Massive Data Management

Organizers: Cyrus Shahabi, University of Southern California, Takeshi Oishi, University of Tokyo

In recent years, we have been witnessing the unfolding of two phenomena—one concerning social issues and the second concerning technological challenges. We are increasingly facing major environmental, socioeconomic and man-made crises—global warming, earthquakes, hurricanes, oil spills and excessive energy consumption. On the other hand, advances in sensor and IT technologies and their reduced costs make it possible to collect data about the real world and its various phenomena at increasingly exponential rates. Example datasets include high-resolution 2D/3D imagery, distributed climate and pollutant data, traffic and CCTV video data, data from mobile devices, social-networking data. These data can be used to address many global challenges such as climate issues or improve social conditions if we are capable of managing and fusing these datasets for practical applications. These data also offer global benefits by allowing different cultures to apply these vast data to understand their interdependencies and to respond to catastrophic crises more effectively. Unfortunately, today there is a large gap between the data that are needed and the data that are available. These data are not available at the time, space, quality, resolution and in the form needed. Data are not integrated properly with other relevant datasets nor as quickly as needed.

In this session, we focus our attention on the massive datasets that are recently becoming available, such as web-sources (e.g., facebook, tweets, blogs, searches), datasets from professional sensors (e.g., LiDAR, satellite/aerial imagery, CCTV cameras) and participatory sensors (e.g., people's mobile devices collecting GPS, Pictures, audio, etc.), and then discuss some of the new techniques and algorithms developed to deal with these data to enable applications in various domains such as intelligent transportation, climate change and archaeological analysis.

In particular, we start by Prof. Johannes Gehrke who will describe the challenges in organizing, managing, and mining these data, drawing upon his group's experience in developing some of the fastest data mining algorithms and building a scalable platform for micro-simulations of the world's largest transportation networks. Next speaker, Prof. Asanobu Kitamoto will focus on managing a variety of data sources that is spatially distributed on Eurasian Continent for reconstructing *Digital Silk Road*. Then, Prof. Yan Liu will focus on managing the time-series data for modeling and visualizing climate change, and will discuss the central role of the massive data management and data mining techniques in devising potential solutions to climate change. Finally, we end our session with a presentation by Prof. Yasuhide Okamoto who will introduce their attempt to collect and manage massive datasets by utilizing*Cloud Computing* for modeling and rendering of large scale cultural heritage sites.