NJIT

Department of Computer Science

Loch Data and Other Monsters: on Creating Data Ecosytems, the Intelligent Way

Avigdor Gal
Technion - Israel Institute of Technology

Hosted by Shantanu Sharma

Date: Wednesday, September 21, 2022
Coffee: 2:15 PM – 2:30 PM
Seminar: 2:30 PM – 3:30 PM
Location: GITC 4402
Webex Link https://njit.webex.com/njit/j.php?MTID=mb3b10c488b42b2a8dde454e92fd3d133

https://cs.njit.edu/seminars

Abstract:
A data ecosystem offers an alliance-driven infrastructure that enables the interaction of different stakeholders and the resolution of interoperability issues among shared data. Despite years of research in data governance and management, trustability is still affected by the absence of transparent and traceable data-driven pipelines. Data integration is the main facilitator of such data-driven pipelines and matching is a task at the heart of any data integration process, aimed at identifying correspondences among data elements. Matching problems were traditionally performed in a semi-automatic manner, with correspondences being generated by matching algorithms and outcomes subsequently validated by human experts. Human-in-the-loop data integration has been recently challenged by the introduction of big data and recent studies have analyzed obstacles to effective human matching and validation. In this talk, we focus on the tension between human and machine matching. We propose a novel data ecosystem architecture that relies on both human knowledge and machine learning and offer a concrete algorithmic solution for effective data integration within this architecture. In particular, we shall present the limitations of human matching and offer a method for learning to characterize reliable and valuable matching experts.

Bio: Avigdor Gal is The Benjamin and Florence Free Chaired Professor of data science at the Technion - Israel Institute of Technology, where he heads the Big Data Integration laboratory. He specializes in various aspects of data management and mining with about 150 publications in leading journals, books, and conference proceedings. In the current age of big data, his research is focused on developing novel models and algorithms for data integration. In the past he gave keynotes and tutorials in leading conferences in the areas of data and process management. Avigdor Gal is a recipient of the prestigious Yannai award for excellence in academic education, and multiple best paper and test-of-time awards.