Biomedical and Health Informatics Lecture Series

Tuesday, May 18, 2010
12:00 - 12:50 p.m., Room RR-134

Meliha Yetisgen-Yildiz, PhD

Assistant Professor, Division of Biomedical and Health Informatics
University of Washington, Seattle

"Corpora Generation for Clinical/Biomedical Natural Language Processing"

A flood of high throughput genetic testing tools (genome wide single nucleotide polymorphism chips, whole exome sequencing and on the horizon whole genome sequencing) is currently available. These high throughput technologies allow researchers to get a very good picture of the underlying genetic variability in a population of patients. In order to understand the genetic contribution to clinical conditions of interest (diseases or pre-disposition to disease), it is key to characterize and codify patients whose genetic background is known or in other words to create a coded representation of their phenotype. Relying on the coded problem lists in medical records does not work well to define the phenotype. Because the coding is usually generated by the medical billing process rather than by the clinical care process, it is not as precise as one ideally would like to see. Furthermore, the coding does not capture key information such as family history, diet, exercise, and other factors that influence health (e.g. smoking, drinking). Such information is generally represented in free text and not recorded as discrete data elements. In my research, I focus on applying statistical natural language processing approaches to extract information from medical records and biomedical literature that will be used in translational genetic research. One of the main challenges is to produce high quality annotated corpora in an efficient and effective way. In this talk, I will talk about my current work on corpora generation by using machine learning approaches and crowd sourcing environments.

Meliha Yetisgen-Yildiz received her BS degree on Computer Engineering and Information Science from Bilkent University (Ankara, Turkey) and MS degree on Computer Engineering from Middle East Technical University (Ankara, Turkey). She received her PhD from the University of Washington with a thesis on biomedical text mining in December, 2007. Before joining to BHI, she worked as a post-doctoral researcher at UW and as a text mining researcher at Kiha, Inc. Her current research interests include clinical natural language processing, biomedical text mining, and information extraction.

* *

NOTE: Podcasts from MEBI 590 Lecture Series talks for this quarter are available at
http://courses.washington.edu/mebi590/schedule.htm

Podcasts from previous quarters are available at
http://courses.washington.edu/mebi590/past.lecture.schedules.html