

**Amgen Seminar Series in Chemical Engineering**  
in  
Cherry Auditorium, Kirk Hall, 1 PM

**Presents on February 18, 2010**

**Cross-document Cross-lingual Information Extraction and Tracking**

By

Professor Heng Ji  
Department of Computer Science  
Queens College & The Graduate Center of City college of NY

Most current information extraction analyzes documents in isolation. The net result is a set of disconnected, inaccurate and often redundant annotations, because events are repeated in many news stories. In this talk we will present a new task of cross-document cross-lingual information extraction and tracking and its evaluation metrics. From enormous multi-lingual documents we identify important person entities which are frequently involved in events as 'centroid entities'. Then we link the events involving the same centroid entity along a time line. We will also present a system performing this task and our current approaches to address the main research challenges. We will discuss how we can take advantage of redundancy to improve the accuracy of relation and event annotation, by means of

- Cross-document event coreference resolution
- Event ranking by salience and novelty, and
- Event organization by participant, time, and place
- Name translation
- Knowledge Discovery from Google Ngrams
- Domain Adaption Techniques for Applying Information Extraction to Scientific Literature

This series at the University of Rhode Island is made possible through the generosity of Amgen, West Greenwich, R.I.