

App ID #: 899

Mentor: Bunescu, Razvan

Email: rbunescu@uncc.edu

Title: Associate Professor

Department: Computer Science

Co-mentor: No

Community engaged research: No

Title: The CS1 Compendium: Acquisition, Visualization, and Knowledge Tracing

Description: The overall aim of this project is to:

Extend a knowledge base of the coding concepts introduced in an introductory course in programming, e.g. coding in Python. This knowledge base, called the CS 1 Compendium (CS1C) will connect coding concepts (e.g. loops, assignments, variables, ...) through edges representing conceptual relationships, such as knowing concept A requires knowledge of concept B. The CS1C will also contain nodes for lectures, assignments, quizzes, questions, and answers, all interconnected.

We already have an initial, working version of this graph.

Store the CS1C in a non-relational database, such as neo4j.

Build a tool that visualizes the database, including capabilities for focusing on a particular part of the knowledge graph, for summarizing a large section of it, or for querying it using LLM-based capabilities.

Utilize the CS1C knowledge base to do knowledge tracing, i.e. trace the knowledge acquired by a simulated student (as well as their misconceptions) as they go through course materials.

We currently have an initial approach based on graph neural networks.

Looking for: Only 160 hours over an academic semester (~10h/wk)

of positions available: 2

Anticipated Student Learning Outcomes: Depending on which part of the project the student contributes to, they will:

Develop skills for developing ontologies and representing knowledge into a graphical format.

Learn about storing data in non-relational databases and accessing and visualizing it through Python code.

Develop and evaluate machine learning (ML) models in PyTorch, using existing libraries.

The students will learn to conduct research in a collaborative setting; they will write weekly meeting reports; they will run empirical evaluations; they will write project reports and ideally contribute to the writing of a paper to be submitted for peer-review at a relevant venue.

Upon a successful completion of the project, the student will have developed skills in knowledge representation, databases, machine learning (PyTorch) and language processing (LLMs) that are highly valued in academic and industry settings.

Training Description: No training required.

Training hours: 0

Mentoring plan: We will have weekly faculty-student meetings, with weekly reports and code to be submitted on <https://cci-git.charlotte.edu/>.

The student will also interface with a PhD student who has already done initial work on the project. The student will be expected to also attend the bi-weekly group AI readings.

Applicant Requirements: Student is expected to be comfortable with coding and have working knowledge of data structures and algorithms. Student needs to be self-motivated and have a passion for education and an strong interest in natural language processing (NLP) and machine learning (ML).

Applicant Preferences: Completion of (or registration into) a natural language processing course (e.g. ITCS 4101) or machine learning course (e.g. ITCS 3156) would be a significant plus.

Specific Time considerations/conflicts: No conflicts.