

Use of Basic Classes in Database Structure Design:

Research Concept

Don Drumtra

University of Texas at Austin

January 6, 2003

Version 1.1

Approved by the Examining Committee

February 27, 2004

Dr. Fran Miksa, Chair

Dr. Randolph Bias

Dr. Larry Browning

Dr. Glynn Harmon

Use of Basic Classes in Database Structure Design

Research Concept

Chapter 1- Introduction

The suggested research explores how business organization database designers determine the basic classes they use in their information and data structures, what concepts and theory they should have available to inform their decisions, what concepts and theory they actually use to make their decisions, and why there might be a gap between their knowledge and their practice. This research also explores basic classification philosophy¹ and theory that could potentially help designers' future decision-making process and how they might be convinced to apply that knowledge to improve data structure decision-making.

The Nature of the Problem

Decision-makers in business organizations need information in order to make decisions. Such information must be accurate, sufficient, and available at the right time if it is to be useful in decision-making. Information systems, particularly computer information systems and associated databases, have been a major tool decision-makers have used to provide the tools. In 1990, Efraim Turban² wrote,

¹ Philosophy in this regard is based on the Oxford English Dictionary definition, "7. With *of*: The study of the general principles *of* some particular branch of knowledge, experience, or activity" (*Oxford English Dictionary*, 1989, Philosophy, n).

² Efraim Turban was the Lumpkin Distinguished Professor of Business at Eastern Illinois University and an expert in management information and decision support systems.

Managers may access thousands of databases all over the country. Many organizations, private or public, are no longer making major decisions without some computerized analysis. . . . Corporations are developing distributed systems that enable easy accessibility to data stored in multiple locations. Various information systems are being integrated with each other and/or with other automated systems. Managers can make better decisions because they have more accurate information at their fingertips. (Turban, 1990, p. v)

The structuring of databases and their contents are important if they are to be useful tools for business decision-making. Rob and Coronel³ wrote, “A good database does not just happen, the structure of its contents must be designed carefully. . . . A well-designed database facilitates data management and becomes a valuable information generator” (Rob & Coronel, 2000, p. 7).

Even though designing the structure of a database is not a trivial task, several scholars have suggested that field database designers may not treat database design with the seriousness it deserves. Inmon⁴ wrote:

It is a sad fact of life that when a person learns the mechanics of data base design, he or she often feels equipped to design an application with no further study. Perhaps the architects of today’s data base management systems have done a disservice by making it too *easy* to construct data bases. The freedom the beginner has in constructing data bases usually results in a poor design, which in turn triggers all sorts of other problems. (Inmon, 1981, p. ix)

Nevertheless, as the examples which follow in the background section of this concept paper show, even experienced designers may push ahead with a design without adequate consideration of the nature of the entities being modeled. This is particularly true at the beginning of a project. The need to get the design product started and the need to show initial

³ Peter Rob was a database design, implementation, and management professor and researcher in Middle Tennessee State University and a prolific author of works dealing with database management. Carlos Coronel was an instructor also at MTSU teaching databases and database techniques.

⁴William Inman was an expert consultant, author, and lecturer on data warehousing and was considered by some to be the “father of data warehousing.” He was co-creator of the Corporate Information Factory and creator of the Government Information Factory—Web resources for building information systems.

progress can result in very costly corrections later in the design life cycle. Inmon captured this thought well:

In many respects, the most critical moment in the life of the development of a system is its design. Once the designer casts the form of data “in concrete,” it becomes increasingly expensive (in terms of labor, money, machine processes, etc.) to change the design. The longer the system development process goes, the more expensive and distasteful change becomes, however the system is designed. This hard fact of life is usually learned by management after one (sometimes two) really expensive failures. The philosophy of “build it now, make it run later” has proven over and over again to be an open invitation to all types of failure. (Inmon, 1981, pp. ix-x)

This structuring at the initial stages of database design is important. It is at these stages that the fundamental or basic levels of the structure are put in place as a foundation upon which the rest of the data structure is built. It is important that the basic levels correctly model the concepts the data are to represent to reduce the chances that models that fail to reflect external reality. Without close modeling, the information delivered to decision-makers may be incorrect leading to decisions that are not in the best interests of the organization.

What is a basic level? In a biological structure, or biological taxonomy, the basic level is obvious—it is the species, the class that contains entities or animals, such as dog and cat, which we commonly recognize. But, although it is easy to determine the entities at the basic level in a biological taxonomy,⁵ it may not be as easy to determine the entities in the general case as the examples below indicate.

This research explores the concept of basic classes, how it applies to data structure design, how database designers categorize, structure, and classify the representations of entities in their databases at the basic level—that is the level in the structure upon which the rest of the

⁵ Actually, as discussed later in this proposal, distinction between animals even at the species level is not trivial and is subject to considerable controversy.

structure depends, and what concepts might be useful to inform database designers in the initial stages of database design.

Two Examples

The idea for this research came from two near disasters resulting from design of data structures for business decision-making at the basic level.

Illinois Bell Telephone Company

The first occurred in 1963 when the Traffic Department of Illinois Bell Telephone Company decided to automate their manual system for estimating the number of special service communications lines needed between telephone offices in Illinois. These estimates were used in installation of cables in concrete ducts under streets in Chicago, installation of cables on telephone poles in the Chicago suburbs, and the installation of radio systems in the rest of the state. Errors in the estimates could have resulted major construction costs that the company might not have been able to recover from customer revenue. Therefore, the analysts spent considerable time and energy in testing the automated system against the existing manual system. After several months of testing, revision, and retesting, the results automated estimates continually failed to match those obtained from the manual system. Although the concepts modeled by the automated system seemed identical to those of the manual system, the results indicated that something was different.

The basic concept and basic class used for estimating the required number of cables was an entity called a “span.” Traffic Department estimators translated their service estimates into

specific communication cable estimates, which they then allocated to specific spans. Engineers then combined the estimates in each span to come up with an overall estimate of cable and other construction requirements needed to expand the span's capacity.

After the failure of the automated program, the system designers⁶ began to question their basic concept and entity—that of the span. Was it indeed basic? Were there other entities that interfered with its use for estimating? Discussions ensued with engineers in the three major divisions of the company—Chicago Area, Suburban Area, and State Area. These engineers assured the designers that spans were basic and that they had been used successfully for years in all three divisions to estimate the needed cables to provide the service. As the designers delved more into specific cases, they developed a general definition of what a span was. They interviewed engineers who defined the span as the basic connection between telephone offices and expressed their conviction that that the span was basic because it was a basic connection between two telephone offices. As may be seen, the definition was circular. The span was basic because the engineers defined it as basic. The real problem was not identified until the engineers attempted to define the span in physical terms.⁷

⁶ Today professional who do this sort of design work are called information engineers and information architects.

⁷ Their views were as follows: Suburban Area engineers: A span always connects physically adjacent telephone offices. If three offices (A, B, and C) are in line geographically, there will be a span between A and B and a span between B and C, but no span between A and C. Service from A and C would be provided by two spans, A to B and B to C; Chicago Area Engineers: A span always connects two telephone offices but they are not necessarily adjacent. If three offices are in line, there would be a span between A and B, B and C, and perhaps A and C. Service between A and C would be provided by one span, from A to C, if there were existing lines directly from A to C, if not service would be provided by two spans, from A to B and from B to C; and State Area engineers: A span connects telephone offices where carrier systems⁷ terminate. If three offices are in line, there would be spans between A and B only if a carrier system connects them. If carrier systems connected A with C and B with C but not A with B, the service from A to B would be provided and by the two existing spans—A to C and C to B. A carrier system is a means of combining many communications lines together for transmission over long distances using radio, coaxial cable, or other media. Unlike the communications lines themselves, it requires expensive equipment at the terminals but makes up for the expense by savings on the transmission path. Lines are only available at the terminal points.

The span was basic in each division, but the definitions and underlying concepts of the span was different in each division. Although engineers performing manual calculations could handle the ambiguity when calculating the number of cables needed in each span, the three different concepts of spans caused computational problems when modeled for automated estimating. Ultimately, the data structure had to be redesigned to take into account the different concepts of a span in each of the companies divisions.

Department of Defense

The second near disaster was in the late 1980s. The U.S. Department of Defense (DOD) was developing a new database design to replace the design that DOD had developed for punched cards in the 1960s. DOD awarded Planning Research Corporation (PRC) 36,000 engineering hours to complete the design. The final design ultimately was an entity-relationship data structure with over 700 relational tables and perhaps 2,000 data element entities and attributes. Engineers from the research organization of the Defense Communications Agency (DCA) suggested to PRC that the basic class or element for the design was probably the transmission link—a communications path between two DOD locations. The suggestion was based on quote from a DCA engineer's research paper, "Transmission is basic; everyone uses transmission."

Based on the suggestion that transmission was fundamental, PRC designers began to develop the database structure with transmission as the basic class. At the most fundamental level of the design, characteristics of the transmission entities included the two physical location entities at the ends of the links connected with the assumption that information traveled both ways. Additional entities and attributes related to transmission. After several hundred hours of

design, some of the relationships became very cumbersome and the design was in danger of becoming unmanageable.

Exploration of the difficulties revealed that the basic concept of the data structure—the transmission link—was at fault. Some links, such as radio broadcast links connected more than two locations. These same broadcast links only passed data in one direction. Some new locations under construction did not have any links connected to them yet needed communications. These were indications that the transmission link might not be basic. But if the transmission link was not the basic class, then what was? Engineers finally concluded that the physical locations were basic and links that interconnected the locations were relationship entities rather than basic entities.

Lessons

The impact of these design problems was the expenditure of thousands of dollars in redesign of the data structure so that information could be satisfactorily stored and retrieved. The potential impact was millions of dollars that might have had to be spent if the data structure problems were not discovered early in the database design. The cause of the problem was in the basic class or highest element of the database design taxonomy—the span in the first example and the transmission link in the second.

Research Questions

The cases discussed above, seems to identify several question sets that are worthy of exploration:

1. How do database designers determine the basic classes or entities upon which to build their data structures or taxonomies? What are the processes they use? What guides do they say they use? Do they actually use them? If they do not, why not?
2. What concepts or general principles from organization, classification, and taxonomic literature apply to or are imbedded in the designer's process?
3. What concepts or general principles from organization, classification, and taxonomic literature are might be useful to designers' determination of basic classes or entities upon which taxonomies may be built?
4. What guidance could be provided to help designers make better choices in fundamental classes at the basic level to best support business decision-makers and other users? What incentive could we use to get designers to use such guidance?

Summaries of Concepts Important to the Research

There are several theoretical and practical concepts that are important and potentially impact this research. Each will be summarized here.

Database Structures

Many volumes have been written on design of database structures and systems including the James Martin series (Martin, 1976, 1977, 1983; Martin & Leben, 1995), Teorey and Fry (1982), Rob and Coronel (2000; 2003), Brackett (1987), and the classic by C. J. Date of IBM research (1981) which he popularized a couple of years later (Date, 1983). A reading of these texts show a movement from structures of the networked element-attribute databases models of the 1970s and 1980s, such as the Conference on Data Systems Languages (CODASYL) model

(CODASYL--The Stored-Data Definition and Translation Task Group, 1977) to the structures of the relational table driven entity-relationship models in use today.

These texts all seem to base their database design discussions on the assumption that the designers understand the data structures they are working with. They instruct the database designer to take great care in the design of their database structures without providing guidance on how to do that. They state the database designer must understand structure of the data to be modeled, such as Inmon's advice above, but provide little help in the division of the domain into basic classes be they spans, locations, or other fundamental classes of the domain.

Object Oriented Systems

Object oriented systems provide an alternative approach to representing data. These systems have some features that may shed some light on this proposed research. Helen Klein, in her University of Wisconsin Ph.D. dissertation examined "whether modeling commonly used in business domain concepts in object oriented systems results in well-structured systems" (Klein, 1998, Abstract). In object orientation, "each class should represent, if designed as a popularly prescribed, some real world concept" (Klein, 1998, p. 3). Features of potential interest include encapsulation, inheritance, and polymorphism.

SGML, HyTime, and Topic Maps

Another way of structuring data is through structuring data by internal and external reference. SGML or Standard Generalized Markup Language "is a set of rules for designing

markup languages that describe the structure of a document” (Taylor, 1999, p. 65). SGML “makes possible information-system handling systems for language built on application that specialize in input, management, and output functions replacing applications that control the information and format at every stage” (Alschuler, 1995, p. xiv).

HyTime is an superstructure of SGML that allows it to “deal with time-based multi/hypermedia” ((DeRose & Durand, 1994, p. xvi). SGML tends to be text based with hyperlinks linking documents together. HyTime expands the SGML concept to include other forms of media—film, audio and video tape, etc.—and provides a convention for establishing relational documents separate from the original documents. These documents seem to be much like metadata used in libraries for decades. They may, however, provide access points to many related documents rather than just one.

Topic Maps are a third international standard that further refines the structure provided by HyTime. The standard “provides a standardized notation for interchangeably representing information about the structure of information resources used to define topics, and the relationships between topics” (ISO/IEC, 2002, p. iii). These hypermedia documents can be visualized as sitting above the base of information objects and providing links to the topics in the items at the base(Rath, 2001)

Like databases, and object oriented structures, topic maps provide an alternate way of representing data. However, the structure of that data must be determined by data structure analysts and the problem of identifying the basic classes still remains.

Ontologies

Currently database designers are looking toward domain ontologies to describe the semantic structure of information. According to Fensel, Ontologies provide a “shared and common understanding of some domain that can be communicated between people and application systems” (Fensel, 2000, p. 11). They go beyond standard database schemas in four ways:

- “A language for defining ontologies is syntactically and semantically richer than common approaches for databases.
- The information that is described by an ontology consists of semi structures natural language texts and not tabular information.
- An ontology must be a shared and consensual terminology because it is used for information sharing and exchange.
- An ontology provides a domain theory and not the structure of a data container. (Fensel, 2000, p. 1).

Ontologies, however, have a broader definition than that suggested by Fensel. The Oxford English Dictionary defined ontology as “The science or study of being; that department of metaphysics which relates to the being or essence of things, or to being in the abstract” (*Oxford English Dictionary*, 1989, Ontology). Arlene Taylor added that it provides a “systematic account of existence” (Taylor, 1999, p. 160). Fensel’s use of the term appears to be focused on a formal approach to defining the boundaries of the information or domain that is available to the computer software managing the information—an computer agent “is only able to perceive the part of the world that is allowed by its ontology” (Taylor, 1999, p. 160)

The concept of an ontology is an important concept for understanding structures of information. It is like the context within which the information resides. It provides the semantic structure within which the data is managed. For example, within the ontology of a pet store, a cat

is feline animal. Within the ontology of construction workers, a cat is a large earth mover built by Caterpillar Tractor Company.⁸

It should then be clear that narrow ontologies may be expected to have less ambiguity and simpler structures. Conversely broad ontologies would be expected to have complex data structures.

Basic Classes and Prototypes

Eleanor Rosch, a professor and cognitive scientist of the Department of Psychology at the University of California, Berkeley, is given credit for important fundamental research into the concepts of basic classes and prototype theory. She has written several articles on the basic classes. (Such as 1973; 1975a; 1975b; 1977; 1978). Rosch's work is built on the earlier work of Roger Brown (Brown, 1958) in the 1950s and Brent Berlin in the 1960s (Berlin, 1968). Rosch and Berlin joined forces in the 1970s to continue research on the topic (Berlin, 1978). There is substantial additional literature in building biological and botanical taxonomies that also may apply to this research. Such theory includes the concepts from Linnaeus in the 18th Century through the works of Agassiz (1962), Mayr (Mayr, Linsley, & Usinger, 1953), Blackwelder (1967), Redford (1986), Woodland (1997), etc. This research has shown that basic categories may be important to understanding how things are organized and, for my research, how basic levels are determined for data structure and database design.

⁸ The current name of the firm is Caterpillar Incorporated.

John Dupré defines a natural kind as “a class of objects defined by common possession of some theoretically important property (generally, but not necessarily, microstructural)” (Dupré, 1981, p. 68) as cited in (Bryant, 2000, p. 113).

“I suggest that suitably scientific properties can be understood as *accounting for* patterns of regularity or behavioral characteristics” (Bryant, 2000, p. 114).

I suggest that there is not necessarily a unique taxonomy of the natural world. Even at this level, the world still displays an inherent flexibility and indeterminacy. Alternative patterns of behavioral characteristics run through nature and these alternate patterns can be accounted for by different important physical properties, which in turn produce different natural kinds (Bryant, 2000, p. 114).

Conceptual and Operational Definitions

Conceptual and operational definitions are important to categorizing the data gathered in the research. According to Svenonius, definitions may be conceptual or operational.

A conceptual definition is one that is intentional or connotative: it characterizes what is to be defined in terms of its properties. . . .

An operational definition, on the other hand, is constructive. It specifies rather than characterizes. What it specifies is a set of operations is a set of steps to be followed to identify what is being defined. . . . Operational definitions can be looked on as empirical correlates of conceptual definitions. (Svenonius, 2000, p. 32)

According to Babbie, there are two important definitions in scientific inquiry: nominal and operational. Babbie’s nominal definition is similar to the conceptual definition.

A nominal definition is one that is simply assigned to a term without any claim that the definition represents a “real” entity. Nominal definitions are arbitrary. . . . Most nominal definitions represent some consensus, or convention about how a particular term is to be used.

An operational definition . . . specifies precisely how a concept will be measured—that is the operations we will perform. An operational definition is nominal rather than real but it has the advantage of achieving maximum clarity about what a concept means in the context of a given study. In the midst of disagreement and confusion over what a term “really” means, we can specify a working definition for the purposes of an inquiry. (Babbie, 2001, p. 125)

As categories are established for analysis of the data gathered, definitions will be developed for the categories when these are significant to the research. Such definitions, will allow comparison of the data from this research with that of other researchers and with the theory identified in applicable literature. Initially definitions are nominal or conceptual, allowing organization of the data for analysis and simple comparison with concepts in the literature. These definitions will be tailored to the actual data gathered. If results indicate that measurement data might be helpful or if additional research might be needed in a more controlled environment so that quantitative techniques may be used, or that analysis might benefit, some of the definitions may be operationalized.

Methodology Summary

A study of database designers is required to address the six research questions of the first set of the questions above. This research gathers data through two stages. The first is a review of specific cases studies on database projects available in the literature. The second consists of a study of practices based upon interviews and surveys of database designers. Since the purpose of this research is to study the self-reported behavior of how database designers do their work, rather than to gather numerical data on any specific technique, it seems that qualitative methods should be used to gather and analyze the research data.

Interviews

Interviews are conducted in two stages. To provide a test of the survey content and interview questions, pilot data is gathered from practitioners on the UT Austin Campus

responsible for design UT databases. Based on the results, surveys and interview questions are refined and data is gathered applied from organizations outside the campus. Practically of finding interviewees would suggest the sample would be one of convenience rather than random. It would begin with public and non-profit organizations followed by business organizations. Interviews would be opened-ended, using qualitative techniques, and taking 45 minutes to an hour.

Data Analysis

All the questions require analysis of the data gathered in the survey and interviewing portion of this research. For the first question set, analysis consists of organizing the results of data gathered to find out similar and dissimilar elements. These similarities and differences will provide insight into the needed to determine the designers' process and information about the guides they use.

Analysis for the second question set requires a synthesis of the results of question set one and the background of previous research and theories to determine what work of other scholars is applicable to and what is imbedded in the current research results.

The third question set requires an analysis of other scholarly work in basic classes to determine if elements that are not imbedded in current research could be useful in determine basic classes in practice.

Finally, the results of the above determination are translated into guidance for use in the field with consideration of the bias that designers would have to not use them.

Consideration of Theory and Other Research

In the analysis process, results of the analysis of the survey and interview data are compared with applicable theory and other research. As mentioned in the concepts summary above, there are several threads of theory and prior research that potentially impact on the research of this proposal from the disciplines of computer science, library and information science, cognitive science, psychology, and philosophy. Each of these disciplines provides insight into various aspects of establishing the fundamental basic structure of information and knowledge structures. Various views of classification are available in the literature.

This consideration is expected to be an iterative process. Based on the results of the data gathered during the pilot study, the researcher will review the literature to determine concepts important to understanding the data. These concepts will be used to refine the survey and interview questions for the main study. Then, based on the results of the new data, the researcher will apply the previous concepts, as well as any others that may be applicable to determine if any can inform the practitioner's data design procedures. During each or the iterations, categories may be adjusted to make the most sense of the data.

Results

Information Science includes aspects of an academic discipline and a professional field. I believe that academic research should be influenced by input from professionals and other practitioners in the field. Similarly, the results of research should provide opportunities for application to the problems of and used by professionals in the field. There needs to be a close

connection between academic research and professional tools. Marcia Bates⁹ captured this notion well during a paper delivered in 1999 specifically about that connection: “the holder of a Ph.D. in a professional field must be able to do basic research AND select research questions that may contribute—if not this century then the next—to the practice of that socially useful activity that is a profession” (Bates, 1999, p. 114).

The results of this research are intended to contribute original knowledge to information science on database designers use of basic classes in practice and to contribute to the activity of that profession. It is expected that the results of the research will provide new insight into how practitioners determine the basic classes upon which they design their database structures. From that insight, hopefully, there will be information available from the literature to inform the processes they use.

⁹ Marcia Bates, a professor of information science at University of California, Los Angeles, has conducted extensive research into practices and application of the concepts of information organization.

REFERENCES

- Agassiz, L. (1962). *Essay on Classification*. (Original 1859). In E. Lurie (Ed.), *Papers on classification*. Cambridge MA: Belcamp / Harvard University Press.
- Alschuler, L. (1995). *ABCD...SGML: A user's guide to structured information*. London: International Thompson Computer Press.
- Babbie, E. R. (2001). *The practice of social research* (9th ed.). Belmont, CA: Wadsworth/Thompson Learning.
- Bates, M. J. (1999). *The role of the Ph.D. in a professional field*. Retrieved October 14, 2003, from <http://www.gseis.ucla.edu/faculty/bates/phdrole.html>
- Berlin, B. (1968). *Tzeltal numerical classifiers*. The Hague: Mouton.
- Berlin, B. (1978). Ethnobiological classification. In E. Rosch, B. B. Lloyd & S. S. R. Council (Eds.), *Cognition and Categorization* (pp. 9-26 23 cm). Hillsdale NJ: Lawrence Erlbaum.
- Blackwelder, R. E. (1967). *Taxonomy: A text and reference book*. New York: John Wiley and Sons.
- Brackett, M. H. (1987). *Developing data structured databases*. Englewood Cliffs, NJ: Prentice-Hall.
- Brown, R. (1958). How should a thing be called? *Psychological Review*, 65, 14-21.
- Bryant, R. (2000). *Discovery and decision: Exploring the metaphysics and epistemology of scientific classification*. Madison, NJ: Fairleigh Dickenson University Press.
- CODASYL--The Stored-Data Definition and Translation Task Group. (1977). Stored-data description and data translation: A model and language. *Information Systems*, 2(3), 95-148.
- Date, C. J. (1981). *An introduction to database systems*. Reading, MA: Addison-Wesley.

- Date, C. J. (1983). *Database: A primer*. Reading, MA: Addison-Wesley.
- DeRose, S. J., & Durand, D. G. (1994). *Making hypermedia work : a user's guide to HyTime*. Boston: Kluwer Academic.
- Dupré, J. (1981). Natural kinds and biological taxa. *Philosophical Review*, 90, 86-90.
- Fensel, D. (2000). *Ontologies: A silver bullet for knowledge management and electronic commerce*. Berlin: Springer-Verlag.
- Inmon, W. H. (1981). *Effective data base design*. Englewood Cliffs, NJ: Prentice-Hall.
- ISO/IEC. (2002). *ISO/IEC 1320 Topic Maps*. New York: American National Standards Institute (ANSI).
- Klein, H. A. (1998). *Does modeling real world objects in object-oriented systems result in well-structured systems?* Unpublished Dissertation, University of Wisconsin, Ipselanti.
- Martin, J. (1976). *Principles of data-base management*. Englewood Cliffs, NJ: Prentice-Hall.
- Martin, J. (1977). *Computer data-base organization* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Martin, J. (1983). *Managing the data-base environment*. Englewood Cliffs, NJ: Prentice-Hall.
- Martin, J., & Leben, J. (1995). *Client/server databases: Enterprise computing*. Upper Saddle River, NJ: Prentice-Hall PTR.
- Mayr, E., Linsley, G. E., & Usinger, R. L. (1953). *Methods and principles of systematic zoology*. New York: McGraw-Hill.
- Oxford English Dictionary*. (1989, 2000 May 4). Retrieved 2003 December 15, from <http://www.oed.com>
- Rath, H. H. (2001). *Topic Maps: Introduction and Allegro*.

- Redford, A. E. (1986). *Fundamentals of plant systematics*. New York: Harper and Row.
- Rob, P., & Coronel, R. (2000). *Database systems: Design, impementation, and management* (4th ed.). Cambridge, MA: Thompson Learning.
- Rob, P., & Coronel, R. (2003). *Database systems: Design, impementation, and management* (5th ed.). Cambridge, MA: Course Technology, Thompson Learning.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4, 328-350.
- Rosch, E. (1975a). Cognitive reference points. *Cognitive Psychology*, 7, 532-547.
- Rosch, E. (1975b). Cognitive representation of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233.
- Rosch, E. (1977). Human categorization. In N. Warren (Ed.), *Studies in cross-cultural psychology*. London: Academic Press.
- Rosch, E. (1978). Principles of categorization. In E. Rosch, B. B. Lloyd & S. S. R. Council (Eds.), *Cognition and Categorization* (pp. 9-26). Hillsdale NJ: Lawrence Erlbaum.
- Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, MA: MIT Press.
- Taylor, A. G. (1999). *The organization of information*. Englewood, CO: Libraries Unlimited.
- Teorey, T. J., & Fry, J. R. (1982). *Design of database structures*. Englewood Cliffs, NJ: Prentice-Hall.
- Turban, E. (1990). *Decision support and expert systems: Management support systems* (2nd ed.). New York: Macmillan Publishing.
- Woodland, D. W. (1997). *Contemporary Plant Systematics* (2nd ed.). Berrien Springs MI: Andrews University Press.

Proposed Committee Members

Dr. Francis M. Miksa, Professor, School of Information. Chair Dr. Miksa is an expert in organization, classification, and taxonomies. His research interests center on classification theory, the organization of materials for access, and the history of library and information science.

Expected Contributions to this Research: It is expected that Dr. Miksa's extensive knowledge of and organization and classification theory and his experience in its application will inform all aspects of this research.

Dr. Larry Browning, Professor, School of Communications, is an expert on ethnographic research. His current research projects include examining technology production and how Information Communication Technologies are used in the business world. His consulting work includes both public and private organizations such as Shell Oil, SYNTRO Corporation, Tokyo Electron of America, 3M Corporation, Internal Revenue Service, Digital Equipment Corporation, and Texas Electric Cooperatives. He was Cofounder and Vice President of Partnerwerks, a research, training, and consulting firm specializing in high technology organizations.

Expected Contributions to this Research: It is expected that Dr. Browning's knowledge of ethnographic research principles and his experience with application of those principles in practice will inform this research in the data gathering and analysis stages as well as developing guidelines for basic level organization in using information technologies.

Dr. Glynn Harmon, Professor, School of Information. Among Dr. Harmon's research interests include human cognition with focus on informatics and education for information science. In addition, he has worked as research associate in the Center for Documentation and Communications Research, Case Western Reserve University.

Expected Contributions to this Research: It is expected that Dr. Harmon's knowledge of cognition and his research experience will inform this researcher in the understanding of the results from the interviews and the applicability of cognitive theory to them in general and specifically in answering the questions in the first and second sets.

Dr. Randolph Bias, Associate Professor, School of Information. Dr. Bias has worked as a usability engineer in industry for over 20 years, with companies such as Bell Labs, IBM, and BMC Software, helping software developers make human-computer interfaces, including web sites, user friendly.

Expected Contributions to this Research: It is expected that the experience Dr. Bias has with industry in organizing and presentation of information considering usability aspects will provide this researcher insight into user concerns particularly during the work on question 4 dealing with developing guidance for information architects.

Dr. Don Turnbull, Assistant Professor, School of Information. Dr. Turnbull's research is focused on Database Design and Management, Human-Computer Interaction, Data structure, Information Storage and Retrieval, Online Searching/Information Retrieval, Systems Analysis

Expected Contributions to this Research: It is expected that the knowledge and experience Dr. Turnbull has in databases and data structure as well as research with profit and non-profit organizations will help this researcher to understand the data gathered to answer the first question set.