

COMPUTER SCIENCE
SEMINAR

*Shaping your research so it can transition to practice: a view
from data integration research*

Arnon (Arnie) Rosenthal
MITRE

Abstract: Data integration technology aims to transform data from the providers form to a form consumers can use, and also to merge data from multiple providers. Relevant theory and algorithms appeared as early as the 1980s, but until very recently the transition to products was highly disappointing

After an introduction to the challenges of data integration, we examine research areas whose results were difficult to transfer. From these, we identify two generic tactics for formulating good research problems whose theoretical results will also be exploitable by product planners and by development teams. First, if you have a theoretical results that applies to constrained, simplified problems, extend it to be /somewhat /useful on systems that violate the constraint. Second, work downstream first tackle the last challenge that blocks creating runnable code with large user bases otherwise your results may stay on the shelf for decades (e.g., schema matching circa 1985-2008).

Finally, we will examine how well research has aligned with needs (i.e., areas desperately needing models and technique to clarify them. While there has been some terrific (and terrifically useful) recent research on data integration (e.g., IBM, Microsoft, Google Research), we will describe our pain points terrifically important challenges for tractable research problems have not yet been formulated (let alone solved).

Friday, February 27, 2009, 3:00 pm
Mathematics and Science Center: W301

MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY