

COLLOQUIUM

Taming the Big Data Elephant with Query Explanations

Sudeepa Roy
University of Washington

Abstract: In recent years, the availability of big data has resulted in a growing number of users from a variety of backgrounds interested in identifying and interpreting the general trends and anomalies of large datasets. This presents an imminent requirement of sophisticated data analysis tools that can provide qualitative information based on query answers on such datasets. In this talk, I will describe my current research on developing a principled framework for explaining query answers in terms of intervention (explanations are changes in the database that change the observed query answers). I will present our solutions to core challenges in this task such as obtaining concise descriptions of explanations, handling inherent dependencies of database tuples, and achieving real-time efficiency in large explanation spaces. Then, I will briefly talk about my research in the areas of probabilistic databases, provenance, information extraction, and crowd sourcing. The unifying theme of this research is to address defining characteristics of modern datasets: uncertainty, unreliability, lack of structure, and the effects of human participation. I will conclude with my long-term vision of incorporating techniques to handle these challenges in the generic data explanation framework.

Monday, February 23, 2015, 4:00 pm
Mathematics and Science Center: W303

MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY