symposium*. April 17–18, 2023. Richmond, VA, USA.

2. William A. Ingram, **Jian Wu**, and Edward A. Fox. "Maximizing Equitable Reach and Accessibility of ETDs." In: *the 2023 ACM/IEEE-CS Joint Conference on Digital Libraries*. **JCDL 2023**, June 26 – June 30, 2023. Santa Fe, NM, USA.

3. William A. Ingram, **Jian Wu**, and Edward A. Fox. "Electronic Theses and Dissertations: A Research Corpus of Scholarly Big Data." Poster Exhibition in: *the 24th International Conference on Grey Literature*, **GL2022**, December 5, 2022. Virtual.

4. Sarah Rajtmajer, Christopher Griffin, **Jian Wu**, Robert Fraleigh, Laxmaan Balaji, Anna Squicciarini, Anthony Kwasnica, David Pennock, Michael McLaughlin, Timothy Fritton, Nishanth Nakshatri, Arjun Menon, Sai Ajay Modukuri, Rajal Nivargi, Xin Wei, C. Lee Giles. "A Synthetic Prediction Market for Estimating Confidence in Published Work." Poster Exhibition in: *1st International Conference of Science & Innovation*, **ICSSI 2022**, June 7-9, 2022. National Academy of Sciences, Washington, D.C., USA

5. **Jian Wu**, Saad Hasnain Khan, Winston Shield, MD Rashed Ul Hoque, Jiang Li, Morgan Edwards, Jeremiah Still. "SciPEP: an AI-powered Scientific Misinformation Labeler – A Prototype Design." Poster Exhibition in: *2022 CCI Symposium*. April 4-5, 2022. Richmond, VA.

6. Ryan Hiltabrand, **Jian Wu**. "Quality Assessment of Scholarly Big Data." Poster Exhibition in: *2022 ODU Undergraduate Research Symposium*, Perry Library, Old Dominion University. March 19, 2022.

7. Kunho Kim, and C. Lee Giles. "Financial Entity Record Linkage with Random Forests." Poster Exhibition in: *SIGMOD Workshop on Data Science for Macro-Modeling with Financial and Economic Datasets*, San Francisco, CA, USA, 2016

8. **Jian Wu**, Jason Killian, Huaiyu Yang, Kyle Williams, Sagnik Ray Choudhury, Suppawong Tuarob, Cornelia Caragea, and C. Lee Giles. "PDFMEF: A Multi-Entity Knowledge Extraction Framework for Scholarly Documents and Semantic Search." Poster Exhibition in: *the 8th Annual Postdoc Research Exhibition*, Nittany Lion Inn, University Park, PA, USA, 2015.

9. **Jian Wu**, C. Lee Giles, and Shuyi Zheng. "Improving the Efficiency of CiteSeerX Crawler." In: *Poster Exhibition for Pennsylvania State University CyberInfrastructure Day*, University Park, PA, USA, 2011.

10. **Jian Wu**, Daniel E. Vanden Berk, and Donald P. Schneider. "The UV Properties of A Set of Composite Quasar Spectra and the Baldwin Effect." In: *Bulletin of the American Astronomical Society (the 215th American Astronomical Society, AAS Meeting)*, Vol. 42, p.582, 2010.

11. **Jian Wu**, Daniel E. Vanden Berk, and Donald P. Schneider. "Continuum and Emission Line Properties of Composite Quasar Spectra from SDSS." In: *Bulletin of the American Astronomical Society (the 215th American Astronomical Society, AAS Meeting)*, Vol. 39, p.796, 2007.

12. **Jian Wu**, Jane C. Charlton, Toru Misawa. "Modelling Three Quasar Intrinsic Narrow Absorption Line Systems." In: *Bulletin of the American Astronomical Society (the 207th American Astronomical Society, AAS Meeting)*, Vol. 37, p.1186, 2005.

Research Reports

1. Md Reshad Ul Hoque, Jiang Li, **Jian Wu**. "SciEv: Finding Scientific Evidence Papers for Scientific News." In: *arXiv:2205.00126*, 2022

2. **Jian Wu**, Rajal Nivargi, Sree Sai Teja Lanka, Arjun Manoj Menon, Sai Ajay Modukuri, Nishanth Nakshatri, Xin Wei, Zhuoer Wang, James Caverlee, Sarah M. Rajtmajer, and C. Lee Giles. "Predicting the Reproducibility of Social and Behavioral Science Papers Using Supervised Learning Models." In: *arXiv.2104.04580*, 2021.

3. Alipourfard, N., Arendt, B., Benjamin, D. M., Benkler, N., Bishop, M. M., Burstein, M., Bush, M., Caverlee, J., Chen, Y., Clark, C., Dreber Almenberg, A., Errington, T. M., Fidler, F., Fox, N., Frank, A., Fraser, H., Friedman, S., Gelman, B., Gentile, J., Giles, C. L., Gordon, M. B., Gordon-Sarney, R., Griffin, C., Gulden, T., Hahn, K., Hartman, R., Holzmeister, F., Hu, X. B., Johannesson, M., Kezar, L., Kline Struhl, M., Kuter, U., Kwasnica, A. M., Lee, D-H., Lerman, K., Liu, Y., Loomas, Z., Luis, B., Magnusson, I., Miske, O., Mody, F., Morstatter, F., Nosek, B. A., Parsons, E. S., Pennock, D., Pfeiffer, T., Pi, H., Pujara, J., Rajtmajer, S., Ren, X., Salinas, A., Selvam, R. K., Shipman, F., Silverstein, P., Sprenger, A., Squicciarini, A. M., Stratman, S., Sun, K., Tikoo, S., Twardy, C. R., Tyner, A., Viganola, D., Wang, J., Wilkinson, D. P., Wintle, B., & **Wu, J.** (2021). Systematizing Confidence in Open Research and Evidence (SCORE). *SocArXiv*. https://doi.org/10.31235/osf.io/46mnb

4. Alexander G. Ororbia, Ankur Mali, **Jian Wu**, Scott O'Connell, David Miller, and C. Lee Giles. "Learned Iterative Decoding for Lossy Image Compression Systems." In: *arXiv,1803.05863*, 2018.

5. Jonathan Gelbord, Caryl Gronwall, Dirk Grupe, Daniel E. Vanden Berk, and **Jian Wu**. "Exploring Multiwavelength AGN Variability with Swift Archival Data." In: *arXiv, 1505.05248*, 2015.

6. **Jian Wu** and Daniel E. Vanden Berk. "Fitting the Continuum Component of A Composite SDSS Quasar Spectrum Using CMA-ES." In: *arXiv, astro-ph.1312.7356*, 2013.

Ph.D. Thesis:

· Title: "A study of UV, Optical and X-ray Emission from quasars.", 2011
· Contribution: The work utilizes big data, compiled from archives of space observatories and ground based telescopes, to investigate important correlations between physical parameters of Active Galactic Nuclei and quasars.
· Thesis committee: Donald P. Schneider (chair), Jane C. Charlton (co-chair), Yuexing C. Li, William N. Brandt, Daniel E. Vanden Berk (special member), Paul Sommers (external member)

System Documentation

· **Jian Wu**, et al. "CiteSeerX Internal Operation Manual." 200+ pages.

Media Reports

· ODU – August 31, 2023. President to Host Sept. 8 Symposium on Artificial Intelligence in Education. `https://www.odu.edu/article/president-to-host-sept-8-symposium-artificial-intelligence-education`
· 13NewsNow – August 22, 2023. A media report on "More teachers plan to use artificial intelligence in classrooms." `https://tinyurl.com/jianwuaiteaching`
· WTKR-NFK (CBS) – NewsChannel 3 News at 6, Tuesday, February 23, 2023. A Workshop on "How Are Misinformation and Disinformation Related to You? – Perspectives from Researchers and Educators." `https://bit.ly/41XJvZl`.
· Joe Garvey. "ODU Recognizes Outstanding Faculty, Administrators at Awards Dinner." News@ODU. April 25, 2022. `https://www.odu.edu/news/2022/4/faculty_and_administ#.YzrDiuzMJHs`
· Tiffany Whitfield. "Computer Scientist Earns Young Professor Award." Discovery Monarch Research. Summer, 2022. `https://indd.adobe.com/view/f94cf009-aeee-4063-8f59-2525ad499882`
· Jessica Hallman. "IST academic search engine awarded 'Best Open Source Project' by BCS". IST Story. April 19, 2022. https://www.psu.edu/news/information-sciences-and-technology/story/ist-academic-search-engine-awarded-best-open-source/
· Tiffany Whitfield. "ODU Computer Science Professor Receives British Computer Society Award for Pioneering Platform". Inside ODU. April 21, 2022. `https://www.odu.edu/about/odu-publications/insideodu/2022/04/21/kudos1`
· Bethan Davies. "Research Reproducibility Concerns Can Be Resolved Using AI Betting Agents". AZORobotics. March 18, 2022. `https://www.azorobotics.com/News.aspx?newsID=12822`
· Penn State. "Scientists tap AI betting agents to help solve research reproducibility concerns". **EurekAlert! AAAS**. March 17, 2022. `https://www.eurekalert.org/news-releases/946747`
· Sherry DiBari. "Professors Study Cybersecurity's Role in Curbing the Spread of Misinformation". **News@ODU** August, 13, 2022. `https://www.odu.edu/news/2021/8/cybersecurity#.YkW5NDfMKgU`

CONFERENCE AND SEMINAR TALKS

1. "Towards Automatically Understanding Scientific Papers." Invited by Bill Ingram at Virginia Tech. November, 2022.

2. "Query Log Analysis of Academic Search At Scale." In *Introduction to Information Retrieval*. November 8, 2022. Virtual. Invited by Dr. Behrooz Mansouri at South Maine University.

3. "Fake News Detection – An Introduction." In *2022 ODU REU summer camp*. July 15, 2022.

4. "Fundamentals of Machine Learning." In *2022 ODU REU summer camp*. June 29, 2022.

5. "Towards Training Computers To Understand Human Language." In *2021 ODU CS Data Science Summer camp*. August, 2021.

6. **Jian Wu**, William A. Ingram, Edward A. Fox. "Towards Aiding Research by Improving Access to Electronic Theses and Dissertations from Multiple Domains." In *CNI Fall 2021 Virtual Membership Meeting.* (**CNI 2021**). December 7–9, 2021.

7. "CiteSeerX: Mining Scholarly Big Data." *Invited talk*, presented at

   - School of Software Engineering, Dalian University of Technology, Dalian, Liaoning, China, 2019
   - School of Information Management, Wuhan University, Wuhan, Hubei, China, 2019
   - MITRE Corporation, Hampton, VA, April, 2019
   - Department of Electrical & Computer Engineering, Old Dominion University, April, 2019
   - Computer Science Department, Old Dominion University, November, 2018

8. "CiteSeerX: 20-year Service for Scholarly Big Data." *Invited talk*, presented at

   - School of Software Engineering, Dalian University of Technology, Dalian, Liaoning, China, 2019
   - School of Information Management, Wuhan University, Wuhan, Hubei, China, 2019

9. "CiteSeerX: AI in A Digital Library Search Engine." *Invited Speaker by Mary Beth Rossen for IST590.* College of Information Sciences and Technology, Pennsylvania State University, University Park, PA, December, 2018.

10. "Automatic Identification of Research Articles from Crawled Documents." *Special Interest Group in Informatics Computation luncheon seminar*, IST, Pennsylvania State University, University Park, PA, USA, 2015.

11. "Experiences with Large Scale Information Extraction for Scholarly Big Data." *Invited talk for the Open Syllabus Project Hackathon*, Columbia University, New York City, NY, USA, 2014.

MEMBERSHIPS

Board Networked Digital Library of Theses and Dissertations (NDLTD) Board of Directors	Fall 2022 – present
ACM Membership	*2019 – present*
IEEE Membership	*2018 – 2000*
Schreyer Institute for Teaching Excellence	*2017 – 2018*
Penn State Postdoc Society	*2012 – 2017*

SERVICES

Grant Review Panel

1.	National Science Foundation, GRFP	*2024*
2.	National Science Foundation, REU	*2023*
3.	National Science Foundation, IIS	*2022*
4.	National Science Foundation, IIS	*Twice, 2018*

Program Committee Members:

1. iConference 2024.

2. REUNS 2023. 8th National Workshop for REU Research in Networking and Systems

3. Sci-K 2023. International Workshop on Scientific Knowledge colocated with the Web Conference

4. DIRA 2022. The second workshop on Drawings and abstract Imagery: Representation and Analysis (DIRA) colocated with ECCV 2022

5. AACL-IJCNLP 2022: The 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing

6. EEKE 2021-2023: Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents

7. CLEF 2021: Conference and Labs of the Evaluation Forum

8. SDP 2020-2022: International Workshop on Scholarly Document Processing

9. JCDL 2018–2023: ACM/IEEE Joint Conference on Digital Libraries

10. WOSP 2020: International Workshop on Mining Scientific Publications @JCDL

11. BIRDNL 2019: Joint Workshop on Bibliometrics-enhanced Information Retrieval and Natural Language Processing for Digital Libraries

12. ICIIT 2018: International Conference on Intelligent Information Technology

13. BigScholar 2018: The KDD 2018 Workshop on Big Scholarly Data

14. NAACL 2018: The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

15. ICDAR 2017: The 14th IAPR International Conference on Document Analysis and Recognition

16. DSMM 2017: Data Science for Macro-Modeling with Financial and Economic Datasets

17. ACL 2017: Annual Meeting of the Association for Computational Linguistics

18. NAACL 2016: The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

19. SBD 2015: The 29th AAAI Conference Workshop on Scholarly Big Data

20. NLPCC 2015: The 4th CCF Conference on NLP and Chinese Computing

21. DyNo 2015: The 1st International Workshop on Dynamics in Networks

Conference Committee Chairs, Session Chairs, or Mentors

1. Workshop Co-chair. International Workshop on Artificial Intelligence for Science of Science **Ai4SciSci 2023**

2. Workshop Committee Chair for **JCDL 2019**

3. Tutorial Committee Chair for **JCDL 2020**

4. Doctoral Consortium Mentor for **JCDL 2018**

5. "Semantics and Linking" Session Chair for **JCDL 2018**

6. "Citation Analysis" Session Chair for **JCDL 2017**

Invited Workshops:

1. *Ensuring Scholarly Access to Digital Records Workshops* (Invited by Bill Ingram at Virginia Tech)

2. "Citation Analysis" *The ACM/IEEE-CS Joint Conference on Digital Libraries* (**JCDL 2017**)

Ad Hoc Reviewers:

1. SAC 2022: ACM/SIGAPP Symposium On Applied Computing

2. ACL 2022: Annual Meeting of the Association for Computational Linguistics

3. SDU 2021: International Workshop on Scientific Document Understanding

4. BIRDNL 2017: Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries

5. WWW 2017, 2016, 2015, 2011: International World Wide Web Conference

6. TPDL 2017, 2016, 2013: International Conference on Theory and Practice of Digital Libraries

7. SIGIR 2012, 2011: The 35th Annual International ACM SIGIR Conference

Journal Article Reviewer:

1. TKDD 2023: ACM Transactions on Knowledge Discovery from Data

2. COMSIS 2021: Journal of Computer Science and Information Systems

3. JASIST 2021: Journal of the Association for Information Science and Technology (twice)

4. Nature Machine Intelligence 2020

5. PeerJ 2020

6. Scientometrics 2019

7. Transactions on the Web 2019

8. Computer Science Review 1019

9. JOI 2019: Journal of Informatics

10. IJDL 2018: International Journal on Digital Libraries

11. ACM TKDE 2016: ACM Transactions on Knowledge and Data Engineering

12. JOCS 2015: Journal of Computational Science

13. ACM TKDE 2015: ACM Transactions on Knowledge and Data Engineering

14. SCIENCE CHINA Information Sciences 2015

15. ACM TOIS 2015: ACM Transactions on Information Systems

16. IEEE TMM 2014: IEEE Transactions on Multimedia

17. JBI 2014: Journal of Biomedical Informatics

18. DAMI 2014: Data Mining and Knowledge Discovery

Tutorial Organizer:

1. "Information Extraction for Scholarly Digital Libraries." *A Tutorial in the Joint Conference on Digital Libraries*, (**JCDL 2016**), Newark, NJ, USA.

University/College/Department Level Services:

1. PhD Prelim exam committee for Satvik Chekuri — Virginia Tech, December 8, 2023

2. PhD candidacy exam committee for Bhanuka Mahanama — ODU, Nov 9, 2023

3. Graduate Studies Committee member — ODU, Spring 2023 – present

4. *Reign-in-Science* event — ODU, October 7, 2023

5. Provost-appointed AI Implementation Workgroup — ODU, October–December, 2023

6. PhD candidacy exam committee for Kehinde Ajayi — September 25, 2023

7. Member of the ODU Artificial Intelligence (AI) Implementation Workgroup — ODU, September, 2023

8. AI Symposium panelist — ODU, September 8, 2023

9. GTAI Evaluator Day 2 — ODU, August 25, 2023

10. MS thesis defense committee for Emily Escamilla — ODU, July 25, 2023

11. PhD candidacy exam committee for Mohan K. Sunkara — ODU, July 18, 2023

12. PhD candidacy exam committee for Md Javedul Ferdous — ODU, July 20, 2023

13. ODU Summer Machine Learning Boot Camp — ODU, May 8–12, 2023

14. PhD Prelim exam committee for Bipasha Banerjee — Virginia Tech, May 4, 2023

15. PhD candidacy committee for Kritika Garg — ODU, April 17, 2023

16. MS thesis defense committee member for Manoj Kumar Angadi — PSU, March 23, 2023

17. College of Sciences Science Friday Speaker — March 17, 2023

18. Co-hosting 22 Grassfield High School students to visit the ODU Computer Science Department March 16, 2023

19. Lead the effort of the 2023 ODU Misinformation and Disinformation Workshop and Panel Discussion ODU, Spring 2023

20. Teaching load committee chair: lead a successful proposal to change teaching load from 2+2 to 2+1 ODU, Fall 2021 – Fall, 2022

21. Mentor of CS410/CS411 Blue Team — ODU, Fall 2022

22. PhD thesis defense committee member for Md Ariful Haque — ODU, October 12, 2022

23. ODU New Student Convocation — ODU August 25, 2022

24. MS thesis defense committee member for Javaid Akbar Manzoor — Virginia Tech, September 23, 2022

25. PhD thesis external examiner and committee member for Sherry Maynard — University of West Indies, September, 2022 – February, 2023

26. PhD Skill Requirement Presentation for Yasith Jayawardana — ODU, September 7, 2022

27. PhD thesis defense committee member for Md Reshad Ul Hoque — ODU, July 21, 2022

28. CS faculty for the Spring 2022 ODU Graduation Commencement Ceremony — ODU, May 7, 2022

29. MS thesis defense committee chair for Sami Uddin — ODU, March 4, 2022

30. PhD candidacy committee chair for Muntabir Choudhury — ODU, November 17, 2021

31. PhD proposal defense committee member for Md Reshad Ul Hoque — ODU, November 3, 2021

32. PhD thesis defense committee member for Shawn M. Jones — ODU, August 5, 2021

33. ODU Data Science Summer Camp invited speaker — ODU, August 13, 2021

34. MS defense committee external member for Sai Ajay Modukuri — PSU, June 24, 2021

35. MS defense committee external member for Rajal Nivargi — PSU, June 22, 2021

36. CS faculty for the Spring 2021 ODU Graduation Commencement Ceremony — ODU, May 9, 2021

37. MS defense committee external member for Arjun Menoj Menon — PSU, March 15, 2021

38. MS defense committee external member for Sai Raghav Reddy Keesara — PSU, March 8, 2021

39. Trick-or-Research Open House — ODU, 2019, 2020

40. PhD dissertation committee member for Sawood Alam — ODU, December 4, 2020

41. MS defense committee external member for Sree Sai Teja Lanka — PSU, November 12, 2020

42. PhD candidacy committee chair for Richard Pates — ODU, August 7, 2020

43. PhD dissertation committee member for Mohamed Aturban — ODU, July 23, 2020

44. PhD candidacy committee member for Bathsheba Farrow — ODU, July 9, 2020

45. PhD candidacy exam committee member for Md Reshad Ul Hoque — ODU, November 25, 2019

46. PhD candidacy exam committee member for Mohamed Aturban — ODU, August 22, 2019

47. PhD candidacy exam committee member for Sawood Alam — ODU, August 21, 2019

48. PhD candidacy exam committee member for Shawn M. Jones — ODU, August 7, 2019

49. PhD candidacy exam committee member for Alexander Nwala — ODU, April 29, 2019

50. PhD dissertation committee member for Lulwah M. Alkwai — April 4, 2019

51. Great Computer Challenge Judge — ODU, 2019

52. Scholar's Day 2019 Presenter — ODU, 2019

53. Advisor of a team to participate the Nittany AI Challenge — Penn State, 2018

54. Data Science Faculty Teaching Group — Penn State, 2017-2018

55. Course Committee Chair of DS440: Data Science Capstone — Penn State, 2017-2018

56. Masters Advisory Committee on Data Science — Penn State, 2017-2018

Community Engagement

1. Coach of a team of 3 elementary school students and won the second place of the 2023 Great Computer Challenge (GCC) Video Editing Competition   Virginia Beach Public School, 2023

2. Director of the Computer Club for Artificial Intelligence (CCAI) Club   Old Donation School, Virginia Beach, 2023-2024

3. Scoring Appraiser of the Destination Imagination Team Challenge   Atlee High School, Mechanicsville, March 18, 2023

4. My Research and Teaching at Old Dominion University, a recruitment presentation at Norview High School with Nirmala Karunarathna (Nimi)   Norview High, May 2022

5. Virginia State Science and Engineering Fair (VSSEF) competition judge   Virtual, April 9, 2022

MENTORING

Graduate Students:

1. Saad Khan [ODU MS student on misinformation]

2. Sumit Kumar [PSU MS student on Citation intent classification]

3. Manoj Kumar Angadi [PSU MS student on CiteSeerX]

4. Raj Ratn Pranesh [PSU MS student on information extraction]

5. Md Sami Uddin [ODU MS student on citation parsing]

6. Javaid Manzoor [VT MS student on document segmentation]

7. Sai Ajay Modukuri [PSU MS student on retraction paper classification]

8. Lamia Salsabil [ODU PhD student on replicability project 2021-]

9. Kehinde Ajayi [ODU PhD student on materials science data compilation 2021-]

10. Sai Raghav Reddy Keesara [PSU MS on Extraction-Ingestion-System 2019-2020]

11. Sree Sai Teja Lanka [PSU MS on p-value extraction 2020]

12. Himarsha Jayanetti [ODU PhD student on metadata extraction]

13. Muntabir Choudhury [ODU PhD student on text mining]

14. Xiaohan Wang [UIUC Master's student on entity extraction]

15. Richard Pates [ODU PhD student on ETD Mining]

16. Pei Wang [ODU Master's student on NLP]

17. Xin Wei [ODU PhD student on NLP]

18. Alexander C. Nwala [ODU PhD student on web crawler scheduler]

19. Yasith Jayawardana [ODU PhD student on web crawler scheduler]

20. Gavindya Jayawardena [ODU PhD student on web crawler scheduler]

21. Md Reshad Ul Hoque [ODU PhD student on fake news detection]

22. Bharath Kandimalla [PSU PhD student on neural network methods for entity matching]

23. Jing Zhao [PSU MS student on data cleansing]

24. Shaurya Rohatgi [PSU MS student on CiteSeerX system]

25. Athar Sefid [PSU PhD student on entity-linking between CiteSeerX and IEEE Xplore Database]

26. Kunho Kim [PSU PhD student on entity-linking between CiteSeerX and Web of Science]

27. Agnese Chiatti [PSU PhD student on semantic entity extraction from scientific documents]

28. Kaiyu Wang [PSU MS 2016, principal mentor on semantic entity extraction]

29. Yanbo Sun [PSU MS 2016, principal mentor on compressing video data with ANN]

30. Douglas W. Jordan [PSU MS 2016 on developing RESTful API and NoSQL for CiteSeerX]

31. Wenfang Xu [PSU PhD student]

32. Yuqing Dong [PSU MS 2015 on crawling Google Scholar author profiles]

33. Po-Yu Chuang [PSU MS 2017 on entity-linking between CiteSeerX and Web of Science]

34. Yuan-Hsin Chen [PSU MS 2015 on improving CiteSeerX ranking function]

35. Jayanthi Thankamani [PSU MS 2015 on developing focused search engine]

36. Venkata Neppalli [UNT MS on developing a supervised document type classifier]

37. Alexander G. Ororbia II [PSU PhD student on web crawling]

Undergraduate Students:

1. Dominik Soós [ODU 2021 on ETD Parsing, later on disinformation]

2. Bill Chen [PSU 2021 on CiteSeerX]

3. Li-Zhi Lin [ODU 2021 on data linking]

4. Kevin Kuo [PSU 2021 on developing and developying the NextGen Seer]

5. Jason Chhay [PSU 2021 on developing and deploying the NextGen Seer]

6. Winston Shields [ODU 2021 on patent figure search engine]

7. Ryan Hiltabrand [ODU 2020 on data quality assessment]

8. Kyle Canady [ODU 2020 on information extraction]

9. Martin Gryder [ODU 2019 on image search]

10. Gunnar Reiske [ODU 2019 on domain knowledge entity extraction]

11. Andrew Mayes [ODU 2019 on implementing keyword voting on CiteSeerX]

12. Bradley Dash [ODU 2018 on large scale synthesized citation parsing dataset]

13. Vincent J. Trost [PSU 2018 on entity extraction from scholarly papers]

14. Wei Cheng [PSU 2018 on entity matching between digital libraries]

15. Jianyu Mao [PSU 2018 on topical classification]

16. Jeffery Cao [PSU 2018 on Evaluation of multiple keyphrase extraction models]

17. Seth Wilcox [PSU 2017-2018 on Evaluation of multiple keyphrase extraction models]

18. Allen Ge [PSU 2017 on entity-linking between CiteSeerX and IEEE Xplore Database]

19. Scott O'Connell [2016-2017 on designing Deep Neural Network JPEG decoder]

20. Jason Killian [PSU BS 2015, principal mentor on developing PDFMEF]

21. Huaiyu Yang [Vanderbilt University, funded by NSF REU program on developing PDFMEF]

22. Lichi Li [PSU BS 2013]

High School Students:

1. Andrew Zhang, Figure search		Norman North High School, Norman, OK
2. Alyssa Vargas, Data verification	Governor's School of Science and Technology, Newport News, VA
3. Bryce Kemmey, Deepfake		Massaponax High School, Fredericksburg, VA
4. Huaiyu Yang, Developing PDFMEF		State College Area High School, State College, PA

TEACHING EXPERIENCE

Computer Science Department, Old Dominion University

1. CS418/518: Web Programming — *Fall 2019-2023, Spring 2021*
2. CS450/550: Database Concepts — *Spring 2021-2023*
3. CS734/834: Information Retrieval — *Spring 2019, Fall 2021, Fall 2023*
4. CS795/895: Mining Scholarly Big Data — *Fall 2018, Fall 2020*
5. CS795/895: Deep Learning and Natural Language Understanding — *Spring 2022*
6. CS795/895: Fundamentals of Deep Learning — *Spring 2023*
7. CS800: Research Methods (invited speaker) — *Spring 2020-2022*

Information Sciences and Technology, Pennsylvania State University

1. IST590: Applied Cognitive Science Lab (invited speaker) — *Spring 2019*
2. IST210: Organization of Data (2 sessions) — *Spring 2018*
3. IST140: Introduction to Application Development — *Spring 2018*
4. IST210: Organization of Data (2 sessions) — *Fall 2017*
5. IST140: Introduction to Application Development — *Fall 2017*
6. IST210: Organization of Data — *Spring 2017*
7. IST441: Information Retrieval and Search Engines (Co-teach) — *Spring 2014-2017, 2020-2022; Fall 2021*
8. CSE598B: Performance Evaluation for Clouds (Invited talk) — *Spring, 2014*

Department of Astronomy and Astrophysics, Pennsylvania State University

1. ASTRO11: Elementary Astronomy Lab (Instructor) — *Spring 2006*
2. ASTRO001: Astronomical Universe (TA for Dr. Mercedes Richards) — *Spring 2006*
3. ASTRO001: Astronomical Universe (TA for Dr. Mercedes Richrads) — *Fall 2005*
4. ASTRO001: Astronomical Universe (TA for Dr. Robin Ciardullo) — *Fall 2005*
5. ASTRO001: Astronomical Universe (TA for Dr. Pablo Laguna) — *Spring 2005*
6. ASTRO001: Astronomical Universe (TA for Dr. Kevin Luhman) — *Spring 2005*
7. ASTRO001: Astronomical Universe (TA for Mangala Sharma) — *Spring 2005*
8. ASTRO292: Distance Universe (TA for Dr. Donald Schneider) — *Spring 2005*
9. ASTRO001: Astronomical Universe (TA for Dr. Steinn Sigurdsson) — *Fall 2004*
10. ASTRO291: Solar System (TA for Dr. Donald Schneider) — *Fall 2004*

TECHNICAL SKILLS

- **Ten years** work experience under Linux/Unix environment; Latex and MS Excel.
- **Five years** work experience of building/maintaining production MySQL databases and Apache Solr, debugging and optimizing ETL work flows, based on scholarly big data.
- **Five years** work experience of search engine architecture and infrastructure, deploying and implementing web application features
- **Five years** work experience of designing, coding, and testing LAMP website powered by MySQL databases and Apache Solr, using frameworks such as Django and Spring.
- **Five years** programming experience with Python; familiar with load balancing, virtual environment, firewall (e.g., iptables), and file systems.
- **Three years** work experience of managing software projects on open source software platforms, e.g., GitHub.

- **Two years** experience of analyzing logs using MapReduce; Deep Learning architectures of RNN and CNN on image data; experience with Amazon AWS, Microsoft Azure Cloud, Google Cloud, and Google Analytics; Experience with NLP tools, Bash, Java, R, Ruby on Rails, RESTful API.

- **Backgrounds** in Physics, Math, and Statistics; Familiar with ML, NLP, ANN, IR, and genetic algorithms.