

# UW CENTER FOR PATTERN ANALYSIS AND MACHINE INTELLIGENCE

## GRADUATE SEMINAR SERIES

### Automatic Document Topic Identification Using Hierarchical Ontology Extracted from Human Background Knowledge

**Speaker:** Mostafa Adel

**Date:** February 6, 2013

**Time:** 4:30 pm – 5:00 pm

**Place:** E5 (5128) Refreshments will be served

#### **Abstract :**

The rapid growth in the number of documents available to end users from around the world has led to a greatly-increased need for machine understanding of their topics, as well as for automatic grouping of related documents. This constitutes one of the main current challenges in text mining. In this work, a novel technique is proposed, to automatically construct a background knowledge structure in the form of a hierarchical ontology, using one of the largest online knowledge repositories: Wikipedia. Then, a novel approach is presented to automatically identify the documents' topics based on the proposed Wikipedia Hierarchical Ontology (WHO). Results show that the proposed model is efficient in identifying documents' topics, and promising, as it outperforms the accuracy of the other conventional algorithms for document clustering.