



Thesis Defense

Computer Science Master's Program

“Semantic Structuring of Digital Documents: Knowledge Graph Generation and Evaluation”

By Erik Luu

Abstract:

In the era of total digitization of documents, navigating vast and heterogeneous data landscapes presents significant challenges for effective information retrieval, both for humans and digital agents. Traditional methods of search and organization often struggle to keep pace with evolving user demands, resulting in suboptimal outcomes such as information overload, irrelevant results, and disorganized data. This thesis introduces a pipeline that leverages principles from cognitive science, graph theory, and semantic computing to generate semantically organized knowledge graphs. By evaluating the combination of different models, methodologies, and algorithms, the pipeline aims to enhance the organization and retrieval of digital documents. The proposed approach focuses on representing documents as vector embeddings, clustering similar documents, and constructing a connected and scalable knowledge graph. This graph not only captures semantic relationships between documents, but also ensures efficient traversal and exploration. The effectiveness of the pipeline is demonstrated through a series of experiments, highlighting its potential to improve information retrieval and knowledge organization in the digital age.

Date: Monday, June 3rd, 2024

Time: 2:30 PM – 4:00 PM

Location: 14-232b

Committee: Dr. Canaan, Dr. Khosmood, and Dr. Kurfess

