FINAL REPORT OF FACULTY RESEARCH GRANT 2006

"VISUAL DATA REPRESENTATION AND INFORMATION EFFICIENCY IN KNOWLEDGE DISCOVERY"

By
Dr. Roger Fang
Computer and Information Science

June 2007

I. Restatement of Problem Researched

While new information technologies make data warehousing and data mining functions that are crucial to business operations more available, we also find that an important issue regarding the effectiveness of user access of such applications is mostly left either unsolved or overlooked. As an extension to a prior FRG-funded research "Knowledge Discovery in a University Data Warehouse," this project is intended to investigate and resolve issues related to the information efficiency of data warehousing systems.

II. Brief Review of the Research Procedure Utilized

The purpose of this project is to extend our previous work funded by FRG in fall 2004 to study and develop an effective multi-dimensional data access interface with certain visual representation technologies and, therefore, to improve the information efficiency of a data warehouse system.

To completely build a data warehouse to house all data extracted from SIS, we began to implement a multi-dimensional database using Microsoft SQL Server 2005 in spring and summer 2006. We have explored several user interface methods of information access of the data warehouse, including Microsoft Analysis Services, Microsoft Business Intelligence, Microsoft Visual Studio 2005, and Excel 2003. Using these tools users can find information like "Display the total number of morning classes, afternoon classes, and evening classes by day, department, and semester" or "Find the total number of enrollments and total number of withdraws by instructor, course, and semester" or "What is the most likely major a student will change to if he is currently majored in Computer Science with a GPA between 3.0 and 3.25?"

III. Summary of Findings

Depends on the tools, users can write OLAP queries directly in multi-dimensional extension (MDX) language, or via a graphical user interface to "drag-and-drop" dimensional data attributes to form queries. The final information output can be represented in a two to three dimensional pivot table format and/or a pivot chart. The

tools selected by the project also allow users to drill-down, roll-up or slice a multi-dimensional cube to browse the derived data in any appropriate detail level.

IV. Conclusions and Recommendations

Though visualization and data mining has been around and accepted in business and industry for a while, it's still very challenging to present data and information in multi-dimensional format. In this project we have found that three-dimensional and graphical visualization is still the most acceptable format in today's technology to solve business questions. We believe more research work in multimedia communications and visual data representation are required to achieve better and effective data representation to improve information efficiency in data warehousing and data mining applications.