

Provenance Management: Challenges and Opportunities

Juliana Freire

School of Computing
University of Utah
Salt Lake City, Utah, USA

juliana@cs.utah.edu

Abstract: Computing has been an enormous accelerator to science and industry alike and it has led to an information explosion in many different fields. The unprecedented volume of data acquired from sensors, derived by simulations and data analysis processes, accumulated in warehouses, and often shared on the Web, has given rise to a new field of research: provenance management. Provenance (also referred to as audit trail, lineage, and pedigree) captures information about the steps used to generate a given data product. Such information provides important documentation that is key to preserve data, to determine the data's quality and authorship, to understand, reproduce, as well as validate results. Provenance solutions are needed in many different domains and applications, from environmental science and physics simulations, to business processes and data integration in warehouses. In this talk, we survey recent research results and outline challenges involved in building provenance management systems. We also discuss emerging applications that are enabled by provenance and outline open problems and new directions for database-related research.