

# A Text-mining Technique for Literature Profiling and Information Extraction from Biomedical Literature

by Hisham Al-Mubaid

**M**ASSIVE AMOUNTS OF BIOMEDICAL literature are readily available online to researchers in many forms: text abstracts, Medline with more than 16 million biomedical abstracts, full text research articles, databases of protein interactions, dictionaries of gene and protein names, and other electronic databases. Huge amounts of valuable knowledge and useful information are embedded in these resources and available to be properly extracted, discovered, and utilized. There is a great need for computational techniques to utilize and extract the useful knowledge from these resources. A number of systems and software tools have been developed to utilize these extensive resources. Biomedical research has shown that text mining can be effective in this field, making text mining increasingly important and necessary for biology and medicine.

This project aims at investigating and designing effective computational methods for literature profiling to extract and organize important information and related data from the biomedical literature. For that, we implemented new methods for identifying and classifying technical terms and entity names in biomedical texts. The methods are based on machine learning and can be viewed as a word classification task. We use feature extraction techniques like MI (mutual information) and  $X^2$  (Chi-square) to select the key features in the contexts of the terms of interest. Methods were evaluated extensively with a large number of experiments.



**Hisham Al-Mubaid**

## **Publications**

Al-Mubaid, H. and Nguyen H. A., "New Ontology-Based Semantic Similarity Measure for the Biomedical Domain," IEEE Conf. on Granular Computing GrC-2006. Atlanta, GA, May 10–12, 2006.

Al-Mubaid H., "Context-Based Technique for Biomedical Term Classification," Proc. 2006 IEEE Congress on Evolutionary Computation CEC-2006, Vancouver, BC, Canada, Abdelhak Bensaoula, David Starikov, Chris Boney, July 16–21, 2006.

Al-Mubaid, H. and N. Ghaffari, "A New Gene Selection Technique Using Feature Selection Methodology Gene Selection," Proc. CATA-2006, 2006.

Al-Mubaid H. and H. A. Nguyen, "Similarity Computation Using Multiple UMLS Ontologies in a Unified Framework," 22nd ACM Symposium on Applied Computing SAC'07, 2007.

## **Presentation**

Al-Mubaid, H. and H. A. Nguyen, "New Ontology-based Semantic Similarity Measure for the Biomedical Domain," IEEE Conference on Granular Computing, GrC-2006. Atlanta, GA, May 10-12, 2006. 623-28.

## **Funding and Proposals**

Al-Mubaid, H., "Supervised and Adaptive Learning To Improve text Entry for People with Physical Disabilities," Proposal submitted to NSF IIS, December 2006. (*Under review.*)



**SPIRIT OF THE WORLD**—At the entrance to the UHCL Bayou Building, students see Spanish sculptor Pablo Serrano's *Spiritus Mundi*.