Shubham Chatterjee

✉ shubham.chatterjee@mst.edu

 https://github.com/shubham526 • in www.linkedin.com/in/shubham-chatterjee

EDUCATION

University of New Hampshire, Durham, USA
Doctor of Philosophy (PhD) in Computer Science

Completed: September 2022
GPA: 3.83/4

University of New Hampshire, Durham, USA
Master of Science(MS) in Computer Science

Completed: December 2020
GPA: 3.83/4

University of Calcutta, Kolkata, India
Master of Science (MSc) in Computer Science

Completed: May 2017
GPA: 8.54/10

University of Calcutta, Kolkata, India
Bachelor of Science (BSc) in Computer Science

Completed: May 2015
GPA: 8.43/10

RESEARCH INTERESTS

○ **Neural Entity-Oriented Information Retrieval**, emphasizing the integration of Knowledge Graph semantics in neural IR models. I focus on fine-grained information extraction, query-specific representation learning, and entity ranking to develop intelligent web search systems.
○ **Conversational Information Seeking**, focusing on the development of adaptive conversational agents that offer tailored user experiences by dynamically adjusting to user profiles and histories. I am also interested in mixed initiative in conversational systems to strike a balance, ensuring that the agent's guidance enhances discourse without being perceived as overly intrusive.
○ **Large Language Models** to improve the accuracy of search results, ensuring the generation of trustworthy and contextually relevant content in response to diverse user queries.

RESEARCH EXPERIENCE

Postdoctoral Research Associate, GRILL Lab. Mentor: Dr. Jeff Dalton
University of Edinburgh, Scotland
University of Glasgow, Scotland

October 2023 – August 2024
December 2022 – September 2023

○ Developed a novel neural IR model that integrates Knowledge Graph semantics to learn query-specific document representations. Showed that the approach can outperform several hard neural and non-neural baselines on three large-scale document ranking datasets by 40–60%. Paper accepted at ECIR 2024.
○ In collaboration with a PhD student, I led the development of the Generative Relevance Feedback (GRF) method which leverages Large Language Models to derive query expansion terms from generated text. We demonstrate that the GRF method improves the precision of document retrieval by 5–19% tested across three large-scale document ranking test collections. Notably, our GRF method set new benchmarks in recall across all datasets examined. A paper from this research was published at SIGIR 2023, and another at a CIKM 2023 workshop on LLMs.
○ Collaborated with a PhD student on the development of a method to fine-tune the use of LLM-generated texts in query expansion, addressing the LLMs' "hallucination" tendencies. Our approach emphasized aligning generated documents with relevant ones in target collections, achieving a performance boost of 6–9% in precision and 2–4% in recall across three document ranking test collections. Paper accepted at CIKM 2023.

- Lead a collaborative project with Radboud University, The Netherlands, focused on personalized conversational search. This project centers around addressing the challenges in current dialogue systems that often provide generic responses, overlooking individual user nuances. Recognizing the need for systems to remember and utilize user preferences across multiple sessions, our collaboration is developing an LLM-augmented dialogue construction framework. This framework aims to amass large-scale, multi-session, and diverse personalized dialogue datasets by leveraging LLMs in tandem with crowd workers. Paper accepted at SIGIR 2024.
- Lead a collaborative project with the University of Amsterdam on integrating Knowledge Graph semantics in learned-sparse retrieval models.

Postdoctoral Research Associate, TREMA Lab. Mentor: Dr. Laura Dietz
University of New Hampshire, Durham, USA   *September 2022 – November 2022*

- In collaboration with a PhD student, developed the neural model which integrates BERT-based text embeddings with Wikipedia-based entity embeddings for entity ranking. Showed that the approach improves entity ranking performance by 54% on a large-scale entity ranking test collection. Paper accepted at IJCKG 2023.
- Proposed a new method of learning query-specific latent entity spaces that are useful for IR. Showed that the approach can outperform several hard neural and non-neural baselines on two large-scale entity ranking datasets by 40–60%. Paper under review.

Graduate Research Assistant, TREMA Lab. Mentor: Dr. Laura Dietz
University of New Hampshire, Durham, USA   *August 2017 – August 2022*

- Developed a new method for entity *aspect* linking, a fine-grained information extraction task. Showed that the approach can outperform several hard neural and non-neural baselines on a large-scale dataset by 40%. Paper published at CIKM 2022.
- Developed a method to learn query-specific entity embeddings using BERT. Showed that using query-specific entity embeddings in an entity ranking system can outperform several neural and non-neural entity ranking baselines by 13–42% on two large-scale entity ranking test collections. Paper published at SIGIR 2022.
- Demonstrated that fine-grained entity aspects can improve entity ranking performance by 50% over the state-of- the-art on a large-scale entity ranking test collection. Paper published at SIGIR 2021.
- Designed a joint probabilistic model for passage ranking which captures the joint relevance of (1) entity to query, (2) passage to query and (3) passage to entity. Achieved a 21% improvement as compared to state-of-the-art baselines for the task. Paper published at ICTIR 2019.

TEACHING EXPERIENCE

Assistant Professor, Department of Computer Science
Missouri University of Science and Technology, Rolla, MO   *August 2024 – Present*

- CS 2500 | Algorithms   Undergraduate

Teaching Assistant, Department of Computer Science
University of New Hampshire, Durham, USA   *August 2017 – August 2020*

- Assisted with the following courses:
  - CS 415 | Introduction to Computer Science I   Undergraduate
  - CS 416 | Introduction to Computer Science II.   Undergraduate
  - CS 414 | From Problems to Algorithms to Programs.   Undergraduate
  - CS 417 | From Programs to Computer Science.   Undergraduate
  - CS 515 | Data Structures and Introduction to Algorithms.   Undergraduate
  - CS 619 | Introduction to Object-oriented Design and Development.   Undergraduate
  - CS 671 | Programming Language Concepts and Features.   Undergraduate
  - CS 853 | Information Retrieval.   Undergraduate/Graduate

- CS 870 | Computer Graphics.   Undergraduate/Graduate

○ Led practical labs and recitations for undergraduate computer science students.

○ Graded assignments and tests for 150 undergraduate students.

○ Partnered with faculty members to update and co-develop two introductory computer science undergraduate courses (CS 415/416). Also developed programming assignments for these courses.

○ Developed supplementary materials for both undergraduate and graduate courses to enhance comprehension and deepen subject mastery.

○ Provided regular assistance to students through both in-person and virtual (Zoom) sessions.

○ Successfully mentored numerous female and under-represented students, guiding them towards academic achievements and fostering their research interests.

RELEVANT PUBLICATIONS

Google scholar profile: `https://scholar.google.com/citations?user=DdgpMIQAAAAJ&hl=en`

○ Thong Nguyen, **Shubham Chatterjee**, Sean MacAvaney, Iain Mackie, Jeff Dalton, and Andrew Yates. DyVo: Dynamic vocabularies for learned sparse retrieval with entities. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 767–783, Miami, Florida, USA, November 2024. Association for Computational Linguistics

○ **Shubham Chatterjee**, Iain Mackie, and Jeff Dalton. DREQ: Document Re-Ranking Using Entity-based Query Understanding. In *Proceedings of the 46th European Conference on Information Retrieval*, ECIR '24. Springer, 2024 Best Paper Honourable Mention.

○ Iain Mackie, Sean MacAvaney, **Shubham Chatterjee**, and Jeffrey Dalton. Adaptive Latent Entity Expansion for Document Retrieval. In *Proceedings of the First Knowledge-Enhanced Information Retrieval workshop*, ECIR '24, 2024

○ Hideaki Joko, **Shubham Chatterjee**, Andrew Ramsay, Arjen P. de Vries, Jeffrey Dalton, and Faegheh Hasibi. Doing Personal LAPS: LLM-Augmented Dialogue Construction for Personalized Multi-Session Conversational Search. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '24. Association for Computing Machinery, 2024

○ Mohammad Aliannejadi, Zahra Abbasiantaeb, **Shubham Chatterjee**, Jeffery Dalton, and Leif Azzopardi. iKAT: A Test Collection for Evaluating Conversational and Interactive Knowledge Assistants. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '24. Association for Computing Machinery, 2024

○ Pooja Oza, **Shubham Chatterjee**, and Laura Dietz. Neural Entity Context Models. In *Proceedings of the 12th International Joint Conference on Knowledge Graphs*, IJCKG '23. Association for Computing Machinery, 2023

○ Iain Mackie, **Shubham Chatterjee**, and Jeff Dalton. Generative Relevance Feedback with Large Language Models. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '23. Association for Computing Machinery, 2023

○ Iain Mackie, **Shubham Chatterjee**, and Jeffrey Dalton. Generative and Pseudo-Relevant Feedback for Sparse, Dense and Learned Sparse Retrieval. In *Proceedings of the Workshop on Large Language Models' Interpretation and Trustworthiness (LLMIT)*, CIKM '23. CEUR, 2023

○ **Shubham Chatterjee** and Laura Dietz. Predicting Guiding Entities for Entity Aspect Linking. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, CIKM '22. Association for Computing Machinery, 2022

○ **Shubham Chatterjee** and Laura Dietz. BERT-ER: Query-Specific BERT Entity Representations for Entity Ranking. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22. Association for Computing Machinery, 2022

○ Laura Dietz, **Shubham Chatterjee**, Connor Lennox, Sumanta Kashyapi, Pooja Oza, and Ben Gamari. Wikimarks: Harvesting Relevance Benchmarks from Wikipedia. In *Proceedings of the 45th International*

ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22. Association for Computing Machinery, 2022

- **Shubham Chatterjee**. An Entity-Oriented Approach for Answering Topical Information Needs. In *Proceedings of the 44th European Conference on Information Retrieval*, ECIR '22. Springer, 2022
- **Shubham Chatterjee** and Laura Dietz. Entity Retrieval Using Fine-Grained Entity Aspects. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21. Association for Computing Machinery, 2021
- **Shubham Chatterjee** and Laura Dietz. Why Does This Entity Matter? Support Passage Retrieval for Entity Retrieval. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, ICTIR '19. Association for Computing Machinery, 2019

INVITED TALKS AND GUEST LECTURES

Workshop on Knowledge-Enhanced Information Retrieval, ECIR 2024.	Invited Talk
BBC Data Dates Series, UK	Invited Talk
Curai Health, USA	Invited Talk
Radboud University, The Netherlands	Invited Talk
Glasgow IR Seminar Series, UK	Invited Talk
University of New Hampshire, Durham, USA	Guest Lecture

MENTORING

Neelabha Banerjee | Christ University, India	MSc Data Science
Iain Mackie | University of Glasgow, UK	PhD Computer Science
Paul Owoicho | University of Glasgow, UK	PhD Computer Science
Pooja Oza | University of New Hampshire, Durham, USA	PhD Computer Science

PROFESSIONAL SERVICE

European Conference on Information Retrieval (ECIR).	Program Committee
ACM Conference on Research and Development in Information Retrieval (SIGIR).	Program Committee
ACM International Conference on Information and Knowledge Management (CIKM).	Program Committee
European Chapter of the Association for Computational Linguistics (EACL).	Program Committee
ACL Rolling Review.	Program Committee
ACM Transactions on Information Systems (TOIS).	Manuscript Reviewer
Springer Artificial Intelligence Review (AIR).	Manuscript Reviewer
International Joint Conference on Knowledge Graphs.	Area Chair in IR
SoBigData++[1].	Proposal Reviewer

ORGANIZING

- SIGIR Virtual Forum. PhD Connect event.
- TREC Interactive Knowledge Assistance track.[2] This is shared a task on conversational IR.
- Tutorial on "Neuro-Symbolic Approaches for IR" at ECIR 2023 and SIGIR 2023.[3]

FELLOWSHIPS, AWARDS, AND HONORS

- Best Paper Honorable Mention Award. ECIR 2024.
- Outstanding Reviewer Award. ECIR 2024.

———————————————

https://plusplus.sobigdata.eu/
https://www.trecikat.com/
https://github.com/laura-dietz/neurosymbolic-representations-for-IR

- Distinguished Reviewer Award. CIKM 2023.
- Outstanding Reviewer Award. ECIR 2022.
- Dissertation Year Fellowship, University of New Hampshire, Durham. 2021.
- ACM Travel grants to attend conferences (ICTIR 2019, ECIR 2022, SIGIR 2022).
- INSPIRE Fellowship, Government of India. 2012.