

Susan Davidson: Conquering the Tower of Babel in Bioinformatics... and other tales of scientific databases

Over the past decade, biological research has been transformed from a science of the "small" to a science of the "large". Fueled by novel technologies capable of producing massive amounts of data for a single experiment, scientists are faced with an explosion of information which must be rapidly analyzed and combined with other data to form hypotheses and create knowledge. Thus a number of new research challenges have arisen in data modeling and data integration that must be solved to further biological as well as other scientific research.

Scientific data is very different from business data, for which current database technology has been developed. Michael Stonebraker recently argued that the traditional database concept of "one size fits all" is no longer applicable in the database market. Nowhere is this more true than with scientific data, much of which is tree structured either because its type is complex or because it models an inherently hierarchical process or object. In this talk, I will give examples of two different scientific database applications -- linguistic analyses and phylogenetic trees -- and discuss the particular issues that must be overcome in each. I will then discuss the problem of integrating across genomic data sets, and highlight the SHARQ project currently being undertaken by the database group at UPenn.

November 17, 2005, 3 PM Wu & Chen