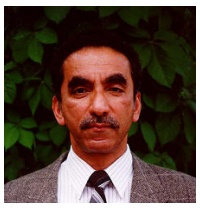
# Welcome

## Seminar Series of the Department of Computer Science Electrical Engineering

## Presents

### *Text Mining Biological Literature for Knowledge Discovery about Genes*

**October 19, 2007. Time 2:00-3:00PM. FH 557.**



**Prof. Sham Navathe**
**College of Computing, Georgia Tech, Atlanta, GA**

**Abstract**

**DNA microarrays can screen thousands of genes in a single experiment identifying altered expression levels for hundreds of genes. Many of these altered genes are often outside the field of expertise of the investigator. Interpretation of such large quantities of information by non-experts using traditional literature research (reading biomedical journals and searching related databases one gene at a time) is slow and inaccurate. This inefficiency hampers the understanding and discovery of the subtle functional relationships between the genes of interest. This talk will present our work that helps in interpretation of the results of microarray experiments. Keywords are identified using a statistical algorithm and are extracted from MEDLINE citations containing specific gene names. This talk will outline the various investigations we have done related to the extraction of meaningful keywords associated with genes and the clustering of these genes based on similarity of function from the entire Medline database We have developed our own clustering algorithm called BEA_PARTITION[1] which has been shown to outperform the common algorithms like K-means, Hierarchical clustering etc. Tests have been conducted with various datasets from microarray experiments from neurological and cardiovascular diseases. We showed the potential for discovering new functional information about genes that is hitherto not represented in public databases. We have further applied the extracted keywords as features to help in classifying literature related to epidemiology at CDC. Current research is directed toward enriching our approach with biomedical ontologies. This work is in collaboration with Ying Liu, UT-Dallas, and Profs. Dingledine (Pharmacology) and Ciliax (Neurology) of Emory University.**

**Bio: Shamkant Navathe is a Professor at the College of Computing, Georgia Tech. He has over 150 refereed publications. He is well-known for his work on database modeling, database conversion, database design, distributed database allocation, and database integration. His current research include human genome data management, intelligent information retrieval, text mining, mobile database applications, engineering and design applications, security risk modeling in information systems, and ontology and web services integration. He was the General Co-chairman of the 1996 International VLDB conference, Bombay, India, program co-chair of SIGMOD 1985, and General Co-chair of the IFIP WG 2.6 Data Semantics Workshop 1995. He has been an associate editor of ACM Computing Surveys, and IEEE Transactions on Knowledge and Data Engineering and many other journals. He is a co-author of the leading database text-book worldwide, Fundamentals of Database Systems, with R. Elmasri. He also co-authored the book "Conceptual Design: An Entity Relationship Approach" (Addison Wesley, 1992) with Carlo Batini and Stefano Ceri.**