Computer Science Colloquium

"Biomedical Text Processing, basic research articles and clinical notes, challenges and opportunities"

Dr. Lucy Vanderwende
Microsoft Research

Abstract

The importance of language processing has been widely recognized in the biomedical research and clinical domains. The opportunity to make contributions in this field is now increasingly recognized by natural language processing researchers as well. Jointly, these communities have created a number of resources for basic research articles, among which is annotated data representing change of state for bio-molecules, made available as the BioNLP shared task. In this talk, I will discuss this data set, outline why this data is particularly challenging, and describe the system that our group at Microsoft Research built to perform the task of extracting bio-events from text. Another considerable challenge is the text processing of clinical records. In this talk, I will discuss the logistic challenges of working with clinical data, and I will present the results of joint work with University of Washington Department of BioInformatics to identify patients with pneumonia based on text processing of clinical notes. We have demonstrated that assertion-based features significantly improve our pneumonia prediction results, which will result in improved health care outcomes.

Biography

Dr. Vanderwende received her Ph.D. in Computational Linguistics from Georgetown University, in Washington D.C. She has worked at IBM in Natural Language Processing (1988 - 1990), and she was a Visiting Scientist at the Institute for Systems Science in Singapore (1991). Since 1992, she has been a manager and senior researcher in the Natural Language Processing group at Microsoft Research. Among the products Dr. Vanderwende has contributed to are the Microsoft Word grammar checker and the Encarta Natural Language User Interface. Early research work includes MindNet, a graphical semantic knowledge base through automatic text extraction, and the analysis of noun compounds. She is currently focusing on a new task definition for Summarization and on fine-grained Information Extraction from general text, and from biomedical text in particular.

Date: Friday, March 30, 2012
Time: 10:30am to 11:30am
Location: ECS South 2.410

Refreshments will be served at 10:15am