



BRIGHAM YOUNG  
UNIVERSITY  
CS Department Colloquium Series



## Raymond J. Mooney

University of Texas at Austin

Thursday, November 16, 2006

1170 TMCB, 11:00 AM

### Learning to Extract Proteins and their Interactions from Biomedical Text

#### Abstract

Automatically extracting information from biomedical text holds the promise of easily consolidating large amounts of biological knowledge in computer-accessible form. This strategy is particularly attractive for extracting data on human genes from the 11 million abstracts in Medline. We have developed and evaluated a variety of learned information-extraction systems for identifying human proteins and their interactions in Medline abstracts. We will present our current best results on identifying names of human proteins using Conditional Random Fields and Relational Markov Networks. We will also present our current best results on identifying interactions between proteins using a Support Vector Machine with an underlying string kernel. Finally, we will summarize results from a recent large-scale application of our techniques, in which we mined 753,459 Medline abstracts to extract a database of 6,580 interactions between 3,737 human proteins. By merging this extracted data with existing databases, we have constructed (to our knowledge) the largest database of known human-protein interactions containing 31,609 interactions amongst 7,748 proteins.

Joint work with Razvan Bunescu, Edward Marcotte, Ruifang Ge, Rohit Kate, Yuk-Wah Wong, and Arun Ramani.

#### Biography

Raymond J. Mooney is a Professor in the Department of Computer Sciences at the University of Texas at Austin. He received his Ph.D. in 1988 from the University of Illinois at Urbana/Champaign. He is an author of over 100 published research papers, primarily in the area of machine learning. He was program co-chair of the 2006 National Conference on Artificial Intelligence, general chair of the 2005 joint Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, co-chair of the 1990 International Conference on Machine Learning, a recipient of the Best Research Paper Award at the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, a former editor of the Machine Learning journal, and a Fellow of the American Association for Artificial Intelligence. His recent research has focused on learning for natural-language processing, text mining, statistical relational learning, active learning, semi-supervised learning, learning for record linkage, and bioinformatics.

**Donuts will be provided**