

EDUARD CONSTANTIN DRAGUT

SERC (Room 348), 1925 N. 12th Street
Philadelphia, PA, 19122

(215)-204-0521
edragut@temple.edu
cis.temple.edu/~edragut

RESEARCH INTERESTS

Data management (retrieval, extraction, integration, and quality), Data Mining, Social Media Mining, AI for Social Good, Cyber-infrastructure for science.

EDUCATION

- Ph.D., University of Illinois at Chicago, Computer Science, July 2010.
- M.S., University of Iowa, Computer Science, 2004.
- Bachelor of Science, University of Bucharest, Computer Science, June 1998

EMPLOYMENT

- | | |
|---------------------------------------------|-----------------------|
| ▪ Associate Professor, Temple University | July 2019 – present |
| ▪ Assistant Professor, Temple University | July 2013 – June 2019 |
| ▪ Chair Graduate Studies, Temple University | Aug. 2018 - present |
| ▪ Visiting Professor, Purdue University | Jan. 2013 – May 2013 |
| ▪ Post Doctorate, Purdue University | July 2010 – June 2013 |

FUNDING

- *The Rehabilitation Engineering Research Center on Augmentative and Alternative Communication*. DHHS Administration for Community Living. Aug. 1, 2025 – Sept. 30, 2030. \$1,500,000. Co-PI. With Stephen MacNeil and Slobodan Vucetic.
- *Proto-OKN Theme 1: Knowledge Graph to Support Evaluation and Development of Climate Models*. NSF. Oct. 1, 2023- Sept. 30, 2026. \$1,500,000. Lead PI. With Longin Latecki (Temple University).
- *III: Medium: Collaborative Research: Extracting and Linking AI Artifacts*. NSF. Dec. 1, 2021 – Nov. 30, 2024. \$1,200,000. Lead PI. With Longin Latecki (Temple University) and Cornelia Caragea (University of Illinois at Chicago).
- *NSF Convergence Accelerator Track H: Next Generation Augmentative and Alternative Communication Technology Powered by Artificial Intelligence*. NSF. Dec. 15, 2022 – Dec. 14, 2023. \$750,000. Co-PI. With Slobodan Vucetic and Stephen MacNeil (Temple) and Christine Holyfield and Elizabeth Lorah (University of Arkansas).
- *NSF Convergence Accelerator Track F: America's Fourth Estate at Risk: A System for Mapping the (Local) Journalism Life Cycle to Rebuild the Nation's News Trust*. NSF. Sept. 20, 2021 – Oct. 1, 2022. \$750,000 Lead PI. With Robert Holbert and Aron Pilhofer (Temple); Lei Guo and Gianluca Stringhini (Boston University); and Cornelia Caragea (UIC).

- *Center for Epistemic Security: Preserving a Democratic Society's Epistemic Security*. The OVPR at Temple University. Jan. 1, 2022 – Dec. 31, 2023. \$195,000. Lead PI. With Slobodan Vucetic, Robert Holbert, Bruce Hardy, Meghna Tallapragada, and David Nickerson.
- *R2P3: Reading Between the Lines Across Thousands of Social and News Media Streams to Predict Events*. The OVPR at Temple University. Jan. 1, 2022 – Dec. 31, 2023. \$195,000. Co-PI.
- *Automatic Interpretation of Digital Pathology Images Using Deep Learning*. The OVPR at Temple University. Jan. 1, 2022 – Dec. 31, 2023. \$195,000. Co-PI.
- *FW-HTF-RL: Personalized Virtual Job Assistants to Prepare Individuals with Neurodevelopmental Disabilities for Entry Level IT Jobs*. NSF. Sept. 4, 2020 – Sept. 3, 2024. \$2,310,269. Co-PI with Slobodan Vucetic (PI), Donald Hantula, and Matthew Tincani (all at Temple University).
- *BIGDATA: F: Collaborative Research: Collective Mining of Vertical Social Communities*. NSF. Jan. 1, 2019 – Dec. 31, 2022. \$733,875 (Temple University \$427,912 and Houston University \$305,963). Lead PI. With Arjun Mukherjee.
- *EAGER: Assessing Influence of News Articles on Emerging Events*. NSF. Sept. 1, 2018 – Aug. 31, 2020. \$300,000. Co-PI with PI Zoran Obradovic.
- *BIGDATA: Collaborative Research: F: Streaming Architecture for Continuous Entity Linking in Social Media*. NSF. Jan. 1, 2016 – Dec. 31, 2018. \$1,083,339 (Temple University \$783,339 and SUNY at Binghamton \$300,000). Lead PI, with Co-PI Yuhong Guo (Temple) and PI Weiyi Meng (Binghamton).
 - REU Supplement: \$16,000. Awarded May 4, 2018.

JOURNAL PUBLICATIONS

1. M. West, M. Tincani, D. Hantula, R. Hong, S. Vucetic, and E. Dragut. *Applying data science practices to facilitate autistic students' access to higher education*. Journal of Autism and Developmental Disorders (**JADD**), to appear.
2. C. Holyfield, S. MacNeil, N. Caldwell, T. O'Neill Zimmerman, E. Lorah, E. Dragut and S. Vucetic. *Leveraging Communication Partner Speech to Automate Augmented Input for Children on the Autism Spectrum Who Are Minimally Verbal: Prototype Development and Preliminary Efficacy Investigation*. American Journal of Speech-Language Pathology (**AJSLP**). 2024.
3. E. Garrison, D. Singh, D. Hantula, M. Tincani, J. Nosek, S. Ray Hong, E. Dragut, S. Vucetic. *Understanding the experience of neurodivergent workers in image and text data annotation*. Computers in Human Behavior Reports (**CHB**). Volume 11, 2023.
4. J. Pan, Q. Zhang, E. Dragut, C. Caragea, and L. Latecki. *DMDD: A Large-Scale Dataset for Dataset Mentions Detection*. Transactions of the Association for Computational Linguistics (**TACL**). 11 1132–1146.
5. M. Tincani, H. Ji, M. Upthegrove, E. Garrison, M. West, D. Hantula, Slobodan Vucetic and E. Dragut. *Vocational Interventions for Individuals with ASD: Umbrella Review*. **Rev J Autism Dev Disord** (2023).
6. J. Alshehri, M. Pavlovski, E. Dragut, and Z. Obradovic. *Aligning Comments to News*

- Articles on a Budget*. **IEEE Access** 11: 18900-18909 (2023).
7. S. S. Bhowmick, E. Dragut, and W. Meng. *TwICS: Lightweight Entity Mention Detection in Targeted Twitter Streams*. The IEEE Transactions on Knowledge and Data Engineering (TKDE), 35(1): 1043-1057 (2023).
 8. L. He, C. Han, A. Mukherjee, Z. Obradovic, E. Dragut. *On the dynamics of user engagement in news comment media*. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. (WIREs), 10(1) (2020)
 9. Y. Dong, E. Dragut, W. Meng. *Normalization of Duplicate Records from Multiple Sources*. In the IEEE Transactions on Knowledge and Data Engineering (TKDE), 14 page. 2018.
 10. J. Yuan, L. He, E. Dragut, W. Meng, C. Yu. *Result Merging for Structured Queries on the Deep Web with Active Relevance Weight Estimation*. **Information Systems**. 64: 93-103. 2017.
 11. E. Dragut, H. Wang, C. Yu, S. Prasad, W. Meng. *Polarity Consistency Checking for Domain Independent Sentiment Dictionaries*. The IEEE Transactions on Knowledge and Data Engineering (TKDE), vol.27, no.3, pp.838-851. 2015.
 12. E. Dragut, B. DasGupta, B. Beirne, A. Neyestani, B. Atassi, C. Yu, and W. Meng. *Merging Query Results from Local Search Engines for Geo-referenced Objects*. ACM Transactions on the Web (TWEB), Volume 8, Issue 4, Article No. 20. 29 pages. 2014.
 13. A. Kruger, R. Lawrence, and E. Dragut, *Building a Terabyte NEXRAD Radar Database for Hydrometeorology Research*, **Journal of Computers & Geosciences**, Vol. 32, Issue 2, pages 247-258. 2006.

PEER-REVIEWED CONFERENCE PROCEEDINGS

14. J. Pan, Q. Zhang, E. Dragut, C. Caragea, and L. Latecki . *FlowLearn: Evaluating Large Vision-Language Models on Flowchart Understanding*. European Conference on Artificial Intelligence (ECAI). 2024. [acceptance rate: 23%].
15. Z. Chen, L. He, A. Mukherjee, and E. Dragut. *Comquest: Large Scale User Comment Crawling and Integration*. The ACM Special Interest Group on Management of Data Conference (SIGMOD). 2024. [acceptance rate: 30%].
16. Abdullah Aljebreen, Weiyi Meng, Eduard C. Dragut. *Analysis and Detection of "Pink Slime" Websites in Social Media Posts*. The 33rd International World Wide Web Conference (WWW) 2024: 2572-2581. [acceptance rate: 19%].
17. J. Pan, Q. Zhang, E. Dragut, C. Caragea, and L. Latecki. *SciDMT: A Large-Scale Corpus for Detecting Scientific Mentions*. The Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING). 2024.
18. U. Ergashev, E. Dragut, W. Meng. *Learning To Rank Resources with GNN*. The 32nd International World Wide Web Conference (WWW). 2023: 3247-3256. [acceptance rate: 19%].
19. S. Saha Bhowmick, E. Dragut, W. Meng. *Globally Aware Contextual Embeddings for Named Entity Recognition in Social Media Streams*. The 39th IEEE International Conference on Data Engineering (ICDE). 2023: 1085-1097 [acceptance rate: NA].
20. S. Saha Bhowmick, E. Dragut, W. Meng. *Boosting Entity Mention Detection for*

- Targetted Twitter Streams with Global Contextual Embeddings*. The 38th IEEE International Conference on Data Engineering (**ICDE**). 2022: 1085-1097 [acceptance rate: NA].
21. Z. Chen, W. Meng, E. Dragut. *Web Record Extraction with Invariants*. Proc. **VLDB** Endow. 16(4): 959-972 (2022).
 22. A. Schneider, L. He, Z. Chen, A. Mukherjee, E. Dragut. *COIN - an Inexpensive and Strong Baseline for Predicting Out of Vocabulary Word Embeddings*. International Conference on Computational Linguistics (**COLING**) 2022: 3984-3993 [acceptance rate: 33.4%].
 23. C. Shen, C. Han, L. He, A. Mukherjee, Z. Obradovic, E. Dragut. *Session-based News Recommendation from Temporal User Commenting Dynamics*. The International Conference Series on Advances in Social Network Analysis and Mining (**ASONAM**) 2022: 163-170 [acceptance rate: 19.6%].
 24. J. Alshehri, M. Stanojevic, P. Khan, B. Rapp, E. Dragut, Z. Obradovic. *MultiLayerET: A Unified Representation of Entities and Topics Using Multilayer Graphs*. **ECML/PKDD** (2) 2022: 671-687 [acceptance rate: 23%].
 25. J. Alshehri, M. Stanojevic, E. Dragut, Z. Obradovic. *On Label Quality in Class Imbalance Setting -A Case Study*. **ICMLA** 2022: 1666-1671 [acceptance rate: 25%].
 26. K. Tumarada, Y. Zhang, F. Yang, E. Dragut, O. Gnawali, A. Mukherjee. *Opinion Prediction with User Fingerprinting*. The 6th International Conference Recent Advances in Natural Language Processing (**RANLP**). Sept. 2021. [acceptance rate: 14%].
 27. M. Hosseiniama, E. Dragut, D. Bumber, A. Mukherjee. *On the Usefulness of Personality Traits in Opinion-oriented Tasks*. The 6th International Conference Recent Advances in Natural Language Processing (**RANLP**). Sept. 2021. [acceptance rate: 14%].
 28. F. Yang, E. Dragut, A. Mukherjee. *Improving Evidence Retrieval with Claim-Evidence Entailment*. The 6th International Conference Recent Advances in Natural Language Processing (**RANLP**). Sept. 2021. [acceptance rate: 14%].
 29. A. Aljebreen, W. Meng, and E. Dragut. *Segmentation of Tweets with URLs and its Applications to Sentiment Analysis*. The 35th AAAI Conference on Artificial Intelligence (**AAAI**). Feb. 2021. [acceptance rate: 21%]
 30. L. He, C. Shen, A. Mukherjee, S. Vucetic, E. Dragut. *Cannot Predict Comment Volume of a News Article before (a few) Users Read It*. International AAAI Conference on Web and Social Media (**ICWSM**). Jan. 2021. [acceptance rate: 21.2%]
 31. J. Alshehri, M. Stanojevic, E. Dragut and Z. Obradovic. *Stay on Topic, Please: Aligning User Comments to the Content of a News Article*. The 43rd European Conference on Information Retrieval (**ECIR**). March 2021. [acceptance rate: 24.1%]
 32. F. Yang, E. Dragut, and A. Mukherjee. *Predicting Personal Opinion on Future Events with Fingerprints*. International Conference on Computational Linguistics, (**COLING**). Dec. 2020. [acceptance rate: 33.4%]
 33. F. Yang, E. Dragut, and A. Mukherjee. *Claim Verification under Positive Unlabeled Learning*. The International Conference Series on Advances in Social Network Analysis and Mining (**ASONAM**). Dec. 2020. [acceptance rate: 17.8]
 34. Y. Zhang, F. Yang, Y. Zhang, E. Dragut and A. Mukherjee. *Birds of a Feather Flock*

- Together: Satirical News Detection via Language Model Differentiation*. International Conference on Social Computing, Behavioral-Cultural Modeling (**SBP**). Oct. 2020. [acceptance rate: NA]
35. S. Zhang, L. He, E. Dragut, S. Vucetic. *How to Invest my Time: Lessons from Human-in-the-Loop Entity Extraction*. The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (**KDD**). 2305-2313. Anchorage, AK, USA, August 2019. [acceptance rate: 20.7%]
 36. M. Hosseinia, E. Dragut, A. Mukherjee. *Pro/Con: Neural Detection of Stance in Argumentative Opinions*. The 12th International Conference Social, Cultural, and Behavioral Modeling (**SBP-BRiMS**). 2019: 21-30. [acceptance rate: 21%]
 37. S. Zhang, L. He, S. Vucetic, E. Dragut. *Regular Expression Guided Entity Mention Mining from Noisy Web Data*. The Conference on Empirical Methods in Natural Language (**EMNLP**): 1991-2000. Brussels, Belgium, Oct. 2018. [acceptance rate: 25.5%]
 38. A. Schneider, J. Male, S. Bhogadhi and E. Dragut. *DebugSL: An Interactive Tool for Debugging Sentiment Lexicons*. The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL- HLT**). New Orleans. June 2018. [acceptance rate: 29%].
 39. A. Schneider, A. Mukherjee, E. Dragut. *Leveraging Social Media Signals for Record Linkage*. The 27th edition of the World Wide Web Conference (**WWW**). Lyon, France, April 2018. [acceptance rate: 14.8%].
 40. F. Yang, A. Mukherjee, E. Dragut. *Satirical News Detection and Analysis using Attention Mechanism and Linguistic Features*. The Conference on Empirical Methods in Natural Language (**EMNLP**), Copenhagen, Denmark, Sept. 2017 [acceptance rate: 25%].
 41. A. Wang, J. Croft, E. Dragut. *Reflections on Data Integration for SDN*. ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization (**SDN-NFV Security**). Scottsdale, US, March 2017. [acceptance rate: unknown]
 42. J. Jupin, J. Shi, E. Dragut. *PSH: A Probabilistic Signature Hash Method with Hash Neighborhood Candidate Generation for Fast Edit-Distance String Comparison on Big Data*. The **IEEE Big Data**. Washington. DC, USA, Dec. 2016. [acceptance rate: 18.7%].
 43. El Kindi Rezig, E. Dragut, M. Ouzzani, A. Elmagarmid and W. Aref. *ORLF: A Flexible Framework for Online Record Linkage and Fusion*. The 32nd International Conference on Data Engineering (**ICDE**). [acceptance rate: 33%].
 44. Q. Liu, E. Dragut., A. Mukherjee, W. Meng. *FLORIN – A System to Support (Near) Real- Time Applications on User Generated Content on Daily News*. The 41st International Conference on Very Large Data Bases (**VLDB**), Hawaii, Volume 8 Issue 12, August 2015. 1944-1955. [acceptance rate: 33%].
 45. A. Schneider, E. Dragut. *Towards Debugging Sentiment Lexicons*. The 53th Annual Meeting of the Association for Computational Linguistics (**ACL**), Beijing, China, July 2015. 1024- 1034. [acceptance rate: 25%]

46. E. Rezig, E. Dragut, M. Ouzzani, and A. Elmagarmid. *Query-Time Record Linkage and Fusion over Web Databases*. The 31st International Conference on Data Engineering (**ICDE**), Seoul, Korea, April 2015. 42 – 53. [acceptance rate: 20%].
47. D. Salt, M. Ouzzani, E. Dragut, P. Baker, S. Rangarajan. *iHUB - An Information and Collaborative Management Platform for Life Sciences*. The 23rd International World Wide Web Conference (**WWW**). Seoul, Korea. April 2014. [acceptance rate: 30%]
48. E. Dragut, P. Baker, J. Xu, M. Sarfraz, E. Bertino, A. Madkour, R. Agarwal, A. Mahmood and S. Han. *CRISS - Computational Research Infrastructure for Science*. The 14th IEEE International Conference on Information Reuse and Integration, (**IRI**), San Francisco, USA. August 2013 (invited paper).
49. L. Cen, E. Dragut and L. Si. *Author Disambiguation by Hierarchical Agglomerative Clustering with Adaptive Stopping Criterion*. The 36th international ACM SIGIR conference on Research and development in information retrieval (**SIGIR**). Dublin, Ireland. July 2013, 741-744. [acceptance rate: 35%].
50. E. Dragut, B. P. Beirne, A. Neyestani, B. Atassi, C. Yu, W. Meng, B. DasGupta, *YumiInt - A Deep Web Integrating System for Local Search Engines for Geo-referenced Objects*. The 29th IEEE International Conference on Data Engineering (**ICDE**), Brisbane, Australia. April 2013 (demo paper) [acceptance rate: around 30%].
51. E. Dragut, W. Hong, C. Yu, P. Sistla, W. Meng, *Polarity Consistency Checking for Sentiment Dictionaries*. The 50th Annual Meeting of the Association for Computational Linguistics (**ACL**), Jeju, South Korea, July 2012 [acceptance rate: 19%].
52. E. Dragut, M. Ouzzani, A. Madkour, N. Mohamed, P. Baker, D. Salt. *Ionomics Atlas - A Tool To Explore Interconnected Ionomic, Genomic and Environmental Data*. The 21st ACM International Conference on Information and Knowledge Management (**CIKM**), Maui, USA, November 2012 (demo paper) [acceptance rate: 40%]
53. E. Dragut, C. Yu, P. Sistla, W. Meng. *Construction of a Sentiment Word Dictionary*. The 19th International Conference on Information and Knowledge Management (**CIKM**), Toronto, Canada, October 2010. [acceptance rate: 31.3%]
54. T. Kabisch, E. Dragut, C. Yu, U. Leser. *Deep Web Integration with VisQI*. The 36th International Conference on Very Large Data Bases (**VLDB**), Singapore, Sept 2010 (demo paper) [acceptance rate: 36%].
55. E. Dragut, F. Fang, P. Sistla, C. Yu, W. Meng. *Stop Word and Related Problems in Web Interface Integration*. The of 35th International Conference on Very Large Data Bases (**VLDB**), Pages 349-360, Lyon, France, August 2009 [acceptance rate: 17.9%]
56. E. Dragut, T. Kabisch, C. Yu, U. Leser. *A Hierarchical Approach to Model Web Query Interfaces for Web Source Integration*. The 35th International Conference on Very Large Data Bases (**VLDB**), Pages 325-336, Lyon, France, August 2009 [acceptance rate: 17.9%]
57. E. Dragut, F. Fang, C. Yu, W. Meng, *Deriving Customized Integrated Web Query Interfaces*. The IEEE/WIC/ACM International Conference on Web Intelligence (**WI**). Milan, Italy, September 2009 [acceptance rate: 34%]
58. E. Dragut, C. Yu, W. Meng. *Meaningful Labeling of Integrated Query Interfaces*. The 32nd International Conference on Very Large Data Bases (**VLDB**), Pages 679-690, Seoul, Korea, September 2006 [acceptance rate: 13.2%]

59. E. Dragut, W. Wu, C. Yu, P. Sistla, W. Meng. *Merging Source Query Interfaces on Web Databases*. The 22nd International Conference on Data Engineering (**ICDE**), Pages 679-690, Atlanta, USA, April 2006 [acceptance rate: 19%]
60. E. Dragut and R. Lawrence. *Reducing the Cost of Validating Mapping Compositions by Exploiting Semantic Relationships*. The International Conference on Ontologies, Databases and Applications of Semantics (**ODBASE**). 882-890, Montpellier, France, Oct. 2006 [acceptance rate: 34.1%]
61. E. Dragut and R. Lawrence. *Composing Mappings between Schemas using a Reference Ontology*. The International Conference on Ontologies, Databases and Applications of Semantics (**ODBASE**), Pages 783-800, Larnaca, Cyprus, 2004 [acceptance rate: 25%]

BOOKS

62. E. Dragut, W. Meng and C. Yu. *Deep Web Query Interface Understanding and Integration*. Synthesis Lectures on Data Management, Morgan & Claypool Publishers, 2012.

WORKSHOPS

63. M. Adamu, Q. Zhang, H. Pan, E. Dragut, L. J. Latecki. Querying Climate Knowledge: Semantic Retrieval for Scientific Discovery A Knowledge Graph Approach for Climate Science. **MANILA@SIGIR**. July 2025.
64. H. Pan, M. Adamu, Q. Zhang, E. Dragut, L. J. Latecki. *ClimateIE: A Dataset for Climate Science Information Extraction*. **ClimateNLP@ACL**. August 2025.
65. M. Hosseinia, E. Dragut and A. Mukherjee. *Stance Prediction for Contemporary Issues: Data and Experiments*. **SocialNLP@WWW**. July 9-10, 2020.
66. M. Stanojevic, J. Alshehri, E. Dragut, Z. Obradovic. *Biased News Data Influence on Classifying Social Media Posts*. **NewsIR@SIGIR** 2019: July 2019.
67. E. Dragut and C. Fellbaum. *The Role of Adverbs in Sentiment Analysis*. Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929-2014) with ACL. Baltimore, MD, USA. June 2014
68. A. R. Mahmood, W. G. Aref, E. C. Dragut, S. Basalamah. *The Palm-tree Index: Indexing with the crowd*. The 1st VLDB Workshop on Databases and Crowdsourcing (**DBCrowd**). Trento, Italy. August 2013.

ARCHIVE

69. L. Wu, X. Du, J. Wu, J. Liu, E. Dragut. *An Accountable Anonymous Data Aggregation Scheme for Internet of Things*. CoRR abs/1803.07760 (2018).

UNDERGRADUATE STUDENT PUBLICATIONS

70. Kunal Waghray. *JSON Schema Matching: Empirical Observations*. In the ACM SIGMOD Student Research Competition. Portland, OR, USA. 2020.
71. Aidan Shea. *A Visual System for Mining Crime Mining across College Campuses*. In the ACM SIGMOD Student Research Competition. Chicago, IL, USA. 2017.

72. S. Wu, J. Male, E. Dragut. *Spatial-Temporal Campus Crime Pattern Mining from Historical Alert Messages*. The 3rd National Workshop for REU Research in Networking and Systems (**REUNS**). Silicon Valley USA. January, 2017.

OTHER PUBLICATIONS

73. X. L. Dong and E. Dragut. *The 10th International Workshop on Quality in Databases – QDB 2012* (workshop report). **ACM SIGMOD Record**. Volume 41 Issue 4. pp. 55-59. Dec. 2012.

PRESENTATIONS

- “Continuous, Gradual Entity Mining from Web Data Streams”, Invited Talk at Texas University at Arlington. Arlington, TX, USA. April 2023.
- “Continuous, Gradual Entity Mining from Web Data Streams”, Invited Talk at IBM Research. Almaden, CA, USA. April 2023.
- “Continuous, Gradual Entity Mining from Web Data Streams”, Invited Talk at University of California Merced. Merced, CA, USA. March 2023.
- “Continuous, Gradual Entity Mining from Web Data Streams”, Invited Talk at Megagon Labs. San Jose, CA, USA. March 2023.
- “Human-in-the-Loop Regular Expression Guided Entity Mention Mining from Noisy Web Data”, Invited Talk at IBM Research. Almaden, CA, USA. August 2019.
- “How to Invest my Time: Lessons from Human-in-the-Loop Entity Extraction”, presentation @ KDD. Anchorage, AK, USA, August 2019.
- “Leveraging Social Media Signals for Record Linkage”, Invited Talk in the Computer Science Department Invited Speaker Series at Binghamton University. March 30, 2018.
- “Entity Matching via Word Embedding,” Invited Talk at NSF BDSpoke Initiative Workshop at Purdue, West Lafayette, March 29, 2017.
- “Towards a Formal Understanding of Sentiment Annotation in Sentiment Lexicons,” Invited Talk: CLUNCH (Computational Linguistics and Lunch) - University of Pennsylvania. Philadelphia, USA. November 11, 2015.
- “FLORIN – A System to Support (Near) Real-Time Applications on User Generated Content on Daily News.” demo presentation, at the 41st International Conference on Very Large Data Bases (VLDB'15), September, 2015.
- “Merging Query Results from Local Search Engines for Geo-referenced Objects”. Invited Talk. Computer & Information Science - University of Pennsylvania. Philadelphia, USA. August 6, 2014.
- “Merging Query Results from Local Search Engines for Geo-referenced Objects”. Invited Talk. Orbitz. Chicago, USA. April 2, 2014.
- “Ionomics Atlas - A Tool to Explore Interconnected Ionomic, Genomic and Environmental Data”, demo presentation, at the 21st ACM International Conference on Information and Knowledge Management (CIKM '12), Maui, USA, November 2012.
- “Polarity Consistency Checking for Sentiment Dictionaries”, poster presentation, at the 50th Annual Meeting of the Association for Computational Linguistics (ACL'12), Jeju, South Korea, July 2012
- “Construction of a Sentimental Word Dictionary”, poster presentation, at the 19th International Conference on Information and Knowledge Management (CIKM'10),

Toronto, Canada, October 2010 “A Framework for Transparently Accessing Deep Web Sources”, Purdue University, Cyber Center Seminars, March 11th, 2010

- “Stop Word and Related Problems in Web Interface Integration”, paper presentation at the 35th International Conference on Very Large Data Bases (VLDB’09). Lyon, France, Aug. 2009
- “Meaningful Labeling of Integrated Query Interfaces”, paper presentation at the 32nd International Conference on Very Large Data Bases (VLDB’06). Seoul, Korea, Sept. 2006
- “Merging Source Query Interfaces on Web Databases”, paper presentation at the 22nd International Conference on Data Engineering (ICDE’06). Atlanta, USA, April 2006
- “Metasearch Engines for e-commerce”, poster presentation at the Industrial Advisory Board Meeting, Chicago, October 2005
- “Merging Source Query Interfaces on the Deep Web”, poster presentation, Second Midwest Database Research Symposium, Chicago, April 16, 2005.

STUDENTS SUPERVISED

- 9 PhD students, 18 MS, 28 Undergraduate students

STUDENT AWARDS AND RECOGNITION

Aidan Shea	CIS: Scott Hibbs Future of Computing Award	2018
Aidan Shea	2 nd runner up at the Student Research Competition, ACM SIGMOD	2017
John Male	Honorable Mention at the Temple URP Symposium Poster Presentation	2016

TEACHNIG METHODS RESEARCH

- **Using Algorithm Animation Technology to Demystify Algorithms.** Teaching with Technology Symposium, Temple University. March 3, 2017.

COURSES TAUGHT

Temple University

- CIS 5590 - Topics in Computer Science: Information Retrieval – Spring 2024
- CIS 5516 - Principles of Data Management – Fall 2023
- CIS 5516 - Principles of Data Management – Fall 2022
- CIS5516 - Principles of Database Systems – Fall 2021.
- CIS 4517/5517 - Data-Intensive and Cloud Computing – Spring 2021.
- CIS 5516 - Principles of Database Systems – Fall 2020.
- CIS 5517 - Data-Intensive and Cloud Computing – Spring 2020.
- CIS 4331 - Principles of Database Systems – Fall 2019.
- CIS5516 – Principles of Data Management - Spring 2019.
- CIS5517 - Data-Intensive and Cloud Computing – Fall 2018.
- CIS5516 – Principles of Data Management - Spring 2018.
- CIS2168 - Data Structures- Fall 2017.
- CIS 5516 - Principles of Data Management - Spring 2017.
- CIS 5590 - Introduction to Information Retrieval - Fall 2016.
- CIS 5516 Principles of Data Management - Spring 2016.
- CIS 4331 Principles of Database Systems – Fall 2015.

- CIS 4362 Application System Development Using Relational Technology – Spr. 2015.
- CIS 5516 - Principles of Data Management – Fall 2014
- CIS2168 - Data Structures and Algorithms - Fall 2013

Purdue University

- CS448 - Introduction to Relational Database Systems - Spring 2013.

COURSE DEVELOPMENT

- **CIS 5590 - Topics in Computer Science: Introduction to Information Retrieval.** The course covers models for information retrieval, techniques for indexing and searching, query expansion, latent semantic indexing, and link analysis and ranking.
- **CIS 4517/5517 - Data-Intensive and Cloud Computing.** The course covers a spectrum of topics from core techniques in data management and analytics to highly-scalable data processing using big data and cloud computing systems. It introduces big data ecosystems such as Hadoop, Spark, Storm and MapReduce; cloud technologies such as Amazon EC2 and Google Cloud; data management tailored to cloud and big data such as No SQL, Google Big Table/Apache HBase, and introductory applications to Big Data and cloud environment.

CURRICULUM DEVELOPMENT

- **Master of Science in Data Science.** (Part of the committee.) It provides students with a strong foundation in algorithmic, computational, and statistical thinking, as well as a firm understanding of computer systems.

PROFESSIONAL ACTIVITIES

Associate Editor

- TKDE – since 2022

Program Committee (by year)

- 2024
 - Co-chair: DaSH-NLP Workshop @NAACL
 - SIGMOD, KDD, ACL, ECAI, NAACL, COLING
- 2023
 - AAAI, ACL, NAACL, EMNLP, WWW, SIGMOD, CIKM, ASONAM
- 2022
 - Local Arrangement Chair at ACM SIGMOD.
 - Chair of Demo Track at VLDB.
 - Co-chair: DaSH-NLP Workshop @EMNLP
 - AAAI, WWW, SDM
- 2021
 - WWW, AAAI, DASFA, NAACL, ACL, EMNLP, EACL, SSDBM, ASONAM.
 - Co-chair: DaSH-NLP Workshop @ NAACL and DASH@KDD.
- 2020
 - SIGMOD, VLDB, WWW, ACL, EMNLP, CIKM, AAAI.
 - Co-chair: DaSH Workshop @ KDD.
- 2019

- SIGMOD, WWW, DASFA.
- 2018
 - WWW, EMNLP, NAACL-HLT, SIGMOD,
- 2017
 - SIGMOD Student Research Competition, SIGMOD, ACL, EMNLP, SocialNLP@EACL, IEEE BigData.
- 2016
 - SIGMOD, WWW, ACL, EMNLP.
 - Chair SIGMOD/PODS PhD Symposium.
- 2015
 - SIGMOD, WWW, DASFAA, NAACL-HLT, EMNLP
- 2014
 - Chair ICDE PhD Symposium, Chicago, USA.
 - DASFAA, AICCSA, COLING, ADBIS
- 2013
 - ACL, CIKM (demo track), WWW (Data Mining track); DASFAA; AICCSA.
- 2012
 - ICDE; CIKM (session chair);
 - Co-chair (with Dr. Xin Luna Dong, AT&T) International Workshop on Quality in Databases (QDB) in conjunction with VLDB 2012 (report in SIGMOD Record);

Journal Reviewer:

Very Large Data Bases Journal (VLDBJ); Transactions on Knowledge and Data Engineering (TKDE); Transactions on the Web (TWEB); Distributed and Parallel Databases; Transactions on Services Computing (TSC); Information Systems Journal (ISJ); Transactions on Knowledge Discovery from Data (TKDD); International Journal of Information Security (IJIS); Word Wide Web Journal (WWWJ), ACM Transactions on Interactive Intelligent Systems (TiiS); Wiley/WIREs Data Mining and Knowledge Discovery; Computers & Geosciences.

Panelist:

- 2024: NSF SaTC
- 2023: NSF ISS, SaTC.
- 2022: NSF IIS
- 2021: NSF IIS
- 2020: NSF IIS and AI Research Institutes
- 2019: NSF IIS
- 2018: NSF SBIR/STTR, External Reviewer IIS, External Reviewer DoE
- 2017: NSF IIS, SBIR/STTR
- 2016: NSF IIS, NSF SBIR/STTR
- 2015: NSF IIS, DoE

NON-ACADEMIC EXPERIENCE

TRONN SOFTWARE, July 2000 – November 2001

Project manager

- Managed the development of internet e-commerce web sites; supervised 6 people.
- Designed and built an online car-commerce application for PORSCHE ROMANIA.

IPACRI ROMANIA SRL, September 1998 – May 2000

- **Analyst-programmer consultant** at ELSAG BANKLAB SPA in Rome, Italy.
Participated in two large banking projects for Cassa di Risparmio di Perugia; involved in analysis, development, deployment and maintenance.

SINORG ROUMANIE SA, May 1998 – September 1998

- **Programmer** - Developed applications for local public administrative areas in French provinces.