



**Biomedical and Health Informatics Lecture Series**  
Course Website: [Link](#)

**Tuesday, April 17, 2012**  
**12:00 - 12:50 p.m., Room T-663**

**Cosmin Adrian Bejan, PhD**

Senior Fellow, Post-doctoral Researcher  
Biomedical and Health Informatics  
Department of Biomedical Informatics and Medical Education  
University of Washington, Seattle

**“Classification of medical assertions and  
its application to phenotype identification”**

The availability of comprehensive electronic medical records that include narrative reports provides an opportunity for natural language processing technologies to play a major role in clinical research. One of the main advantages of employing these technologies is the automatic extraction of relevant clinical information to identify critical illness phenotypes and to facilitate clinical and translational studies of large cohorts of critically ill patients. In the first part of this talk, I will describe a supervised learning approach to assertion classification, a task formulated such that each medical concept mentioned in a clinical report (e.g., pneumonia, chest pain) is associated with a specific assertion category (e.g., present, absent, possible). To improve the performance of the assertion classifiers, we introduce several semantic features trying to capture the meaning of specific keywords which convey various forms of information that trigger a specific assertion category. In the second part of the talk, I will present an empirical study on the impact assertion classification has on phenotype identification, a real word application from the clinical domain.

---

Cosmin Adrian Bejan is a senior fellow in the Department of Biomedical Informatics and Medical Education at the University of Washington. Prior to his current position, Cosmin was a postdoctoral fellow in the Institute for Creative Technologies at the University of Southern California. Cosmin received his M.S. and B.S. degrees in computer science from the University of Iasi, Romania. He holds a Ph.D. degree in computer science from the University of Texas at Dallas, where he investigated natural language processing and machine learning methodologies in order to capture the semantics of the event structures that are encoded in text. His research interests are in the areas of natural language processing, biomedical informatics, and machine learning with a focus on event semantics, open-domain and clinical information extraction, and commonsense causal reasoning.

\*\*\*\*\*

**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>