

NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

C S S E M I N A R

Title: Explore-By-Example: A New Database Service for Interactive Data Exploration

Speaker: Yanlei Diao
Professor of Computer Science
Ecole Polytechnique
France

Date/Time: 1 August 2016, Monday, 10:30 AM to 12:00 PM

Venue: SR2, COM1-02-04

Chaired by: Dr Tan Kian Lee, Shaw Senior Professor, School of Computing
(tankl@comp.nus.edu.sg)

Abstract:

Traditional DBMSs are suited for applications in which the structure, meaning and contents of the database, as well as the questions (queries) to be asked, are all well-understood. However, this is no longer true when the volume and diversity of data grow at an unprecedented rate, while the user ability to comprehend data remains (as limited) as before. To address this problem, our project explores a new approach of system-aided exploration of a big data space and automatic learning of the user interest in order to retrieve all objects that match the user interest - we call this new service "interactive data exploration", which complements the traditional querying interface of a database system.

In this talk, I introduce a new framework for interactive data exploration, called "Explore-by-Example", which iteratively seeks user relevance feedback on database samples and uses such feedback to finally predict a query that retrieves all objects of interest to the user. The goal is to make such exploration converge fast to the true user interest model, while minimizing the user labeling effort and providing interactive performance. I discuss a range of techniques and optimizations to do so for linear patterns and complex non-linear patterns. Our user study indicates that our approach can significantly reduce the user effort and the total exploration time, compared with the common practice of manual exploration. I finally conclude the talk by pointing out a host of new challenges, ranging from application of active learning theory, to database optimizations, to visualization.

Biodata:

Yanlei Diao is Professor of Computer Science at Ecole Polytechnique in France. Prior to that, she was Associate Professor at the University of Massachusetts Amherst, USA. Her research interests lie in information architectures and data management systems, with a focus on big data analytics, data stream processing, interactive data analysis, uncertain data management, and sensor and scientific data management. She received her PhD in Computer Science from the University of California, Berkeley in 2005.

Prof. Diao was a recipient of the 2013 CRA-W Borg Early Career Award (one female computer scientist selected each year for outstanding contributions in research and outreach), IBM Scalable Innovation Faculty Award, and NSF Career Award, and she was a finalist of the Microsoft Research New Faculty Award. She spoke at the Distinguished Faculty Lecture Series at the University of Texas at Austin. Her PhD dissertation "Query Processing for Large-Scale XML Message Brokering" won the 2006 ACM-SIGMOD Dissertation Award Honorable Mention. She is currently PC Co-Chair of the IEEE ICDE 2017 and ACM SoCC 2016 conferences. She is Editor-in-Chief of the ACM SIGMOD Record, Associate Editor of ACM TODS, Chair of the ACM SIGMOD Research Highlight Award Committee, member of the SIGMOD Executive Committee and VLDB Endowment, and member of SIGMOD Software Systems Award Committee.