

**THE DEPARTMENT OF COMPUTER SCIENCE & THE COMPUTER SCIENCE  
GRADUATE STUDENT ORGANIZATION (GSOCS) PRESENT**

## **INVITED SPEAKER SERIES**

**Professor Eduard Dragut  
Temple University**

**Friday, March 30th at 12 noon, Engineering Building Room R15**

### **Leveraging Social Media Signals for Record Linkage**

**Abstract:** With the advent of Web 2.0, web sites actively involve visitors in the provision of content, called user generated content. User generated content takes multiple forms, e.g., video sharing, products, social networking, and reviews. We focus on user reviews. With the nascence of social media, entities (e.g., hotels, restaurants, books) presented on web sites are now accompanied by user generated content in the form of reviews. Many data-intensive applications collect and integrate such data from a variety of Web sources. A key task in this process is entity matching, which is the problem of determining the records from these sources that refer to the same real-world entities. Traditional approaches use the record representation of entities to accomplish this task. We argue that this hitherto untapped source of entity information can be used in entity matching. In this talk, we present opportunities, challenges, and preliminary results in entity matching with user generated content.

**Bio:** Eduard Dragut is an Assistant Professor in the Computer and Information Sciences Department at Temple University. He previously was a Postdoctoral Research Associate at Purdue University, Discovery Park, Cyber Center. He completed his Ph.D. degree in Computer Science in July 2010 from University of Illinois at Chicago. His main area of research is Web data management (e.g., extraction, representation, analysis, integration), with particular emphasis on the data from Web databases. He is actively pursuing projects in Entity Mention Detection and Linking in Social Media, Web Data Cleaning, Sentiment Analysis, and Cyber-Infrastructure for Scientific Research.