

FINAL REPORT OF FACULTY RESEARCH GRANT 2004

**“KNOWLEDGE DISCOVERY IN A UNIVERSITY
DATA WAREHOUSE”**

By
Dr. Roger Fang
Computer and Information Science

May 2007

I. Restatement of Problem Researched

In academic field school and department administrators used to face and answer questions like “Which segments of students should we target to improve retention?” or “What are the common characteristics of students who will most likely drop out of our programs?” Unfortunately, these questions cannot be effectively answered and decisions can hardly be made based on general statistical analysis of gigabytes of data stored in a university database. There are simply too much data and too little information for the purpose of decision making. The goal of this project is to establish an exemplary university data warehouse to serve as the foundation and for ongoing research in academic knowledge discovering and decision support areas.

II. Brief Review of the Research Procedure Utilized

In fall 2004 and spring 2005, several existing data warehousing and data mining methodology, techniques and tools were investigated. A graduate course “INFT6903 Emerging Trends in Information Technology: Data Warehousing and Mining” was designed and first taught in spring 2005 in parallel to the research work. The experimental source of data from Student Information System (SIS) of our university was examined and identified during the same semester. In the mean time and through the help of Dr. Morell, Head of the Computer and Information Science, about three gigabytes of data were extracted from SIS and stored in text files. This huge data set includes students’ transcripts and class schedules of twenty years between 1985 and 2004.

To completely protect the extracted data a graduate assistant was hired in summer 2005 to implement programs to encrypt names and social security numbers of all students. Another graduate assistant was hired to cleanse, transform and load all raw data into a staging area of a database, which consists of a few simple structured data tables. During that period of summer time a suite of software tool from Microsoft (i.e., Microsoft SQL Server 2000 and Analysis Services) was acquired and installed. Time was spent in using these brand new software tools in both fall 2005 and spring 2006 while other effort was made in reorganizing the data and designing the schema of the targeted data warehouse. Based on the work up to fall 2005, a paper “Teaching Data Warehousing and Data Mining in a Graduate Program of Information Technology” was submitted and presented

in Consortium for Computing Sciences in Colleges 2006 Mid-South Conference at Christian Brothers University, Memphis, Tennessee.

III. Summary of Findings

The most successful part of this project is the data preparation, which serves as the foundation of more research work of knowledge discovery. The data that have collected contains 20 years of (1) student details like ACT/SAT scores, high school GPA and graduation date, date of birth, college admitted term, admitted school, admitted major, major changes, college graduation date and GPA, etc.; (2) transcript details of every student, such as the semester, course name, section number, credit hours, and grade of each course taken; (3) transcript summary, including student status of each semester, current/cumulative AHRS, current/cumulative EHRS, current/cumulative QHRS, current/cumulative QPTS, and current/cumulative GPA of each semester that the student has registered; and (4) class schedules of every semester between 1985 and 2004, which include semester, course number/name, class day/time, seats, capacity and enrollment, instructor, etc.

IV. Conclusions and Recommendations

With data being prepared and software tools being installed and studied, this project is going to continue on implementing a full scale of data warehouse with online analytic process (OLAP) querying and data mining capabilities. Once the hidden valuable information and knowledge become available, they can be deployed and utilized to support strategic decision makers in student retention, course planning, academic advising, housing development, and even faculty recruitment for the university.