Spring 2015 Seminar Series
Presented by the CS Division

TOWARDS INTERACTIVE LARGE-SCALE DATA ANALYTICS PLATFORM FOR DIVERSE DATA SOURCES

WEDNESDAY MARCH 18, 2015

10:00 AM – HEC 356

"Big Data" research directions are often classified into 3 V's: Too much data (Volume), data is coming from too many sources (Variety), or data is appearing too fast (Velocity) [Stonebraker NIST'10]. I will talk about our work on a new interactive large-scale analytics/fusion platform for diverse data sources that addresses these 3Vs. I will start from the architecture, describe new data models and algorithms, and then dive deeper into several key components. If time permits I go over several significant applications, built on top of this infrastructure, including Hemolix - a Nokia Global 2.25M XPrize 2014 Finalist. The talk concludes with an overview of unique Web-scale datasets used in the experiments and future research plans.

DR. MICHAEL GUBANOV
MIT

Dr. Gubanov is finishing his postdoctoral research at MIT working on a new large-scale interactive data analytics platform for diverse data sources. He earned his Ph.D. from the University of Washington, where his thesis research was invited to publication in a book on major advances in Data Management and Information Retrieval published by Springer in 2013. During the Ph.D. study, he was an intern at Google, where he designed and built one of the first prototypes of a Web-scale machine learning system; and at IBM Almaden Research Center where he co-invented UFO, one of the first data structures that enables Web-scale data integration. Finally, Hemolix, a mobile disease diagnosis system, built on top of this new data analytics platform got to Nokia Global 2.25M XPrize Finals held in 2014 in Los Angeles, CA and earned wide press coverage. To date he co-authored more than 20 conference/journal publications, and published 2 invited Springer book chapters.

Hosted by: Dr. Lotzi Boloni