AN INFORMATION SYSTEM PROFESSIONAL'S PERSPECTIVE ON DATABASES

S. Nataraj, Texas A&M University-Corpus Christi, snataraj@mail.tamucc.edu
A.K. Mukherjee, University of W. Florida, amukherj@uwf.edu
R.J. Meier, Fort Hays State University, rmeier@fhsu.edu

ABSTRACT

This paper presents the results of a survey conducted with the objective of finding out the relative importance of various database techniques and topics as perceived by practitioners. The survey provided practitioner feedback in differentiating topics that students need to be competent in from those that they need to know or merely need to be exposed to. For example, the survey found that over 57% of practitioners believe 'normalization of databases' to be a topic that students should become competent in. 72% of respondents believe that students should be competent in SQL Queries. In contrast, only 10% of respondents believe that students should become competent in CASE tools. Most respondents believe that students need only be exposed to the topic of 'Data Dictionary'. The results of the survey are likely to provide guidance to database instructors about the degree of effort to be allocated to different topics/ techniques in the course. The results have been used to guide curriculum revision in the database course for computer information systems major at a regional university.

Keywords: Database management, database design, databases, information systems

INTRODUCTION

One of the most frequent activities of a Systems Analyst in industry is to design, develop and implement databases. Accordingly, AITP (Association of Information Technology Professionals) and other computing associations require Computer Information Systems (CIS) and Management Information Systems (MIS) graduates to take a course in database theory and practice. In general, the course instructor chooses a popular text and covers topics that he/she considers important. However, in discussions with practitioners it is usually found that most students do not have adequate background in topics that are considered useful to industrial practice.

If a student receives training only in database theory, he/she would need a great deal of learning time to adapt to current needs of the industry. On the other hand, if a student is merely trained to meet current industrial needs, they would lack the theoretical knowledge needed to adapt to the dynamic world of information systems. Clearly, a compromise is important.

Thus a more pragmatic and balanced approach would be to expose students to important ideas and theories while making him/her competent in a few techniques that are of current importance to the practitioner.

This research surveyed information systems (IS) professionals regarding the importance of topics/ techniques the students need to be taught in the area of database design and implementation. The survey instrument had many topics and techniques. The results of the first part of the survey have been presented in Nataraj et al. (1997). The topics covered in the first part of the survey and analyzed in the paper by Nataraj et al. (1997) are listed in Appendix 1.

This paper presents the results of the second part of the survey. First a description of the school and requirements of the CIS program is presented. Then the methodology of the survey is described. Next an analysis of respondents is presented. This is followed by an analysis of results. Finally the discussion of findings summarizes important guidelines for faculty.

THE ENVIRONMENT AND METHODOLOGY

Fort Hays State University, located in western Kansas, offers a major in Computer Information Systems in the undergraduate curriculum within the College of Business. As part of the technical preparation, CIS majors are required to take one database systems course. Others required courses include COBOL, Information Systems Design and Development, MIS and Systems Development Project.

The first step was to identify a list of topics/ techniques to be included in the survey. A comprehensive list of database topics was identified from important textbooks. This included the texts by Awad (1992), Gorman (1991), Pratt (1994), Kroenke (1995), Flores (1981), Elmasri et. Al. (1989), Frost (1984), Chorafas (1989), Cardenas (1979), Fernandez (1981) and Everest (1986). The final list was an abridged version of the above and was prepared in consultation with faculty members in the CIS department.

The second step was to identify suitable respondents. The Department of Computer and Information Systems has an advisory board with members from business and industry including aerospace, oil/gas extraction, finance, telecommunication, insurance and government. The computer advisory board members were part of the target population to answer the survey. The others surveyed were information systems professionals in a large metropolitan city in the state. The respondents were asked to indicate job title, industry type, and size of information systems department. The survey instrument listed a variety of database topics and techniques and asked the respondents to identify the importance category of the listed topic/technique. There were five different importance categories to choose from. The first importance category was "Not important.' The second importance category was 'Important to be exposed to.' It was clarified on the instrument that for such topics, exposure might be achieved through lectures or self-study. However, the learning will not be reinforced through homework/ examinations. The third importance category was 'Important to know.' The survey instrument defined this to be a topic in which a student is expected to become knowledgeable through lectures or self-study. The knowledge is to be reinforced through examinations (quizzes & tests) on this material. However, the students will not perform homework on these. The fourth importance category was 'Important to be competent in.' The survey instrument defined this to be a topic in which a student is expected to become competent. It was clarified that this will involve two steps. First, they will be exposed to the material through lectures/self study. Then, the learning of this material will be reinforced through homework assignments and examinations. The fifth category was that of 'No Comment.' This was to be used if the respondent was unable to comment on the importance of a particular topic/technique.

Different industries use different terms when referring to particular topics. Sometimes two different organizations use the same term in different ways. In order to make sure that respondents understood the meaning of the topic in the manner intended by the authors, a glossary was included with the survey instrument. This glossary provided a detailed definition of the term/topic/technique. A few examples of details provided in the glossary are included in Appendix 2.

ANALYSIS OF RESPONDENTS

Replies were received from fifty-one respondents. These respondents had such job titles as 'programmer' (8%), 'analyst' (14%), 'end user' (12%), 'manager' (25%), 'other' (25%) and 'no response' (16%). A broad range of industries was represented. This included 'manufacturing,' 'utilities,' 'oil and gas extraction,' 'finance,' 'services,' and 'government.'

ANALYSIS OF RESULTS

In this article results are reported about five major topics covered in the second part of the survey. These include the major topics of 'Data Dictionary,' 'Normalization,' 'SQL and Data Manipulation,' 'Distributed DBMS,' and 'Other Topics.' The results of the survey are discussed in terms of sub-topics that make up these five major topics.

Data Dictionary

In this part of the instrument there were 6 sub-topics. The results have been grouped according to the sub-topic.

The first sub-topic was concerned with 'Function of the Data Dictionary.' Around 43% of the respondents considered it important to know . The second sub-topic was concerned with 'Components of Data Dictionary.' The majority of respondents (47%) considered this to be a topic that students should be exposed to . The third sub-topic 'Data Dictionary-Users' was similarly found important enough only to be exposed to by 53% of the respondents . Around 51% and 45% of respondents considered 'Data Dictionary-Interfaces' and 'Data Dictionary-Structure' to be important to be exposed to, respectively. Only 12% of respondents felt that 'Evaluating Data Dictionary' was a topic that students should become competent in.

Normalization

A majority of respondents indicated that each sub-topic in this broad topic was important enough for students to become competent in. Around 57%, 45%, 43%, 50%, and 51% of respondents considered the sub-topics 'What is Normalization,' 'First Normal form,' 'Second Normal form,' 'Third Normal form,' and 'Storage Performance Considerations' respectively, to be important enough for students to become competent in.

SOL & Data Manipulation

The third major topic was concerned with SQL & Data Manipulation. This major topic had five sub-topics. Around 59% of respondents reported the sub-topic 'SQL – Data Definition' to be important enough for students to be competent in. This percentage increased to 61% for the sub-topic 'SQL – Data Control' and to 68% for the sub-topic 'SQL – Manipulation Commands'. An overwhelming 72% of respondents considered the sub-topic 'SQL – Queries' important enough for students to become competent in . Most respondents (63%) also believed that students should become competent in the sub-topic 'SQL – Joins'.

Distributed DEMS

This major topic had five sub-topics. In general, the four sub topics 'Distributed DEMS – Advantages,' 'Distributed DBMS – Disadvantages,' 'Distributed DEMS – Design Criteria' and "Distributed DBMS – Planning Considerations' were considered to be important for students to know. In contrast, most respondents (45%) considered the sub-topic 'Distributed DBMS – Security' important for students to become competent in.

Other Topics

This major topic had a collection of eight sub-topics. Most respondents (61%) considered 'Planning a database application' important for students to know. Around 18% of respondents considered the sub-topic 'Structured English' important enough for students to become competent in. In contrast, only 2@ of respondents found the sub-topic of 'Pseudocode' to be important enough for students to be competent in. The sub-topics of 'Code Generators' and "CASE tools' were, in general, reported to be important for students to know. The sub-topic of 'Database Administration' was largely considered to be important to know while the sub-topic of 'Choosing a DBMS' was generally considered important to be exposed to. Most respondents (41%) felt that the sub-topic of 'Commercial Database Systems' was important to be exposed to.

DISCUSSION OF FINDINGS

The original goal of this survey was to differentiate between topics that students needed to be competent in from those that were less important. This would enable instructors to prioritize topics/ techniques and better allocate resources. Topics that students need to be competent in would be allocated more effort than topics that students need to know. Similarly, topics that they need to know will get more effort than topics that they need to be exposed to.

The first step was to identify the most important topics. These are topics that students should become competent in. It was considered reasonable to select those topics for this list where the highest percentage of respondents had considered the topic to be 'Important to be competent in.' The results are presented in Table 1.

Eleven out of twenty nine sub-topics were considered crucial enough for students to become competent in. The most important topics for students to be competent in appears to be 'SQL – Queries' and 'SQL – Manipulation Commands.' Other topics in which students should become competent, in decreasing order of importance, are 'SQL – Joins,' 'SQL – Data Control,' 'SQL – Definition,' 'What is Normalization,' 'Storage Performance Considerations, 'Third Normal Form,' 'Distributed DBMS – Security,' 'First Normal Form,' and 'Second Normal Form.'

Table 1: Topics/ techniques that are important for students to be competent in.

No.	Topic/Technique	% of Respondents
1.	SQL – Queries	72%
2.	SQL – Manipulation Commands	68%
3.	SQL – Joins	63%
4.	SQL – Data Control	61%
5.	SQL – Data Definition	59%
6.	What is Normalization?	57%
7.	Storage Performance Considerations	51%
8.	Third Normal Form	50%
9.	Distributed DBMS – Security	45%
10.	First Normal Form	45%
11.	Second Normal Form	43%

The next step was to identify the second most important topics. These are topics that students need to know. Topics where the highest percentage of respondents had considered the topic to be 'Important to know' were identified for this list (Table 2).

There were 9 sub-topics out of 29 that were classified as being important for students to know. The list was headed by the two sub-topics 'Distributed DBMS – Planning considerations' and 'Planning a Database Application.' Other sub-topics that are important for students to know, in decreasing order of importance, are 'Distributed DBMS – Disadvantages,' 'CASE Tools,' 'Distributed DBMS – Advantages,' 'Code Generators,' 'Distributed DBMS – Design Criteria,' 'Structured English,' and 'Database Administration.'

Table 2: Topics/ techniques that are important for students to know.

No.	Topic/Technique	% of Respondents
1.	Distributed DBMS – Plg. Considerations	62%
2.	Planning a Database Application	61%
3.	Distributed DBMS – Disadvantages	53%
4.	CASE Tools	51%
5.	Distributed DBMS – Advantages	48%
6.	Code Generators	47%
7.	Distributed DBMS – Design Criteria	45%
8.	Structured English	43%
9.	Database Administration	40%

The final step was to identify those topics that are at the third level of importance. In the survey these were identified as topics that students need to be exposed to. Topics where the highest percentage of respondents had considered the topic to be 'Important to be exposed to' were included in this list (Table 3).

Table 3: Topics/ techniques that are important for students to be exposed to.

No.	Topic/Technique	% of Respondents
1.	Data Dictionary – Users	53%
2.	Pseudocode	53%
3.	Data Dictionary – Interfaces	51%
4.	Data Dictionary – Components	47%
5.	Evaluating Data Dictionary	47%
6.	Data Dictionary – Structure	45%
7.	Data Dictionary – Function	43%
8.	Commercial Database Systems	41%
9.	Choosing a DBMS	40%

Five sub-topics related to Data Dictionary were found to be in the third importance category of being a topic that students should be exposed to. Other topics that were only important enough for students to be exposed to includes 'Pseudocode,' 'Commercial Database Systems' and 'Choosing a DBMS.'

Many of the results have come as a big surprise to the authors. For example, contrary to the author's expectations, the sub-topics of 'CASE tools' and 'Planning a Database Application' did not show up in the list of topics that students need to be competent in. While sub-topics

related to Data Dictionary are currently emphasized in the database course, the results of the study suggest that these are topics that students only need to be exposed to. Thus less emphasis needs to be put on the sub-topics related to Data Dictionary. The authors of the study had always considered the sub-topic of 'Storage Performance Considerations' to be important enough for students to know and were considerably surprised to find the sub-topic in the list of topics that students should be competent in. The sub-topic of 'Distributed DBMS – Security' had always been considered by the authors to be important for students to be competent in. The authors had expected the sub-topics of 'Choosing a DBMS, and 'Commercial database systems' to be found important for students to know. The ranks of 8 and 9 for these two sub-topics in the list of techniques to be exposed to came as a big surprise to the authors. This was also surprising in view of the job advertisements, which appear to ask for experience in commercial database skills.

CONCLUSION

A survey was carried out to identify the importance of various database techniques/topics from the perspective of practitioners. The main purpose was to differentiate in a meaningful way between topics that students need to be competent in, topics that students need to know and those that they need to be exposed to.

The study shows that the topics covered in a traditional database course are familiar to practitioners who were chosen to respond to the survey. Topics pertaining to SQL-Queries, SQL-Joins, SQL Data Manipulation, First Normal form, and Second Normal form are some topics/ techniques that students need to be competent in. Planning a database application, CASE tools, and Structured English are some topics that students need to know. Data Dictionary, Pseudocode and Commercial Database Systems are some topics that students need only to be exposed to.

It is expected that the categorization presented in this article would provide instructors with some thumb rules regarding allocation of time and effort to the topics.

REFERENCES

1. Awad, E.M. and Gotterer, M.H. Database Management, Boyd & Fraser Publishing Company, Massachusetts, 1992.
2. Cardenas, A.F., Database Management Systems, Allyn & Bacon, Massachusetts, 1979.
3. Chorafas, D.N., Handbook of Database Management and Distributed Relational Databases. Tab Books Inc., Pennsylvania, 1989.
4. Gorman, M.M., Database Management Systems: Understanding and Applying Database Technology, QED Information Sciences, Massachusetts, 1991.
5. Elmasri, R. and Navathe, S.B., Fundamentals of Database Systems, Benjamin/ Cummins Publishing Co., New York, 1989.
6. Everest, G.D., Database Management Objectives, System Function Administration, McGraw Hill, New York, 1986.
7. Femandez, E.B., Summers, R.C. and Wood C., Database Security and Integrity, Addison-Wesley Publishing Co., Reading, Massachusetts, 1981.
8. Flores, Ivan, Database Architecture, Van Nostrand Reinhold Co., New York, NY, 1981.
9. Frost, R.A., Database Management Systems, McGraw Hill, New York, 1984.
10. Kroenke, D., Database Processing, Prentice Hall, 1995.

11. Nataraj, S., Hassett, M. & Mukherjee, A., "Database Techniques and Topics - What Practitioners Consider Important," International Business Schools Computing Quarterly, Vol. 9 No. 1, Spring 1997, pages 51-60.
12. Pratt, P.J., and Adamski J.J., Database Systems Management and Design, Boyd & Fraser Publishing Co., Massachusetts, 1994.

APPENDIX 1
Topics analyzed in paper by Nataraj et al. (1997)

1. Limitations of File Processing Systems
2. DBMS Characteristics
   - a) Data integration
   - b) Data shareability
   - c) Data independence
   - d) Data Hierarchy
3. DBMS Environments
   - a) Mainframe based database systems
   - b)Micro-computer based database systems
   - c) Network based database systems
   - d)Distributed database systems
   - e) Object oriented database systems
4. Database Systems Life Cycle
5. Primary DBMS Functions
   - a) Natural User Interface
   - b) Support of logical transactions
   - c) Data Integrity
   - d) Concurrent Processing Controls
   - e) Recovery from failure
   - f) Database security
6. Data Models
   - a) Entities and Attributes
   - b) Relationships in a hierarchy
   - c) Bubble charts
   - d) Entity Relationship diagrams
   - e) Schemas
   - f) Object oriented data modeling
7. Models for representing relationships
   - a) Hierarchical model
   - b) Network model
   - c) Relational model
   - d) Comparison of data models

APPENDIX 2
Sample Glossary

1. Data Dictionary: A tool that is used to store descriptions of the entities, attributes, relationships, programs, and so on, that is associated with an organization's database. A data dictionary may also include data that are external to the DBMS.
2. Normalization: A process that assigns attributes to entities in such a way that data redundancies are eliminated or reduced.
3. First Normal form: A relation is in the first normal form if it does not contain repeating groups.
4. Second Normal form: A relation is in the second normal form if it is the first normal form and no non-key attribute is dependent on only a portion of the primary key.