IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4710

Dimensional Modeling: Initial Approach for Identifying and Classifying Patterns

Mary Elizabeth Jones
Drexel University
College of Information Science & Technology

RESEARCH PROBLEM

It is recognized that software design is a difficult and time-consuming process that requires skill and experience. Software engineers continually seek methods for improving the efficiency and effectiveness of the software design process. For example, software engineers have used structured and object-oriented design techniques to improve the software design process. These techniques were developed to enhance the software design process by using a consistent standard approach. Yet, concentration on software design techniques alone has not guaranteed successful software design solutions [Brooks, 1995]. Ward Cunningham concurs:

> *A growing number of us in the object-oriented community feel we have misplaced our collective attention for some time. We no longer need to focus on tools, techniques, notations, or even code. We already have in our hand the machinery to build great programs. When we fail, we fail because we lack experience. [Fowler, 1997]*

The dilemma remains – how do software engineers obtain experience in order to create effective solutions?

As software engineers implemented more and more systems and as the concept of a software pattern was recognized, software engineers became more aware of the similarities within systems. The software engineering community realized they could benefit from capturing and understanding software design similarities within and across various systems. Those similarities formed the basis for software patterns that can be reapplied when designing future systems. With each subsequent system, patterns are used, enhanced, and adapted to improve their usefulness for future systems. Thus patterns capture and combine successful design techniques as well as the experience of software engineers.

In the area of data warehousing, practitioners are trying to help software engineers obtain the experience necessary to create effective solutions. Several books provide data warehouse design solutions that are organized by domain [Kimball, 2002], [Kimball, 1998], [Adamson and Venerable, 1998], [Silverston, 2001]. The purpose of these books is to demonstrate data warehouse design techniques by example. These books are intended to assist the software engineer in understanding data warehouse design by studying and learning from the examples of those more experienced in data warehouse design and implementation. But the examples represent specific approaches to specific situations and therefore the examples may not address the software engineer's particular design problem. The data warehousing community has not used these existing design solutions to identify and classify patterns. Currently, the data warehousing community lacks patterns - a formal organized representation of recurring strategies worthy of following, applying, and repeating when designing data warehousing applications.

RESEARCH OBJECTIVES

The published design solutions are valuable. However, they are usually example solutions to a specific data warehousing problem. The next step is to analyze those solutions with the intent of seeking and extracting design techniques that commonly occur within and across application domain areas. Those repeating occurrences – which will be referred to and documented as data warehouse design patterns - can be reused on future design efforts.

The objective of this research project is to identify data warehousing design patterns, classify those patterns in a way that communicates their usefulness in a usable way, and determine if those patterns will help in designing a data warehouse. This research project will evaluate if the identified patterns help the data warehousing design process by testing the following null hypotheses:

- H0: Data warehousing patterns have no impact on the time to design a data warehouse schema.
- H0: Data warehousing patterns have no impact on the correctness of data warehouse schema design.

LITERATURE OUTLINE

The literature review

- Presents the historical and evolutionary perspectives of databases and software design patterns
- Defines dimensional modeling and software design patterns
- Presents the lessons learned from software design pattern experts in terms of the benefits and difficulties with software design patterns
- Provides an overview of the varieties of existing software design patterns

METHODOLOGY

The objectives of this research project are to identify and classify data warehousing schema patterns and then evaluate their impact on the design process. A classical pretest post-test experiment will be used to evaluate the impact of patterns on the software design process.

Two attributes will be used to evaluate the impact of patterns on data warehousing design solutions. They are the: 1) time it takes to create a solution and 2) correctness of the solution. Once the data warehousing patterns are identified and classified, a pretest post-test experiment will be conducted to evaluate the time it takes to solve a data warehouse design problem and the correctness of the data warehouse design solution. Therefore, the independent variable will be defined as "exposure to patterns" and the dependent variables will be: 1) time and 2) correctness. In other words, this pretest post-test experiment will be evaluating how exposure to patterns impacts the time it takes to solve a data warehouse design problem and the correctness of the solution.

For this experiment, the experimental units are graduate students. They will be selected from the population of graduate students enrolled in Drexel University's College of Information Science and Technology (IST). IST graduate students can pursue a Master of Science in Library and Information Science, Master of Science in Information Systems, a dual Master's of Science degree in Library and Information Science and Information Systems, Master of Science in Software Engineering, and a Doctor of Philosophy in Information Science and Technology.

The graduate students for this experiment must have an understanding of database concepts and system analysis as well as a rudimentary understanding of data warehousing design. Graduate students enrolled in the Applied Information and Database Technology course (INFO 607) are taught dimensional modeling and design techniques, which is used for the design of data warehousing databases. The prerequisites for this course are Database Man-

agement II (INFO 606) and Information System Analysis (INFO 620). Thereby making the graduate students taking INFO 607 well suited to participate in this experiment. Therefore, IST graduate students enrolled in INFO 607 will provide the data for this research project. This course typically has 25 to 30 students per class.

This pretest post-test experiment will use two groups – 1) a control group that will receive no exposure to the data warehousing patterns and 2) an experimental group that will receive exposure to the data warehousing patterns. The control group will be taught dimensional modeling and specific design examples will be reviewed and discussed. The experimental group will be taught dimensional modeling and will also be taught data warehousing design patterns. The subjects will be randomly assigned to either the control group or the experimental group using the simple random sampling technique. Students will be serially numbered. A random number table will be used to assign the students to either the control group or the experimental group.

EXPECTED OUTCOME

It is expected that this research project will add to the body of data warehouse design and software engineering knowledge by:

- Identifying and classifying data warehousing schema patterns
- Evaluating the efficiency and efficacy of data warehousing design schema patterns when used by graduate students of data warehousing

In terms of evaluating the efficiency and efficacy of data warehousing design schema patterns it is expected that the time to design a data warehouse schema is less for software engineers that have learned and applied data warehousing design patterns than those who have not learned and applied data warehousing design patterns. Also, it is expected that the correctness of the data warehouse schema design is more correct for those who have learned and applied data warehousing design patterns than those who have not learned and applied data warehousing design patterns.

BIBLIOGRAPHY

Adamson, Christopher and Venerable, Michael. Data Warehouse Design Solutions. New York: John Wiley & Sons, Inc., 1998.

Brooks, Frederick P. The Mythical Man-Month. Reading: Addison Wesley Longman, Inc., 1995.

Fowler, Martin. Analysis Patterns: Reusable Object Models. Menlo Park: Addison-Wesley, 1997.

Kimball, Ralph; Reeves, Laura; Ross, Margy; and Thornthwaite, Warren. The Data Warehouse Lifecycle Toolkit. New York: John Wiley & Sons, Inc., 1998.

Kimball, Ralph; and Ross, Margy. The Data Warehouse Toolkit, 2nd ed. New York: John Wiley & Sons, Inc., 2002.

Silverston, Len. The Data Model Resource Book, Vol. 1, revised ed. New York: John Wiley & Sons, Inc., 2001.