

Faculty Candidate Seminar – Professor of Engineering Practice



Ruby Y. Tahboub

Visiting Assistant Professor
Purdue University, Dept. of Computer Science

Wednesday, April 20, 2022

10:30 A.M. – 11:30 A.M.

BHEE 317

Introduction to Data Science

Practice Abstract: In this teaching demo, I will give an introduction to data science. We will discuss examples of data science applications (e.g., the missing bullet holes on the missing airplanes in World War II and special education) and the overall data science process. After that, we will go over various data engineering tasks (e.g., filtering and aggregating data) and process real data about COVID-19 cases in US colleges using Pandas library in Jupyter notebook.

Code examples will be available here: <https://drive.google.com/drive/folders/1OJK8HKObgGnu-4TdgcCWb9rWYjKjCgUO?usp=sharing>

Architecting Query Compilers for Diverse Workloads

Lecture Abstract: More data management systems have embraced the compilation of query plans to native code to leverage modern hardware platforms to their fullest. Yet, there is an ongoing debate about how best to architect such query compilers. Query compilation is challenging, requiring techniques fundamentally different from traditionally interpreted query execution. In this talk, I will draw attention to an old but underappreciated idea known as Futamura projections, which fundamentally link interpreters and compilers. Guided by this idea, realizing efficient query compilation can be very simple, using techniques that are no more difficult than writing a query interpreter in a high-level language. To support compiling relational workloads, I will discuss the design of the LB2 query compiler implemented in this generative style that is competitive with the best compiled query engines.

Furthermore, spatial query engines are implemented as extensions to relational query engines to leverage optimized memory, storage, and evaluation. Still, the performance of these extensions is often stymied by the interpretive nature of the underlying data management, generic data structures, and the need to execute domain-specific external libraries. Compiling spatial queries to native code is a desirable avenue to mitigate existing limitations and improve performance. I will discuss the design of LB2-Spatial that extends LB2 with spatial indexing structures and operators to support compiling spatial workloads.

Bio: Ruby Tahboub is a Visiting Assistant Professor in the Computer Science Department at Purdue University. She earned her Ph.D. from Purdue University in 2019. Her research work unites data management and generative programming to efficiently compile query code at runtime.

Ruby is passionate about teaching and making the learning process more effective. She aims to help students construct the link between abstract concepts and practice while minimizing the cognitive load. To that end, she incorporates visualization and storytelling in her teaching materials. She received Purdue's prestigious Raymond Boyce Graduate Teacher Award for her excellence in teaching.