A methodology for the evaluation of source database quality in a decision support system or data warehouse: development and testing

M. Pamela Neely
Doctoral Symposium
Abstracts

*ADAPTIVE LEARNING SYSTEMS: AN AGENT-BASED MODEL FOR SMALL BUSINESS ENTREPRENEURS TO IMPLEMENT IT*  
Deyaa David
Zhejiang University Business – Baruch College, Graduate Center
New York New York
New York, New York 10010  
E-mail: david_d@baruch.cuny.edu

Adaptive Learning Systems (ALS) have garnered increasing attention in recent years from academia, practitioners and trade press. The number of agent conferences and World Wide Web agent sites developed lately evidence this. It is part of the agent paradigm. "Adapting" the research world that provides new ways of analyzing, designing and implementing complex software systems (Jennings et al. 1998). Although varied in description, these systems are characterized by behaviors that are adaptable by self-organization of actors and flexible learning the user's preferences, styles and cognitive levels thereby having positive forms of interaction/support. ALS are user-centered and have the potential to revolutionize the way users interact with computers overcoming many of the limitations of current systems. A review of the literature indicates that researchers have been complaining about the lack of these systems' use as practical tools for real world problems (Mans, 1994; Hoek 1996; Nakao 1996; Jennings et al. 1996; Bradshaw 1997) This research agrees with these researchers and identifies a domain that can significantly benefit from this technology.

The authors depict that ALS hold the potential to facilitate the decision-making process regarding Information Technology (IT) implementation. Due to the potential benefits that IT provides, small business entrepreneurs (SBEs) are faced with daunting task of implementing IT. To them the decision to adopt IT is magnetic. However, the decision-making process poses many challenges. Additionally, failure to adopt IT can mean an important decision regarding IT implementation can be detrimental to their business. Generally, SBEs cannot withstand such risks. In short, the decision to adopt IT is weighted complex, challenging and risky. Prior research indicates that in order to minimize these risks, support tools must be developed. These tools should be adaptable to the user's performance and habits (Chang, 1998; Argyros, 1994; El-Hajj et al., 1993; Thau, 1993). ALS have the characteristics needed to develop tools to support the SBE during IT implementation decision-making process. Flexible, self-adapting systems capable of accomplishing tasks on behalf of the user. Consequently, ALS can be used as a tool for SBE IT implementation decision-making.

With an adaptive learning system (ALS) framework, small business entrepreneurs in making effective IT implementation decisions. The framework is based on an adaptive decision support system. This research adopts a two-phase approach. Phase 1: Review of the literature and conceptual model. Phase 2: Validation of the model.

This model contains the following attributes: user centered, adaptable, communicative, autonomous, flexible, self-organizing, central, knowledge base, and degrade gracefully.

**Phase 2: As proof of concept, an ALS based on the conceptual model will be developed and validated. Evaluation of agent-based systems should not be a task that is only means a simple task. Agent-based systems or validation requires validation methods are not standardized. Adapting Victor et al. (1990) validation of knowledge-based systems approach to agent must be used. A standard for validation of ALS Victor et al. (1990) if the system has a standard for validation of ALS Victor et al. (1990) it is used as a standard in the decision making process. Hence, it is important to distinguish several major features of the ALS in this effective IT system and can be used as a reference to the SBE. This research will integrate the effectiveness of an ALS in Adaptive Learning Systems. Adaptive Learning Systems refer to the ALSTool (2000). This tool is used in the decision-making process. The results will be compared with a laboratory experiment. The experimental will use a predefined agent and a personal assistant manager developed for tight regulation. The validation will follow steps outlined by Victor et al. (1990) that ALSTool is reliable and robust. And the constraints for measurement to be used by the experiment will be developed from previous research and evaluation. Based on the general research model the following hypotheses were developed.

1. ALSTool will be easier to use than a non-adaptive IT tool in terms of user interface.
2. ALSTool will lead to a greater understanding of the user than a non-adaptive IT tool. These results are especially true for female users.
3. ALSTool will lead to a greater perceived usefulness of the IT system than a non-adaptive IT tool. The effects will be true for all technical and non-technical users.
4. Use of the ALSTool will lead to an improved perceived performance of the decision-maker. This is especially true for non-technical users.
5. Use of the ALSTool will lead to a greater perceived ease-of-use of the IT system than a non-adaptive IT tool. The effects will be true for both technical and non-technical users.
6. ALSTool will explicitly lead to more competitive tools of the IT system than the non-adaptive IT tool. The effects will be true for both technical and non-technical users.
7. The explanations given by ALSTool will allow for greater user satisfaction than the non-adaptive IT tool. The effects will be true for both technical and non-technical users.
8. The quality of decisions provided by ALSTool implementing tool will be at least as good as decisions provided by a human agent.

The research hopes to contribute to the existing agent research theories and applications of agents. Based on the literature, the level of practical tools for real world problems provide a realistic application to a heterogeneous body of users to exploit IT. It will significantly contribute to research designed to assist small business entrepreneurs in making advantage of IT.

Future research will include a field test of ALSTool and a longitudinal study to explore other factors that are learned by the constraints of the environment.

References available upon request.
A METHODOLOGY FOR THE EVALUATION OF SOURCE DATABASE QUALITY IN A DECISION SUPPORT SYSTEM OR DATA WAREHOUSE: DEVELOPMENT AND TESTING

M. Parvin Norty
Center for Technology in Government
State University of New York, Albany
518-442-3924
Fax 518-442-3886
norty@ctg.albany.edu

OVERVIEW

Decision Support Systems (DSS), Executive Information Systems (EIS) and Data Warehouses (DW) are construct to support management decision making. And yet, if the underlying data is of poor quality, the decisions resulting from the use of these tools may be faulty. These decisions may lead to directions that are merely inconvenient or they may lead you down a path that is potentially catastrophic (Arnold 1998). The accuracy, consistency, and completeness of data in a DW or DSS has been a concern of the output of these systems. As much as 75% of the effort in building a DSS, DW or EIS is spent on issues associated with cleaning and maintaining the data (Arte 1998). The development of a methodology, potentially coupled with a software tool, which would guide the system developer towards database problems, could reduce the development cost and increase the reliability of the data in the DW, DSS, or EIS. In the auditing profession, audit plans are commonly used to direct the auditors through the process of detecting problems within the data that could lead to material misrepresentations of the financial position of a company or even prevent fraud. A similar approach could be used when detecting discrepancies within a database for use in a decision support tool.

This research is in the very early stages. I have defined the problem to be addressed and have outlined the general framework for the research. I am in the process of identifying the literature in auditing and database quality to support the direction that I am taking. Additional potential areas to be explored include judgment theory (Phillips 1994), decision making theory (Oliver 1977) and theories of error detection (Klein 1997). The outcome of this research will contribute to the successful development of decision support systems by allowing a system developer to be more focused on the quality of the data to be migrated into a new system. Better quality data, detected in a timely, cost-effective manner, will result in a better decision support tool.

RESEARCH PROBLEM

Historically, data has been collected in On-Line Transaction-Processing (OLTP) systems for day-to-day operations. These systems can have accuracy, timeliness, consistency and completeness database quality problems, to name a few of the data quality dimensions found in the literature (Wang and Strong 1996, Redman 1998, Tayi and Ballou 1998). Despite these data quality problems, the data is used to produce financial statements, make purchasing decisions, and make sales calls on customers (Murphy and Groomein 1997, Wang 1998). Frequently, the data is examined by knowledgeable individuals prior to the release of the data for reporting purposes (Ballou and Tayi 1996). These individuals will take into account the source or reputation of the data, the reasonableness of the data, and prior history with the data (i.e., historically this particular free form field has held a billing code while that one has a patient diagnosis). The decision to report the data is based on an individual’s judgment as to its fit for the purpose intended (Tayi and Ballou 1998).

With the advent of DSSs, EIS and DWs, we find that data being used for secondary purposes (Chauhurt 1997). The data is extracted from a source database and placed in a separate system for purposes of management decision making. As the data is separated from its source, it is also separated from many of the facilities necessary to judge the quality of the data. Examination of each individual data source should be made prior to migration of the data to the decision support system. The literature discusses various potential frameworks for assessing data quality in databases (Hoffermeer 1997, Willshire 1997, Kahn and Strong 1998, Kaplans, Krishnan et al. 1998, Kaplan 1998, Sorrey and Wang 1998). None of the models or frameworks specifically guide a decision support developer through the steps of evaluating individual databases for quality nor do they focus on the database structure which must also be evaluated prior to migration of data into a DSS. The purpose of this research is to develop a methodology from the experiences of experts in the field. Once the methodology has been developed, it will be tested to determine if more data errors are detected with the methodology than without it.

In a financial audit, where every company in be audited is different, every data source is different. The same process that is used in a financial audit could be used in the evaluation of data sources for DSS. The development of a methodology to process data sources in a systematic manner will help guide the developer of a DSS towards potential data problems that can be dealt with prior to their inclusion in the DSS. The consistent use of a methodology, much like the use of an audit plan, would guide the developer and promote the discovery of problems.

METHOD

System developers use methodologies to develop information systems (Aston and Fitzgerald 1988). Auditors use audit plans, frameworks, or methodologies when conducting an audit of the financial records of a business. Developers of decision support systems must have the skills to both develop a system and audit the input to the system. The knowledge that they have must be extracted to develop a methodology or potentially a meta model, for the evaluation of source database quality. I plan to:

• Develop an interview to extract the knowledge from data warehouse developers in regards to cleaning data prior to migration to a data warehouse.
• Pilot the interview on a graduate student.
• Interview some number of data warehouse developers. Particular attention will be paid to those experts who are currently using a methodology either with or without a software component.
• Develop a methodology that incorporates the best practices of each of the experts.

Testing will involve a set of databases that were eventually incorporated into a prototype data warehouse. Some number of subjects will evaluate the databases, using the methodology, and the number of database errors found by each individual will be evaluated. The databases will also be processed through three levels of data quality software tools. Finally, a control group will evaluate the databases without a set methodology. Both novices and experts will be used in the testing process.

EXPECTED OUTCOMES

It is anticipated that a methodology will allow the auditor to detect more errors in the databases than an individual working without a methodology. It is also expected that a human, working with a methodology, will detect more errors than a software tool alone. It is possible that the final methodology will incorporate a tool to automate some of the processes, but the novel aspect of this research is the development of a methodology that will make the process of detecting errors more efficient and that there will be savings in time and money in the development of a DSS or DW.

CONCLUSION

Data warehouses, decision support systems and executive information systems rely on data (taxation processing systems for their input. As the data moves from the OLTP environment into the decision support system environment it is also remodeled from the individuals that are knowledgeable of the data and its use. It is imperative that the data be cleansed, prior to the degree possible, prior to migration into the decision support tool. The developer of the DSS, DW or EIS is unlikely to be the primary user of the source data. A systematic approach to evaluating the data, much like the systematic approach used to evaluate data in a financial audit, will allow the developer to detect more of the errors in the database and thus create a more reliable, useful decision support tool. Development and testing of this methodology is the goal of this research.

DOCTORAL SYMPOSIUM PROPOSAL

Vicky Wilson
Coordinator: Graduate Diploma of Science (Information Science); Graduate Diploma of Science (Teacher Librarianship) School of Computer and Information Science, Edith Cowan University