



Department of Computer Science Texas Southern University



MONTHLY RESEARCH SEMINAR

Gene Protein Disease Relationships with the Gene Ontology

by

Hisham Al-Mubaid, PhD

Department of Computer Information Systems, University of Houston, Clear Lake (Texas, USA)

September 24, 2013,

3-4 PM, Nabrit Science 106

Biography

Dr. Al-Mubaid received his Ph.D. degree in Computer Science from University of Texas at Dallas with distinguished dissertation award in 2000. He is currently an Associate professor of Computer Science and Computer Information Systems and Program Chair of Computer Information Systems at the University of Houston, Clear Lake (Texas, USA). Dr. Al-Mubaid's main research interests are centered around natural language processing and bioinformatics and include data mining, machine learning, and biomedical text mining. Dr. Al-Mubaid is working to prove the virtue of the context-based approach to natural language processing (NLP). He developed several algorithms and systems to prove the point including: detecting and correcting context-based errors, word prediction, word classification, text categorization, biomedical term disambiguation and document clustering. He devised a (new) way of learning from 'good' versus 'bad' features in context-based NLP. He is currently writing a book on "*Natural Language Processing: Context-based Approach*". Moreover, Dr. Al-Mubaid, with his students, developed an efficient method for measuring semantic similarity of biomedical terms using multiple ontologies within the UMLS framework. That similarity measure has been developed in CPAN and available freely through: [UMLS-Similarity-0.17 > UMLS::Similarity::nam:http://search.cpan.org/~btmcinnes/UMLS-Similarity0.17/lib/UMLS/Similarity/nam.pm](http://search.cpan.org/~btmcinnes/UMLS-Similarity0.17/lib/UMLS/Similarity/nam.pm).

Besides, he has research projects and publications in learning and educational based research (e.g. *Self-regulated learning*). He serves as program chair and committee member of many regional and international conferences. He also serves in the editorial, technical board and reviewer for several journals. He is in the board of directors of ISCA, a member in ACM, IEEE, IEEE computer society, ACL, and other professional organizations. His teaching experience includes large number of undergraduate and graduate computer science courses.

Abstract

The amount of molecular biology and medical data is humongous and increasing at a very high rate. There is a huge need for methods and techniques for analysis and knowledge extraction from such huge data repositories. Findings and discoveries in this domain are very beneficial to human genomic structure and for understanding of functional associations between diseases and genes at the genomic and proteomic levels. This presentation will demonstrate some methods and techniques for exploring and modeling the relationships among genes, proteins, and diseases using the Gene Ontology.