

Zhiyu Chen

CONTACT INFORMATION

Applied Scientist, Amazon
440 Terry Ave N
Seattle, WA, US, 98109

E-mail: zhiyuchen.ai@gmail.com
Personal Webpage: www.zhiyuchen.com

RESEARCH INTERESTS

Information retrieval, natural language processing, data mining, machine learning

EDUCATION

Lehigh University, Bethlehem, PA, USA **09/2015-08/2022**

Ph.D., Department of Computer Science & Engineering

- Dissertation Title: Dataset Search and Augmentation
- Advisor: Brian D. Davison
- Overall GPA: 3.97/4.0

Ecole Supérieure d'Ingénieurs Léonard de Vinci, Paris, France **09/2014-01/2015**

Exchange Program

- Advisor: Gael Chareyron
- Overall GPA: 15.75/20

Nanjing University of Aeronautics and Astronautics, Nanjing, China **09/2011-06/2015**

B.S., Department of Computer Science & Technology

- Advisor: Dechang Pi
- Overall GPA: 4.2/5.0 (ranking: 1/94)

PUBLICATIONS

Model-based Unbiased Learning to Rank **2023**

D. Luo, L. Zou, Q. Ai, Z. Chen, D. Yin, B.D. Davison

Accepted by Proceedings of The 16th Annual ACM International Conference on Web Search and Data Mining, (WSDM 2023).

Reinforced Question Rewriting for Conversational Question Answering **2022**

Z. Chen, J. Zhao, A. Fang, B. Fetahu, O. Rokhlenko and S. Malmasi

In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022 Industry Track).

StruBERT: Structure-aware BERT for Table Search and Matching **2022**

M. Trabelsi, Z. Chen, S. Zhang, J. Heftin, and B.D. Davison

In Proceedings of the 31st Web Conference, Online, April, 2022 (WWW 2022).

MGNETS: Multi-Graph Neural Networks for Table Search **2021**

Z. Chen, M. Trabelsi, J. Heftin, D. Yin and B.D. Davison

In Proceedings of the 30th ACM International Conference on Information and Knowledge Management, Online, November, 2021 (CIKM 2021).

Neural Ranking Models for Document Retrieval **2021**

M. Trabelsi, Z. Chen, J. Heftin and B.D. Davison

Information Retrieval Journal, October, 2021

- WTR: A Test Collection for Web Table Retrieval** **2021**
Z. Chen, S. Zhang and B.D. Davison
 In Proceedings of the 44rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Online, July, 2021 (SIGIR 2021).
- A Hybrid Deep Model for Learning to Rank Data Tables** **2020**
M. Trabelsi, Z. Chen, J. Hefflin and B.D. Davison
 In Proceedings of the 2020 IEEE International Conference on Big Data (BigData 2020), December, 2020
- Relational Graph Embeddings for Table Retrieval** **2020**
M. Trabelsi, Z. Chen*, J. Hefflin and B.D. Davison*
 In the Seventh International Workshop on High Performance Big Graph Data Management, Analysis, and Mining (BigGraphs 2020), held with IEEE BigData 2020
- Towards Knowledge Acquisition of Metadata on AI Progress** **2020**
Z. Chen, M. Trabelsi*, J. Hefflin and B.D. Davison*
 In Proceedings of the ISWC 2020 Demos and Industry Tracks: From Novel Ideas to Industrial Practice, co-located with the 19th International Semantic Web Conference (ISWC 2020), November, 2020
- Table Search Using a Deep Contextualized Language Model** **2020**
Z. Chen, M. Trabelsi, J. Hefflin, Y. Xu and B.D. Davison
 In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020), Xi'An, China, July, 2020.
- Pretrained Generalized Autoregressive Model with Adaptive Probabilistic Label Cluster for Extreme Multi-label Text Classification** **2020**
H. Ye, Z. Chen, D. Wang and B. D. Davison
 In Proceedings of the 37th International Conference on Machine Learning (ICML 2020), July, 2020.
- Leveraging Schema Labels to Enhance Dataset Search** **2020**
Z. Chen, H. Jia, J. Hefflin, and B.D. Davison
 In Proceedings of the 42nd European Conference on Information Retrieval (ECIR 2020), Lisbon, Portugal, April, 2020.
- Recognizing Quantity Names for Tabular Data** **2018**
Y. Yi, Z. Chen, J. Hefflin and B.D. Davison
 In Joint Proceedings of the First International Workshop on Professional Search (ProfS2018); the Second Workshop on Knowledge Graphs and Semantics for Text Retrieval, Analysis, and Understanding (KG4IR); and the International Workshop on Data Search (DATA:SEARCH'18), pages 68-73. Co-located with SIGIR 2018, Ann Arbor, Michigan, USA, July.
- Generating Schema Labels through Dataset Content Analysis (Best paper award)** **2018**
Z. Chen, H. Jia, J. Hefflin, and B.D. Davison
 In Companion Proceedings of The Web Conference (WWW '18), Presented at the International Workshop on Profiling and Searching Data on the Web (Profiles & Data:Search'18, Co-located with The Web Conference, Lyon, France, April.

| | |
|------------------------|--|
| INDUSTRY EXPERIENCE | <p>Amazon, Seattle, WA, US <i>Applied Scientist Intern</i> 05/2021-09/2021 A research project about conversational question answering. The paper was published in EMNLP 2022 industry track.</p> |
| | <p>Zhuiyi Technology, Shenzhen, China <i>Machine Learning Intern</i> 05/2019-09/2019 Represented the company to join a natural language understanding competition: https://super.gluebenchmark.com/leaderboard. I was in charge of WiC and WSC among all 8 tasks on Super-GLUE. I also contributed ideas to other 2 team members who were working on the rest tasks. Our submission ranked 3rd at the time of submission, including the human baseline ranked 1st.</p> |
| | <p>Bloomberg, Princeton, New Jersey, USA <i>Data Science Intern</i> 06/2016-08/2016 Developed a tool to analyze and visualize the log of Bloomberg Data License.</p> |
| RESEARCH PROJECTS | <p>Domain-Agnostic Dataset Search Lehigh University, Bethlehem, PA, USA <i>Research Assistant</i> 06/2016-05/2022 The goal of this NSF-funded project is to develop a prototype dataset search engine incorporating new techniques for full-content indexing to enable searchers to find data across the web, regardless of domain. My contribution includes: crawled datasets from data portals; developed methods to augment dataset metadata (schema labels); developed new models to rank datasets; created a new benchmark for evaluation.</p> |
| | <p>The PRAISys Project, Lehigh University, Bethlehem, PA, USA <i>Research Assistant</i> 05/2018-08/2018 The PRAISys (Probabilistic Resilience Assessment of Interdependent Systems) platform performs post-event resilience analysis of communities by addressing stochastic interdependencies among infrastructure systems in a probabilistic way. I improved the platform by augmenting the simulator with sub-component analysis.</p> |
| | <p>Discovery of Adverse Drug Reactions from Online Forums, Lehigh University, Bethlehem, PA, USA <i>Research Assistant</i> 09/2015-06/2016 Proposed an algorithm that can automatically extract intra-sentence Drug-ADRs and inter-sentence Drug-ADRs from online medical forums.</p> |
| TALKS | <p>Deep Text Matching and Applications (slides) 2019 Z. Chen Presentation of Depth Study & Graduate Research Seminar Series (GRSS) at the Lehigh University, Bethlehem, USA, May 2019.</p> |
| | <p>Challenges and Progress in Dataset Search (slides) 2018 Z. Chen Presentation at the Eighth BCS-IRSG Symposium on Future Directions in Information Access (FDIA 2018), co-located with the 8th International Conference on the Theory of Information Retrieval, Tianjin, China, September 2018.</p> |

| | | |
|----------------------------|--|---------------------|
| PROFESSIONAL ACTIVITIES | Co-organizer | SemEval 2023 Task 2 |
| | PC Member | ECML PKDD 2023 |
| | PC Member | IJCAI 2023 |
| | PC Member | KDD 2023 |
| | PC Member | SIGIR 2023 |
| | PC Member | ACL 2023 |
| | PC Member | TheWebConf 2023 |
| | PC Member | WSDM 2023 |
| | Reviewer | COLING 2022 |
| | PC Member | SIGIR 2022 |
| | PC Member | TheWebConf 2022 |
| | PC Member | KDD 2022 |
| | PC Member | WSDM 2022 |
| | Reviewer | TKDE 2021 |
| | External Reviewer | SIGIR 2021 |
| | Reviewer | CIKM 2020 |
| | External Reviewer | SIGIR 2020 |
| | External Reviewer | SDM 2019 |
| | Student Volunteer | SIGIR 2018 |
| | External Reviewer | BigData 2017-2021 |
| External Reviewer | WWW 2017 | |
| TEACHING | Teaching Assistant Fundamentals of Programming (CSE 2), Lehigh University, 2016 Fall and Spring 2017 | |
| HONORS AND AWARDS | 2021, ACM SIGIR Student Travel Grant | |
| | 2021, Rossin Professional Development Program Award at Lehigh University | |
| | 2020, ISWC Student Grant | |
| | 2020, ACM SIGIR Student Travel Grant | |
| | 2018, Best Paper Award of International Workshop on Profiles & Data:Search'18 | |
| | 2018, ACM SIGIR Student Travel Grant | |
| | 2015, RCEAS Dean's Fellowship | |
| | 2014, Chinese Government Scholarship from China Scholarship Council | |
| | 2013, First Prize, National Mathematical Contest in Modeling (Jiangsu Division) | |
| | 2012, Honorable Mention for Social Practice | |
| | 2012, Second Prize, 11 th Higher Mathematics Competition of Jiangsu Province | |
| | 2011-2014, Student Scholarship | |
| 2011-2014, Merit Student | | |