# Grace Fan

## Education

| | |
|---|---|
| June 2020 - Present | **Northeastern University**<br>PhD Student in Computer Science<br>Primary Advisor: Professor Renée J. Miller<br>Secondary Advisors: Professor Mirek Riedewald, Professor Wolfgang Gatterbauer<br>Focus Areas: Data Integration, Unionable Table Search, Data Discovery |
| May 2020 | **Brown University**<br>B.S. in Computer Science<br>Undergraduate Advisor: Professor Ugur Cetintemel |
| June 2016 | **Agnes Irwin School**, PA<br>High School Graduation |

## Internships

2022 PhD Intern
**Megagon Labs**, Mountain View, CA
Helped develop research methods for both offline and online phases of problem, and experimented with different design choices, later submitting the paper to VLDB 2023

2018 Systems Analyst Intern
**Amazon Web Services**, Herndon, VA
Helped customers troubleshoot Linux and Networking problems
Assisted customers with internal issues in AWS's services
Created a website to help the team handle customer cases more efficiently.

2017 Research Intern
**Huawei**, Beijing, China
Dealt with database systems and query evaluation
Researched various diagnostic approaches, specifically with the trace facility

2015 Research Intern
**University of Edinburgh**, Edinburgh, Scotland
Helped program a new version of "The Concise Guide to Pharmacology"
Compiled biomedical data and sorted them by various similarities.
Used Python to convert it into HTML that produced a PDF resource book

## Teaching Experience

2021 Computer Science Department at Northeastern University
**Teaching Assistant** for Large-Scale Parallel Data Processing
Taught two lectures, held office hours, and led students in practice-problem sessions

2018 Computer Science Department at Brown University
**Teaching Assistant** for Computer Systems
Helped students understand concepts, plan out their code structure, and debug errors and segmentation faults.

## Research

2022 Semantics-aware Dataset Discovery from Data Lakes
with Contextualized Column-based Representation Learning
Advisors: Jin Wang, Yuliang Li, Dan Zhang, Renée J. Miller
Implemented contrastive learning to learn contextual column embeddings
for dataset discovery tasks in data lakes, and developed
efficient techniques to improve scalability

2022 SANTOS: Relationship-Based Semantic Table Union Search
Advisors: Professor Renée J. Miller, Professor Mirek Riedewald, Professor Wolfgang Gatterbauer
Found a semantic-based solution to unionable table search,
both with an external and a newly-created synthesized Knowledge base

2019 DeepSqueeze:Semantic Big Data Compression Using Deep Learning Models
Advisor: Professor Ugur Cetintemel
Used autoencoders and neural networks to compress relational data tables,
which can lead to better optimized model-based query processing.

2019 Twitter Data Analysis to Track the Spread of Ideas [Link]
Advisor: Professor Ugur Cetintemel
Parsed social media data and created visualizations to facilitate Public Policy research.

## Publications

- **Grace Fan**, Jin Wang, Yuliang Li, Dan Zhang, and Renée J. Miller. "Semantics-aware Dataset Discovery from Data Lakes with Contextualized Column-based Representation Learning." under review for VLDB 2023 [ArXiv]
- Aamod Khatiwada, **Grace Fan**, Roee Shraga, Zixuan Chen, Wolfgang Gatterbauer, Renée J. Miller, Mirek Riedewald: SANTOS: Relationship-based Semantic Table Union Search. SIGMOD Conference 2023 [Preprint]
- **Grace Fan**, Wenfei Fan, Yuanhao Li, Ping Lu, Chao Tian, Jingren Zhou: Extending Graph Patterns with Conditions. SIGMOD Conference 2020: 715-729 [Link]
- Amir Ilkhechi, Andrew Crotty, Alex Galakatos, Yicong Mao, **Grace Fan**, Xiran Shi, Ugur Çetintemel: DeepSqueeze: Deep Semantic Compression for Tabular Data. SIGMOD Conference 2020: 1733-1746 [Link]
- Stephen Ph. Alexander, Eamonn Kelly, Neil Marrion, John A. Peters, Helen E. Benson, Elena Faccenda, Adam J. Pawson, Joanna L. Sharman, Christopher Southan, Peter Buneman, William A. Catterall, John A. Cidlowski, Anthony P. Davenport, Doriano Fabbro, **Grace Fan**, John C. McGrath, Michael Spedding, Jamie A. Davies, "The Concise Guide to Pharmacology 2015/2016." by The British Journal of Pharmacology, Oct 2015 *PMID: 26650438*
- **Grace Fan**, Wenfei Fan, Floris Geerts: "Detecting Errors in Numeric Attributes". WAIM 2014: 125-137. [Link]

## Computer Languages

Python, Java, SQL, C, C++, Go, Scala, LaTeX, React, HTML, Javascript

## Interests and Activities

| | |
|---|---|
| Northeastern University | Member of PhD Women group |
| Brown University | Mentor for the Women in Computer Science; |
| | Member of the Women in Science and Engineering |
| Other Academic Interests | Cognitive Science and Psychology |
| Other Interests | Writing Poetry; Photography & Traveling |