Big Data Curation

(Keynote)
Prof. Renée Miller
Department of Computer Science
University of Toronto

miller@cs.toronto.edu

ABSTRACT

A new mode of inquiry, problem solving, and decision making has become pervasive in our society, consisting of applying computational, mathematical, and statistical models to infer actionable information from large quantities of data. This paradigm, often called Big Data Analytics or simply Big Data, requires new forms of data management to deal with the volume, variety, and velocity of Big Data. Many of these data management problems can be described as data curation. Data curation includes all the processes needed for principled and controlled data creation, maintenance, and management, together with the capacity to add value to data. In this talk, I describe our experience in curating some open data sets. I overview how we have adapted some of the traditional solutions for aligning data and creating semantics to account for (and take advantage of) Big Data.

Biography

Prof. Rene Miller received BS degrees in Mathematics and in Cognitive Science from the Massachusetts Institute of Technology. She received her MS and PhD degrees in Computer Science from the University of Wisconsin in Madison, WI. She is a Fellow of the Royal Society of Canada (Canada's National Academy) and the Bell Canada Chair of Information Systems at the University of Toronto. She received the US Presidential Early Career Award for Scientists and Engineers (PECASE) , the highest honor bestowed by the United States government on outstanding scientists and engineers beginning their careers and the National Science Foundation Career Award. She is a Fellow of the ACM, a former President of the VLDB Endowment, and was the Program Chair for ACM SIGMOD 2011 in Athens, Greece. Her work has focused on the long-standing open problem of data integration and has achieved the goal of building practical data integration systems. She was a co-recipient of the ICDT Test-of-Time Award for an influential 2003 paper establishing the foundations of data exchange.