

## **Computer Science Seminar Series, 2011**

**National Capital Region**

# **Domain Discovery for Web Databases**

**Speaker: Prof. Nan Zhang**  
**Department of Computer Science**  
**George Washington University**

**Friday, March 4, 2011**  
**1:00PM-2:00PM, NVC 325**

### **Abstract**

Many web databases are hidden behind restrictive form-like interfaces which may or may not provide domain information for an attribute. When attribute domains are not available, domain discovery becomes a critical challenge facing the application of a broad range of existing techniques on third-party analytical and mash-up applications over hidden databases. In this talk, I shall discuss the problem of domain discovery over a hidden database through its web interface. I shall show that for any database schema, an achievability guarantee on domain discovery can be made based solely upon the interface design. I shall also describe novel techniques which provide effective guarantees on the comprehensiveness of domain discovery, and present theoretical analysis and experiment results which illustrate the effectiveness of these techniques.

### **Biography**

Dr. Nan Zhang is an Assistant Professor of Computer Science at the George Washington University, Washington, DC, USA. Prior to joining GWU, he was an assistant professor of Computer Science and Engineering at the University of Texas at Arlington from 2006 to 2008. He received the B.S. degree from Peking University in 2001 and the Ph.D. degree from Texas A&M University in 2006, both in computer science. His current research interests include databases, information security and privacy, and data mining. He received the NSF CAREER award in 2008.