

On Schema Discovery

Renée J. Miller

Professor of Computer Science
Bell University Labs Chair of Information Systems
University of Toronto.

April 12, 2012

1:30 PM

Lawson 3102

Speaker Bio

Renée J. Miller received BS degrees in Mathematics and in Cognitive Science from the Massachusetts Institute of Technology (MIT). She received her MS and PhD degrees in Computer Science from the University of Wisconsin in Madison, WI. She received the 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. She received the National Science Foundation Early Career Award. She is a Fellow of the ACM, the President of the VLDB Endowment, and the Program Chair for ACM SIGMOD 2011. Her research interests are in the efficient, effective use of large volumes of complex, heterogeneous data. This interest spans data integration, data exchange, knowledge curation and data cleaning. She is a Professor and the Bell Canada Chair of Information Systems at the University of Toronto. In 2011, she was elected to the Fellowship of the Royal Society of Canada (FRSC), Canada's National Academy of Science.

Presentation Abstract

Structured data is distinguished from unstructured data by the presence of a schema describing the logical structure and semantics of the data. The schema is the means through which we understand and query the underlying data. Schemas enable data independence. In this talk, I consider new challenges in the old problem of schema discovery. I'll discuss the changing role of schemas from prescriptive to descriptive. I'll use examples from Web data publishing and from Business Analytics to motivate the automation of schema discovery and maintenance.