original.

<u>REQUIRED COVER PAGE</u>

APPLICATION FOR FACULTY RESEARCH GRANT

**All questions must be completed to be considered for grant award.

Choose one: [ ] Creative [x] Research	**Date of Last FRG Award** (Semester and Year awarded): _____
	Date of ATU Faculty Appointment (Semester and Year): __Fall 200**&__

1. **Project Title:** ___Knowledge Discovery in a University Data Warehouse___

2. **Name of Principal Investigator/Project Director:** ___Dr. Roger Fang___

3. **School (abbrev):** ___Sys Sci___   4. **Department:** ___Computer and Information Science___

5. **Campus Mail Address:** __Corley 261__   6. **PI/PD Campus Phone:** __498-6082__

7. **Amount Requested: $** __1,800__   8. **Total Cost of Project: $** __1,800__

9. **Does this project involve:**   10. **Duration of Project:** __11/01/04 – 8/01/05__

Yes No
[ ] [X] human subjects?
[ ] [X] animals/animal care facility?
[ ] [X] radioactive materials?
[ ] [X] hazardous materials?
[ ] [X] biological agents or toxins restricted by the USA Patriot Act?
[ ] [X] copyright or patent potential?
[ ] [X] utilization of space **not** currently available to the PI/PD?
[ ] [X] the purchase of equipment/instrumentation/software currently **available** to the PI/PD?

NOTE: If the answer is "yes" to any of the above questions, the investigator must attach appropriate documentation of approval or justification for use/purchase.

SIGNATURES

Department Contribution (if applicable): **$** _____

Account Number: _____   *Jany Morell*   10/12/04
                                             Chairperson   Date

School Contribution (if applicable): **$** _____

Account Number: _____   *John White*   10-12-04
                                             Dean   Date

This Section to be completed by the Office of Academic Affairs
FSBA Committee Award Recommendation: Yes____   No____ FSBA Committee Proposal Rank: _____ of _____ Total Proposals. Recommendation of VPAA: Yes____ No _____ Recommendation of President: Yes____ No _____ Award Date: _____

ABSTRACT

Making mission critical decision is a key factor of corporate success: What are the attributes that influence the percentage of customers returning for more business? Which segments of customers should we target to improve retention? Should additional sections of course X be opened when we know "sixty-eight percent of the time that a student enrolls Y which is a prerequisite course of X, the students also enrolls X in the following semester."? Such similar decisions are fundamental to any organization, including nonprofit corporations like schools, museums, hospitals, militaries, or government agencies at all levels. However, quality decisions depend not only on the decision makers' own experience and knowledge, but mostly on the data and information that are collected and used. Traditionally decision making process relies on intensive statistical approaches, but as a trend and proved best practice, information technology such as data warehousing and data mining has been applied in recent years [1] to create more effective data analysis and more accurate predictive models than traditional approaches. These new technologies emphasize automatic *knowledge discovery* in databases to improve decision quality. Successful business applications of data warehousing and mining include customer retention, fraud detection, cross-selling, and market-basket analysis in virtually every business sector, and gene or protein or new drug identification in life sciences [2], to name a few.

While practitioners expect to see the continued development of applications based on these technologies, more research is needed in this area to develop the underlying mining techniques that are more automated, scalable, and reliable [3]. The goal of this proposed project is to establish an exemplary data warehouse system to serve as the foundation and test bed for ongoing research in knowledge discovering and decision supporting areas.

1

PURPOSE/OBJECTIVES

Through this project, a university data warehouse of students and academic operation records will be established first. Various data mining and knowledge discovering techniques and algorithms such as *k-nearest neighbor*, *neural networks*, and *genetic algorithms* [4] will be compared and selectively applied to create analytical models, which shall emphasize our school information needs of decision making. For example, to predict and minimize the retention risks, or, to discover the success factors and patterns of a given student major to increase the graduation rate.

SIGNIFICANCE/NEED

While the quantity of data in the world roughly doubles every year, more and more organizations are confronted with the challenge of too much data and too little information for decision making. Consequently, the ability of accessing the desired information from an ever-increasing amount of data has become one of the critical success factors for nearly everyone, including a university, to survive as a professional organization in today's competing world. In school, we can easily hear questions like "Should I better major in computer science or mathematics?" or "What are the major factors that predict whether or not the freshman retention rate will keep raising in the next three years?" or "Will five sections of this course satisfy the Fall enrollments?" or "What are the common characteristics of students who most likely will drop out of our programs?"…, etc. Unfortunately, these questions cannot be effectively answered and decisions can hardly be made based on general statistical analysis of gigabytes or terabytes of data stored in a school database. There is more knowledge that is buried deep within the data. Building a data warehouse from several operational databases and using data mining techniques is the most effective and efficient way for an organization to discover hidden knowledge to improve its decision making ability.

2

Evidenced by many successful business cases, we believe the result of this project will strengthen a school's efforts in areas such as student retention, course planning and advising, faculty recruitment, housing development, and so on.

The grant of this project will hire one graduate assistant, pay for publications and additional data from other providers, and software tools that are needed in research and development process of the project.

PROCESS FOR ATTAINMENT OF OBJECTIVES/GOALS

This project will be performed in two stages: research then prototyping. During the research stage, the following tasks will be performed:

1. Existing data warehousing methodology such as [5], [6] and data mining techniques [7], [8] will be investigated.

2. Information needs will be identified and defined.

3. Operational data in our Student Information System (SIS) will be examined and selected. Extra census data and students demographic information from other external sources may be purchased.

4. Software tools (e.g., Microsoft Analysis Services, Oracle Data Mining (ODM), IBM DB2 Intelligent Miner, etc.) and storage equipments will be acquired.

5. An experimental data warehouse with minimal analytical capability will be completed.

The project will continue to finish the following items in the prototyping stage:

6. A full scale data warehouse will be built.

7. Data mining models and algorithms will be determined and implemented.

8. The system's analytical functions, which include uncovering trends, patterns, and relationships from the accumulated data in the data warehouse, will be demonstrated.

3

9. A friendly user interface will be added to present the output information in easy understanding forms and reports.

The proposed timeline for completing the above project activities is scheduled as shown below but will be adjusted, as needed, to correspond to the start of funding:
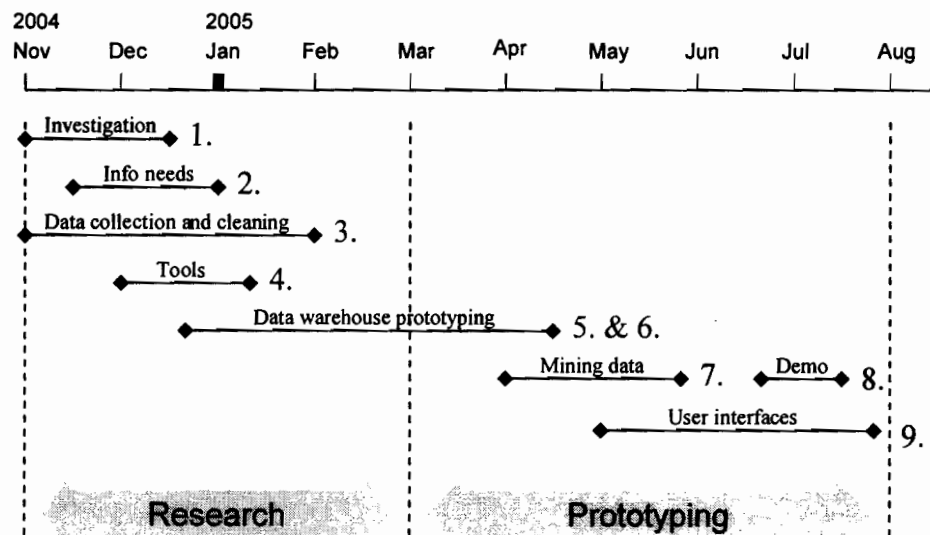
DISSEMINATION OF RESULTS

During the second phase of the project, a prototype of data warehouse system will be presented on campus, followed by some public presentations in conferences such as Annual Mid-South College Computing Conference in 2006, or ALAR Conference on Applied Research in Information Technology, 2006, or other information systems and data warehousing related conferences in mid-2005 or 2006. The completed system can also be deployed on campus to support decision making of users at various offices. In addition, the research and implementation experiences of the project can not only enrich the curriculum of the department by offering a couple of new courses, but also upgrade the department rank to an advanced position in this area.

BUDGET

(See the attached sheet on next page.)

BIBLIOGRAPHY

[1] Westervelt, R. Data Warehousing Gets Active. SearchDatabase.com News, 28 Aug 2003.
http://searchdatabase.techtarget.com/qna/0,289202,sid13_gci921760,00.html

[2] Oracle, Oracle Data Mining Industry Solutions for Database Marketing, Financial Services, and Government.
http://www.oracle.com/technology/products/bi/pdf/odm_ind_sol_dbmktg.pdf
http://www.oracle.com/technology/products/bi/pdf/odm_ind_sol_fin.pdf
http://www.oracle.com/technology/products/bi/pdf/odm_ind_sol_gov.pdf

[3] Apte, C., Liu, B., Pednault, E., Smyth, P. Business Applications of Data Mining. Communication of the ACM, Vol. 45, No. 8, 2002, 49-53.

[4] Adriaans, P., Zantinge, D. Data Mining. Addison-Wesley, 1998, 56-77.

[5] Anahory, S., Murray, D. Data Warehousing in the Real World. Addison-Wesley, 1997.

[6] Sturm, J. Data Warehousing with Microsoft SQL Server 7.0. Microsoft Press, 1999.

[7] Seidman, C. Data Mining with Microsoft SQL Server 2000. Microsoft Press, 2001.

[8] Marakas, G. M. Modern Data Warehousing, Mining, and Visualization. Prentice Hall, 2002.
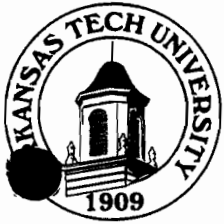
APPLICATION VITA

PI/PD of this project has taught courses in databases and information systems field for several years since 1999 in British Columbia Institute of Technology (BCIT, Canada), Kingston College (Canada), and Arkansas Tech University. Before that, the author of this proposal worked in a Taiwan governmental research lab where he led a project to develop an Informix database to manage millions of customer records in a billing system. No Faculty Research Grants have been applied before by the author in conducting research in this area at Arkansas Tech University.

5

PROPOSED BUDGET
FACULTY RESEARCH GRANT

1. Graduate assistant stipend $ 800
   Fringe benefits @ .4% (4/10 percent) of graduate assistant stipend   4

2. Non-work study stipend $ 0
   Fringe benefits @ .4% (4/10 percent) of non-work study stipend

3. *Supplies (please list items to be purchased and estimated price
   per item including taxes and shipping, if appropriate):

Item No. 1 publication purchase	Estimated Price	$	200
Item No. 2 data purchase	Estimated Price	$	300
Item No. 3 computer software	Estimated Price	$	300
	Total estimated supplies	$	800

4. Travel (please list travel expenditures by date
   and estimated costs):

Travel No. 1 May-June 2005	Estimated Price	$	200
	Total estimated travel	$	200

5. *Capital Outlay (please list items to be purchased and estimated
   price per item including taxes and shipping, if appropriate):

   Total estimated capital outlay $ 0

   TOTAL PROPOSED BUDGET $ ~~1,800~~
       1804

*Items purchased under $2,500 (including taxes and shipping) are considered supply items. Capital Outlay items are those which cost $2,500 or more (including taxes and shipping). Please contact the Purchasing Office for questionable items.

Administration Building Room 200
1509 North Boulder Avenue
Russellville, AR 72801-2222

phone: 479-968-0319
fax: 479-968-0644

www.atu.edu

November 5, 2004

Dr. Roger Fang
Assistant Professor of Computer & Information Science
Arkansas Tech University
Corley Building Room 263
Russellville, AR 72801

Dear Dr. Fang:

Congratulations! Academic Affairs is pleased to announce your application for the Spring 2005 Faculty Research Grant has been recommended by the Faculty Salary Benefits and Awards Committee. Based on this recommendation, Academic Affairs has approved the $1,800 budget for your research of Knowledge Discovery in a University Data Warehouse. Requisitions regarding the grant will be processed through your Dean's office and should be expended by June 30, 2004.

Your research on this project is sure to not only benefit your department, but Arkansas Tech University as a whole. We wish you success with this endeavor.

Sincerely,

Jack Hamm
Vice President for Academic Affairs

Copy: Dr. Larry Morell
Dr. John Watson
File