Community and Data Integration Approach Using Requirement Centric Operational Data Store Model (ReCODS-Model) for Business Intelligence Applications

MUHAMAD SHAHBANI ABU BAKAR & NORSHUHADA SHIRATUDDIN
Graduate Department of Computer Science
College of Arts and Sciences (Applied Science)
Universiti Utara Malaysia
06010 UUM Sintok, Kedah
MALAYSIA
shahbani@uum.edu.my; shuhada@uum.edu.my

Abstract: Building a Business Intelligence (BI) application is very challenging as it is a young discipline and does not yet offer well-established strategies and techniques for the developments process when compared to the software engineering discipline. Furthermore, information requirements analysis for BI applications which integrate data from heterogeneous sources differs significantly from requirements analysis for a conventional information system. Requirement Centric Operational Data Store Model (ReCODS-M) to build BI application that focuses on operational information to support business operations is proposed. In this model, combination of community interaction and data integration approach were used to identify the requirements for developing BI application. Furthermore, how the operational data store can be used for operational and tactical information and can be transferred to a data warehouse for supporting analytical information and decision making is also presented. Finally, to verify and validate the proposed model, the case study approach using web application development in selected subject areas is elaborated.

Keywords: Requirement, Community, Data Integration, Operational Data Store, Data Warehouse, Business Intelligence, Web Application

1. Introduction

Information requirements analysis for Business Intelligence (BI) applications which integrate data from heterogeneous sources differs significantly from requirements analysis for a transactional information system [3].

BI is the process of gathering meaningful information about the subject matter being researched [4]. In information system perspective, BI is a combination of operational data with analytical tools to present complex and competitive information to planners and decision makers [5]. The structures commonly used in BI architecture are an operational data store (ODS), data warehouse (DW) and data mart (DM). To date, there are many topics researched in DW structure (which support analytical information) but fewer studies on ODS structure. ODS is subject oriented, integrated, current valued and volatile collection of detailed data that provides a true enterprise view of information [6]. The major difference between the ODS and DW is ODS contains current and detail data while DW contains summary data to support analytical information for making decision [7]. Moreover, ODS structure can support both operational and analytical information for a decision making.

Since the success of a system application depends on how well it fits the requirements from the users and its environment [1], we propose Requirement Centric Operational Data Store Model (ReCODS-M) to build BI application that focuses on operational information to support business operations. In this model, combination of community interaction and data integration approach were used to identify the requirements for developing BI application.

2. Motivation and Related Works

This paper is based on the concepts of the Corporate Information Factory (CIF) introduced by Inmon [7] and Business Dimensional Lifecycle for Business Requirement suggested by Kimball [8] to develop BI applications. Furthermore, the concept of Information

Pyramid as shown in Figure 1 can be seen as different types of information and different view of users.

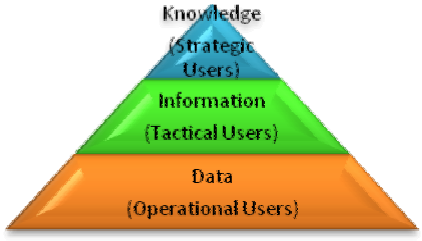

Fig.1 Information Pyramid

Inmon points out the alignment of different DW elements with different requirements as shown in Figure 2. The ODS, DW and older legacy applications are combined to create a common CIF architecture. The raw and detail data is integrated and transformed into an ODS or current detailed level of a DW. As the refined data passes out of the ODS it goes into the current level of the DW. Information processing can be done throughout ODS level, at a current level of detail, or at the data mart level of detail.

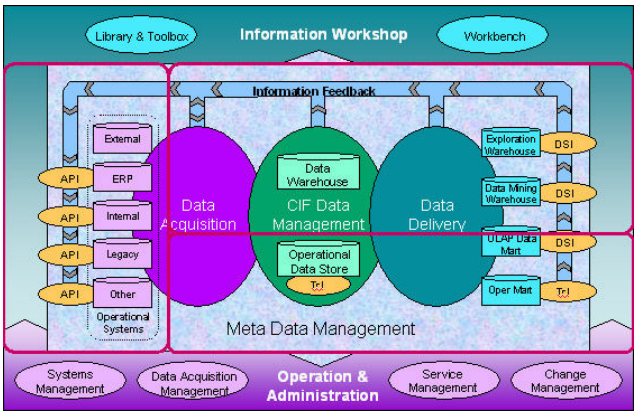

Fig.2 Corporate Information Factory [7]

In contrast, the concept of Business Dimensional Lifecycle for Business Requirement proposed by [8], guides the developer in making strategic choices to prioritize subject areas and how to present required information on the users' screens. The important points to understand here are understanding business requirements, securing solid business sponsorship, defining enterprise-level business requirements and identifying detail subject areas of business requirements.

Basically, existing DW development approaches can be classified within three basic groups; data driven, a goal driven and user driven [9]. The researchers such as [10] and [11] argues that a DW environment is data driven, which are requirements are understood after it is populated with data and being used by the decision support analyst. Moreover, Goal Oriented Requirement Analysis for Data Warehouse (GRAnD) suggested by [12] adopts two different perspectives for requirement analysis - organizational modeling centered on stakeholders and decisional modeling focused on a decision maker. In this approach, DW project must fit with organization business objectives. On the other hand, user-driven or demand-driven approach adopts involvement of end users in data warehousing as suggested by [3]. Data modeling in DW requirement derives a data model directly from a user query requirement without considering the data sources and business goal.

Unfortunately, there are fewer studies about ODS, which support information processing at current details level or operational data in BI applications, especially in ODS requirements compared with studies about requirement analysis in a data warehouse.

3. Role of ODS in BI applications

ODS is an environment where data from a different operational database is integrated [14]. Complementary to the BI architecture, ODS is positioned between the operational systems and the DW as shown in Figure 3. ODS places the tactical operational data needs of an organization by providing current integrated information, whereas the DW supports the strategic needs of an organization by providing more historical data. The ODS contains detail data in particular subject area while the DW contains much summary data in several business processes in an organization.

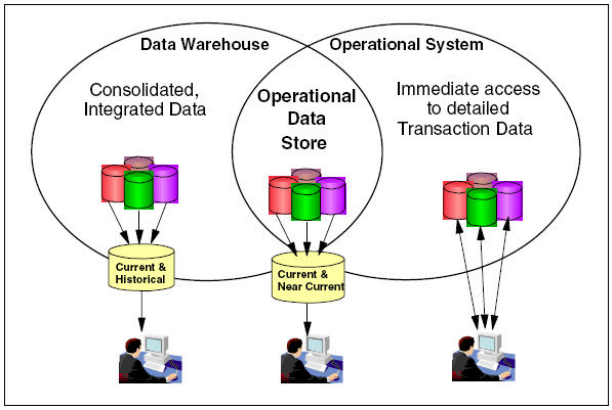

Fig.3 Positioning of ODS [14]

The purpose of ODS in BI applications is to provide the end user community with an integrated view of enterprise data. It enables the user to address operational challenges such as operational integration, Decision Support System on operational data and staging area for eclectic analytical processing [14]. Moreover, there are four classes of ODS to extract data that are, Class 1- online synchronous, Class II – done hourly, Class III – typically every night and Class IV- upon request [7]. These are differentiated by the level of integration between the operational systems and the ODS based on update processes.

4. The proposed ReCODS-M

ReCODS-M is a requirement model to develop BI application, which is focused on operational data. In this model, a two-phase requirements process, at the organizational and subject area levels, is proposed. In between the two-phase requirement process, there is a requirement elicitation process which contains

requirement gathering approach and community collaboration method.

Figure 4 illustrates the proposed model. Each phase is broken down into several sub-activities. For each phase, sub-activity begins with understanding project domain and ends with writing up requirement specification in a different level of details.

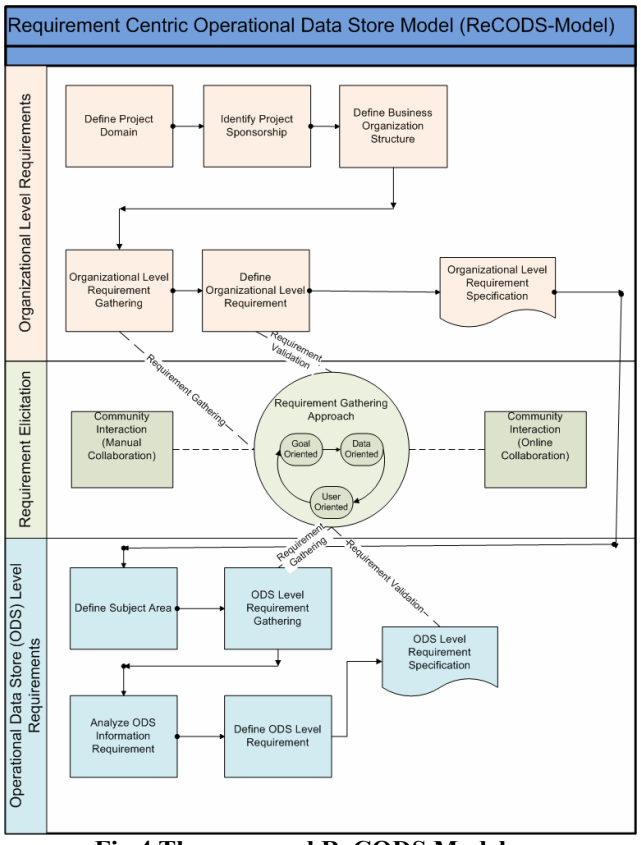

Fig.4 The proposed ReCODS Model

For the requirement elicitation process, three requirement gathering approaches (goal oriented, data oriented and user oriented) to gather requirement from an organizational level and subject area level are suggested. Besides that, community interaction and collaboration using such as interview, survey, data profiling and reports review to gather the requirements are emphasized. In addition, online collaboration for system stakeholder/users to discuss system requirements using forum, chat, blog, email and digital article are utilized.

4.1. Organizational Level Requirements

In the Organizational Level Requirement phase, it starts with understanding project domain from an organization level to build an enterprise information infrastructure and ends with writing up the organizational level requirement specification. The organizational level requirement step gathers a broad and horizontal view of the organization from business point of view. It involves the following six steps.

Step 1: Define Project Domain. This is the process of analyzing the business domain and related process in the specific business area such as insurance, retail, banking/finance, education, and telecommunication. Each project domain has different business functions, requirements, processes and questions depending on the type of business sector. Here domain analysis concept can be utilized to understand the domain knowledge in the specific project domain based on past business processes. The sources of domain knowledge are from technical literature, existing implementations, customer survey, expert advice and current/future requirements.

Step 2: Identify Project Sponsorship. Project Sponsor is the person/organization that is ultimately responsible for the project within an organization. Normally, it involves a senior management post such as chief executive officer, managing director, general manager and the owner of an organization. Typically, project sponsorship is responsible for: 1) championing the project; 2) obtaining budget approval for the project; 3) accepting responsibility for problems escalated by the project manager and 4) document approval for the project. Project Sponsorship can be categorized into three types of organization that are a government agency, business organization and software developer. Each type has a different business motivation for developing BI applications. Good business sponsorship can provide the resources and support to deliver real business value.

Step 3: Define Business Organization Structure. This process involves understanding the organization's vision and motivation, structure and business activities. The information about business organization can be retrieved from organization website, reports and sources from management staffs. By understanding business organization structure, BI developer can understand data sources and information flows in the organization. Business activities in the organization can derive a clear picture for the motivation, guideline, business functions and project scope in BI application.

Step 4: Organizational Level Requirement Gathering. In this step, the focus is on the high level requirement for gathering information process in an organization. Three approaches to gather requirements are used; 1) goal driven; 2) user driven; 3) data driven. Goal driven approach is based on business motivations set up by an organization. The user driven is based on demand from the users and the last approach is based on data profiling on existing organization data sources. In gathering the requirements, community collaboration techniques (such as survey, data profiling, forum, chat, blog, email) are emphasized to convey requirements. Requirement validation is also ascertained in this step.

Step 5: Define Organizational Level Requirement. A high level requirement collected from previous step will be analyzed in this step. Three main processes are proposed: 1) Build Initial Matrix; 2) Conduct Prioritization Session and 3) Write up Summary Requirement. The business processes in the initial

matrix become the major inputs to the requirements for prioritization session. The prioritization process is a meeting involving the BI team and business sponsor/senior management to describe the business process. A prioritization grid is normally used to illustrate the prioritization process which consists of Y axis as business value and X axis as a level of effort. The output of this prioritization process is a list of a business process in priority order. The Business Sponsor will decide several business processes/subject area to be selected for the BI project.

Step 6: Organizational Level Requirement Specification. BI project manager is responsible for writing the summary requirements in an organizational level requirement process and this specification is used for the next detail requirements process.

4.2. Operational Data Store Level Requirement

Operational Data Store Level Requirement focuses on requirement gathering and analysis in a specific subject area. This phase involves the following five steps.

Step 1: Define Subject Area. In this step, a specific subject area which is identified in the previous phase is elaborated. General sub-steps as proposed by [15] are adopted: establish the subject area, collect domain expert, establish the depth and width analysis and define the specific domain objects, relations and constraints. The outputs of this step are taxonomies, standard interfaces, functional models and domain languages as a requirement to develop a software system.

Step 2: Operational Data Store Level Requirement Gathering. Here attention is given to the detail requirements in specific subject areas. In gathering the requirements, community collaboration techniques (such as survey, data profiling, forum, chat, blog, email) are also emphasized to convey requirements.

Step 3: Analyze Subject Area Process. This step involves a process to analyze reports required by organization based on a previous requirement process. The suggested reports obtained from a requirement process are matched with BI users and type of information status. Three groups of BI users are defined: strategic, managerial and operational users. The criteria to define these groups are based on discussion results from the community collaboration and type of reports such as detail level of information, frequency of data updated, summarization of data and type of information users. The reports are also classified by information status: 1) need to have; 2) nice to have and 3) not needed. An example of how the specific reports relate to information status and report users can be shown in Figure 5.

Step 4: Define Operational Data Store Process. An ODS is an environment where data from a different operational database is integrated to provide the end user community with an integrated view of operational and tactical information. Hence, this step focuses on a detailed requirement analysis for the ODS in specific subject areas. Briefly, there are three major tasks in determining the requirements of ODS: 1) ODS Integration; 2) ODS Characteristic and 3) ODS Classification. Figure 6 shows the steps in the ODS requirement process suggested in this model.

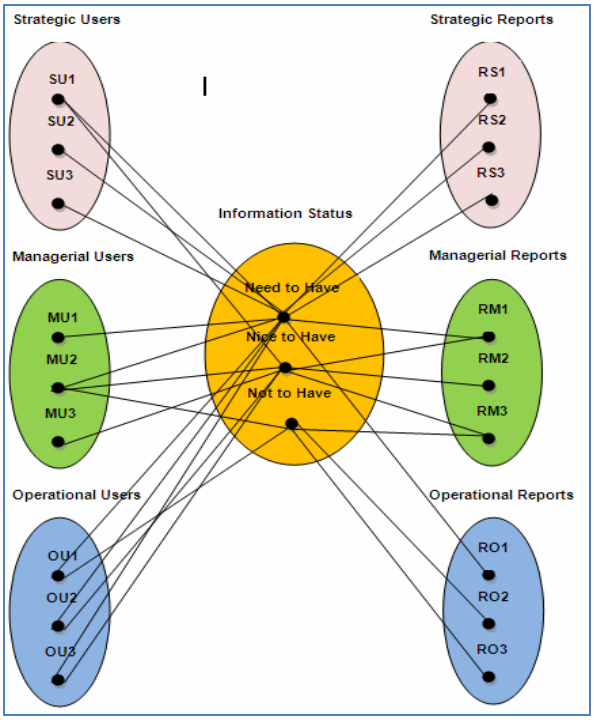

Fig.5 Relationship between reports, report users and information status.

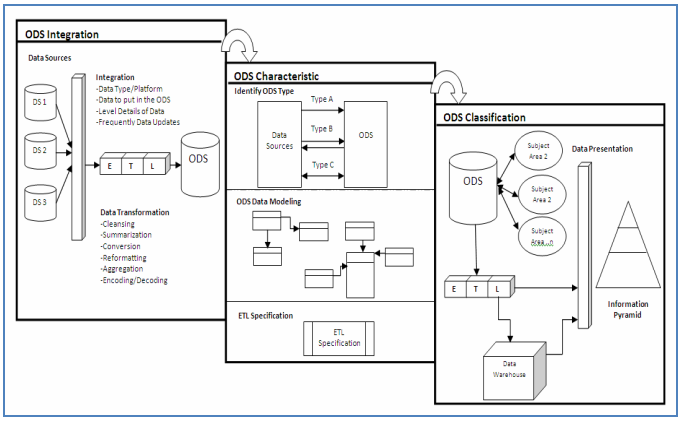

Fig.6 Operational Data Store Requirement Process

Step 5: Operational Data Store Requirement Specification. Requirement specification for the Operational Data Store Level must be written in detail for BI developer. The following requirements must be documented in detail specification: 1) subject area analysis; 2) information requirement analysis (type of report, level of users and type of information status) and 3) ODS Processes (including ODS integration, ODS Characteristic and ODS Classification). It is also suggested that the ODS requirement specifications are documented using a standard notation such as Unified Modeling Language (UML) or existing BI tools.

4.3. ReCODS-M Architecture

ReCODS-M architecture is a way of representing the overall structure of the requirement process for developing BI application that focuses on ODS. Figure 7 shows the architecture, which consists of a set of task component and actors involved for BI application development.

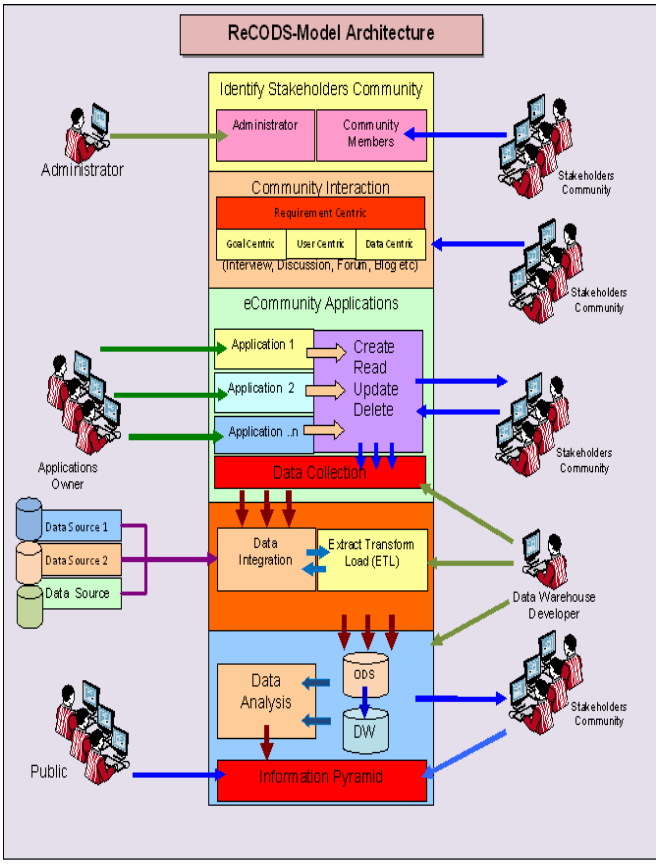

Fig. 7 ReCODS Architecture

5. BI Web Applications as Case Studies

In validating the proposed ReCODS-M, three BI web applications for three organizations were developed as case studies. The organizations involved are a utility company, a telecommunication company and a government agency. Figure 8 depicts a screen shot of one of the applications.

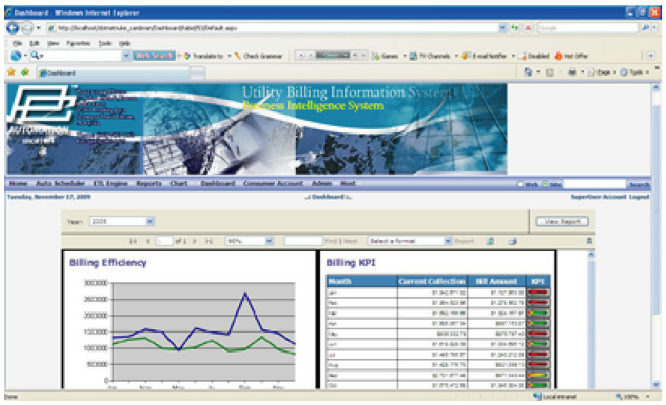

Fig.8 BI Applications Developed Using ReCODS-M

The applications were developed by following all the steps proposed in the ReCODS model. In each of the organizations, a team consisting of developers, sponsors, users and stakeholder community were involved. Table 1 show three case studies were conducted by developed BI applications using ReCODS Model.

Tab. 1 Case Studies Using ReCODS Model

Organization/ Subject Area	Users Community	Business Requirement	Target Information
Utility Company- Water Billing- Billing and Payment	- Director - Engineer - Accountant - System Analyst - Technician - Clerk - Consumers	Consolidation of three dominant applications (Water Billing, Non Revenue Water System and Water Production System) to allow online billing and collection management by user community	Online Billing and Collection Efficiency
Government Agency- Graduate Entrepreneur Profile	- Manager - Program Planner - Loan Officer - Entrepreneur - Suppliers - Customers	Entrepreneur Department requires a history of entrepreneur transaction to evaluate their performance	Online Graduate Entrepreneur Profile
Telecommunication Company- Preventive Maintenance	- Region Manager - Engineer - Accountant - Technician - Clerk - Call Center Officer - Sub Contractors	Account Department requires online monthly preventive maintenance transaction information for a payment claims	Online Preventive Maintenance transactions

Formative evaluations of the applications indicated that these BI applications are usefulness, easy to be used and produced reliable information. Detail discussion on the findings of these evaluations is not included in this paper.

6. Conclusion

The ReCODS model proposed in this paper represents a requirement process for developing BI system that is focused on ODS function, which support operational and tactical information. The model is divided into organizational requirement and operational data store level requirement. Organizational requirement level focuses on a broad and higher level requirement in organization while operational data store requirement emphasizes in a more specific subject area.

References

[1] B. A. Nuseibeh and S. M. Easterbrook, "Requirements Engineering: A Roadmap", In A. C. W. Finkelstein (ed) "The Future of Software Engineering", IEEE Computer Society Press, 2000.

[2] Software Engineering Body of Knowledge, SWEBOK, 2004.

[3] Winter, R., & Strauch, B., *A Method for Demand-driven Information Requirements Analysis in Data Warehousing Projects,* Paper presented at the 36th Hawaii International Conference on Systems Sciences, 2003.

[4] Wu, J. , *What is Business Intelligence?* Retrieved 21 April, 2009, from http://www.dmreview.com/article_sub.cfm?articleId=1924, 2000.

[5] Negash, S., & Gray, P. *Business Intelligence.* Paper presented at the Ninth Americas Conference on Information System (AMCIS 2003), Tampa, Florida, (2003, 4-6 August).

[6] Imhoff, C. *Crystal Clear Customer: The Role of the Operational Data Store*. Retrieved 21 April, 2009, from http://www.intelsols.com/documents/Imhoff_10-02.pdf, 2002.

[7] Inmon, W. H. *Building the Operational Data Store* (2nd ed.): John Wiley & Sons, Inc., 1999.

[8] Kimball, R., & Ross, M. *The Data Warehouse Toolkit:The Complete Guide to Dimensional Modeling* (2nd ed.): John Wiley & Sons, Inc., 2002.

[9] List B., Bruckner R., Machaczek K., and Schiefer, J. A Comparison of data warehouse development methodologies: Case study of the process warehouse. In Proc. DEXA, 2002.

[10] Boehnlein, M., & Ende, A. U.-v. Deriving Initial Data Warehouse Structures from the Conceptual Data Models of the Underlying Operational Information Systems, 1999.

[11] Moody, D. L., & Kortink, M. A. R. *From Enterprise Models to Dimensional Models:A Methodology for Data Warehouse and Data Mart Design.* Paper presented at the International Workshop on Design and Management of Data Warehouses (DMDW'2000), Stockholm, Sweden., 2000, 5-6 June.

[12] Giorgini, P., Rizzi, S., & Garzetti, M. *Goal-Oriented Requirement Analysis for Data Warehouse Design.* Paper presented at the ACM Eighth International Workshop on Data Warehousing and OLAP (DOLAP'05), Bremen, Germany. ,2005.

[13] Schiefer, J., List, B., & Bruckner, R. M. *A holistic approach for managing requirements of data warehouse systems.* Paper presented at the Eighth Americas Conference on Information Systems (AMCIS), Dallas, Texas., 2002, 9-11 August.

[14] Baragoin, C., Marini, M., Morgan, C., Mueller, O., Perkins, A., & Yim, K. H. *Building the Operational Data Store on DB2 UDB Using IBM Data Replication, WebSphere MQ Family, and DB2 Warehouse Manager*. San Jose, California: IBM Corporation.,2001.

[15] Neighbors, J.M. *Software Construction using Components*. Technical Report 160, Department of Information and Computer Sciences, University of California, Irvine, 1980.

[16] Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.,1989.