



Department of Computer Science and Engineering

Presents

**Val Tannen, University of Pennsylvania**

**Data Provenance and Query Languages**

Data provenance is the history of "ownership" of various pieces of information involved in a scientific study or a business data warehouse. Such data is sometimes mapped from data sources through views that can be described using query languages. In this case, scientists or business intelligencers are interested in also tracking through the views the provenance of the data. Data provenance can be used, for example, as part of decision-making in a multi-source environment, when certain facts must be rejected because they can only be derived from "tainted" data, while others are accepted because they can also be derived from trusted data.

Using a novel framework based on the algebraic structure of semirings, we discuss provenance tracking for relational data through relational algebra and datalog views. It turns out that this also generalizes nicely calculations done for incomplete and probabilistic databases, as well as data multiplicity, a surprising connection.

Joint work with T.J. Green, and G. Karvounarakis

**Thursday, October 25, 2007**

**3:30 - 4:30 PM**

**Theater Room 201**

**University at Buffalo – North Campus**

This talk is free and open to the public. Refreshments for attendees after the talk in 224 Bell Hall

For more information, please email [cse-dept@cse.buffalo.edu](mailto:cse-dept@cse.buffalo.edu) or contact (716) 645-3180