RequirementAnalysis Approach for Universiti Utara Malaysia (UUM) Library Data Warehouse

Azman Ta'a[1], Muhammad Shahbani Abu Bakar[2], Mohd Syazwan Abdullah[3], and Haider AbdulKadhim Alwan[4]

[1,2,3,4]*School of Computing, College of Arts and Sciences, University Utara Malaysia, azman@uum.edu.my, shahbani@uum.edu.my, syazwan@uum.edu.my, mr.hayder@yahoo.com*

ABSTRACT

Requirements play a very important role in software development and forms as a backbone of any successful or failure of the IT project. Careless of requirements interpretation would increase cost and hinder the system development to satisfy user's expectation. Therefore, it is necessary to present the requirement in an understandable and meaningful way. This is achieved through a proper requirement process, which gives a complete view of data warehouse (DW) system and represents idea without having to explore into an actual system. This research aims to propose a requirement analysis approach for UUM Library DW system. Snowflake dimension model is used to present the DW structure that derived from a set of requirement lists, which was produced from the requirement analysis process. Various reports were designed and validated by the authorized librarian. Several limitations were discussed, and recommendations for future research were suggested.

Keywords: Requirement Analysis, Data Warehouse, Information Science, Business Intelligence

I   INTRODUCTION

Traditionally, data warehouse (DW)is defined as an integrated and histories collection of data that generally used to support strategic decisions by online analytical processing (OLAP)techniques (Inmon, 2002). A DW refers to a database that is maintained separately from an organization's operational databases and allow for the integration of a variety of application systems. This will support the information processing by providing a solid platform of consolidated historical data for analysis and reporting. Nowadays, most of the existing DW development is industry-based approaches. Thus, most of the methods for developing the DW are tight-up with the particular tools that are provided by the industries.

The differences between operational and DW system required difference approach to develop the system. An operational system still needs more time, more data, more processes, and others to realize the functionalities of the system. With the same effort, the

DW system can begin to be constructed. However, does the DW requirements enough gathered and selected subject's area that is capable enough to be meaningful and implemented? These issues need to be investigated into the DW requirement elicitation and analysis approaches.

In case of implementation in an academic environment such as Library system, the DW must be useful for the students, staffs, and managements. Since the library system has been supporting acquisition, cataloging, borrowing, and other tasks, the DW system must be able to provide analytical information for the benefits of these users. The questions of what is the information required by the users? How to manage the requirement by anticipate the library system, other related system and the users? This paper will answer these questions and present the works on defining the approach for eliciting and analyzing the DW requirements for library DW. Specifically, the paper will identify the information required by the library users, proposes the requirements approach for library DW system, and present the validation method to validate the DW requirements. Since this paper aims to propose an approach to analyze requirements for library DW, the focusing is to manage requirement that relates to main library tasks such as cataloging and borrowing.

The existing DW requirement approaches are based on bottom-up (i.e., user or data driven), top-down (i.e., demand or goal driven), and in-out (process, model or ontology driven). These approaches are summarized as shown in Table 1 (Ta'a, Abdullah and Norwawi, 2010).

Table 1. The DW Requirements Analysis Approaches

Researchers	Approaches
Kimball (1996)	Process-driven
Inmon (2002), Winter and Strauch (2004)	Supply-driven/ Data-driven
Winter and Strauch (2004)	Demand-Requirement driven
Niedrite et al.(2007), Giorgini et al. (2008)	Goal-driven
Mazon, Pardillo, and Trujillo	Model-driven

(2007)	
Romero and Abello (2007), Skoutas and Simitsis (2007)	Ontology-driven

Based on a current library scenario, this research will adopt the process and demand driven approaches, which focusing on the business process and requirement from the student, staff and management. There is no specific approach is used to analyze the requirements for library DW. Therefore, the findings of this research can be used to develop an analytical library DW system. Moreover, the plethora of reporting and analyzing can be produced when DW system is ready to be implemented.

II   DW REQUIREMENT ANALYSIS APPROACH

The requirement analysis approach is adapted from the Requirements Management Control (RMC) methodology by Fábio, Ana and Jaelson (2008). The RMC has three phases of tasks namely Requirements Management Planning, Requirements Specification and Requirements Validation. This approach is shown in Figure 1.
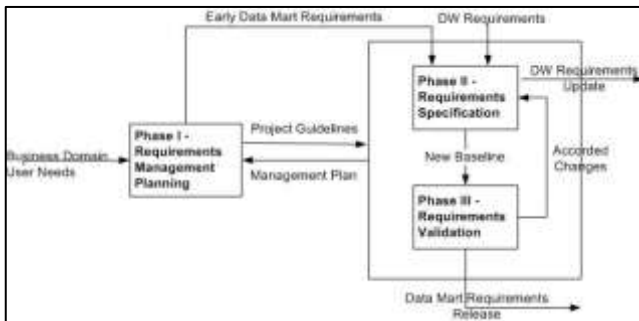

Figure 1. RMC Framework (Fábio, Ana & Jaelson, 2008).

Each phase provides the abstraction level of the DW system in depth, as the DW requirements are collected to form a requirement's baseline. All these phases are assessed by the RMC, which acts as quality assurance of the changes in DW requirements (Fábio, Ana, & Jaelson, 2008). Phase I has provided general guidelines for requirement elicitation and determined the granularity of the data mart. This has been done by interviews and exploration of the existing documentation.

Phase II aims to achieve requirement specifications. In DW development, this procedure involves a cyclic task of acquisition, representation and evaluation of requirements. Thus, an iterative process gives the impression to be more appropriate to support such tasks flow. Figure 2 refers to elicited data mart requirements through a lope of iterations, along which requirements are analyzed, negotiated with the participants, registered and conformed to a broader DW specification.
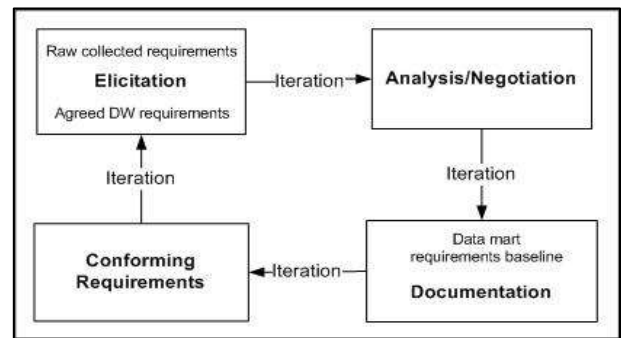

Figure 2.Requirement Specifications

After completing Phase I and II, some misunderstandings or misconceptions regarding the DW features might remain unclear. Sometimes, the DW developer is not confident enough of what information being delivered. Thus, during the review meeting, the data mart final release is presented to all involved parties, and described in terms of the multi-dimension model. This is where the validation process is done by getting feedbacks from the participants and reconfirmed the requirements as implemented in Phase III. Additionally, the requirement analysis and conceptual design are the main stages within the DW design process (Shahbani & Shiratuddin, 2011, Golfarelli, 2009).

III   CASE STUDY: UUM LIBRARY

UUM Library DW requirement is described in detail, and documented using requirement analysis approach as discussed in Section II. The requirement list of the study is presented together with the proposed DW model and reporting. However, the focus of this research is only for Cataloging and Borrowing functional. The library provides access to information resources through direct in person consultation, phone and online via the library's website. The library has 300,000 printed materials, 35 online databases, 44,000 e-books, 254,000 electronic journal titles as well as digitized in-house collections with an emphasis broadly as core business and management disciplines.

A. DW Architecture

The DW system architecture for UUM Library consists of four components: i) data sources; ii) Extract, Transform, Load (ETL) processes; iii) DW structure; and iv) DW applications. These components are closely interacted with the requirements analysis tasks to complete the entire cycle of DW system development. The requirement analysis tasks between components show the transition of library data throughout the workflows of DW processes. This is illustrated in Figure 3.
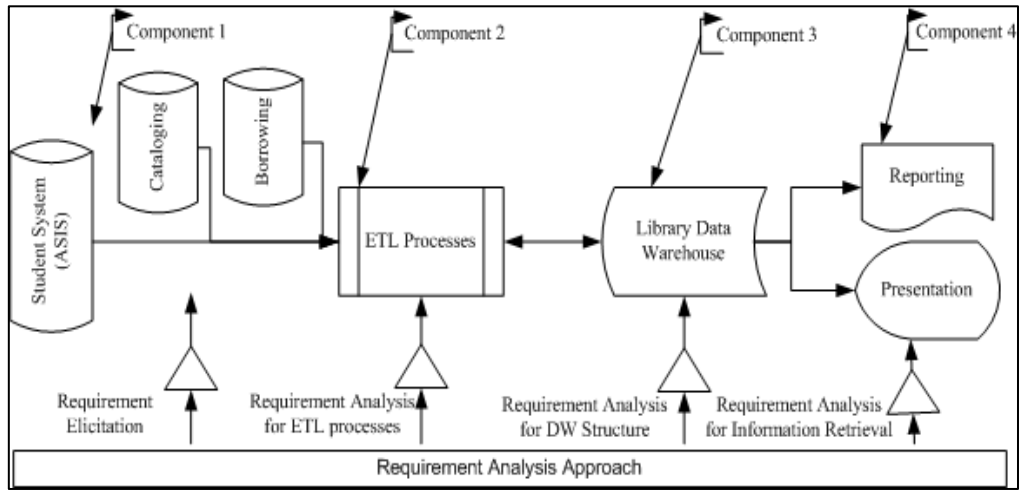
Figure 3. UUM Library DW Architecture

B. **Requirement Definition and Analysis**

Requirements for UUM Library DW was elicited and analyzed by interviewing the various levels of users. The interview session gave more insight of the real situation at UUM Library and better understanding of its functions. The interview was made with Chief of Librarian (i.e., Mr. Saleh Hudin Mustafa), and two staffs from Cataloging department. Generally, the requirements are:

1-How is the behavior of student reading?

2-What is the student study manner or interest?

3-Which kind of books should be bought?

4-What is the number of new book cataloged?

5-How many new books received? (purchase or gift).

6-How is the cataloger performances evaluated?

Haiqing (2010) stated that through analyzing of user requirements, the readers' borrowing, the reading tendency of students' reader, the utilization of library collection and the demand characteristic of literature can be mastered. Therefore, the corresponding counter measure can be taken according to various works of library staffs. The list of user requirements pertaining to the information required by the library is presented in Table 2.

Table 2. The List of User Requirements

Require-ment List	Description	Prio-rity
Reader behavior	To help students arouse reading interest. For instance, improving library conditions continuously, perfecting various serving, buying more books suitable to readers.	1
Readers choosing books	Analyze relations between the reader and the books by using association rule, some interrelatedness quite high rules are discovered.	2
Demand for books and readers	Which books' demand is higher and what kind of service can satisfy reader's need? Meanwhile the Library system is supported the reader and the books.	2
New book received	Through analyzing the kind of new books received. This is helpful to understand the number of books received.	2
Catalogers performance	To know the number of new books received for each cataloger. This help decision- maker to understand the staffs performance and effectiveness.	2
Reader study manner and interest	Through analyzing the borrowing type and the time to different major and different grade student.	1
Books lost	To identify the number of books lost for various reasons.	3

Legend: 1-Urgent, 2-Important, 3-Less Important

The results of the interviews were analyzed, and main information required by the users was classified as the following:

Books historical analysis. Through books historical data's analysis, the information about monthly circulation and instantly circulation will help managers to deeply understanding the book's circulation working rule, and grasp the different time sector of circulation work for internal structure condition of circulation department. Additionally, book losses also can be detected by analyzing the movement of books.

Readers demand analysis. It mainly analyzes books demand situation of readers, traces different services generating different effects, and forecast books demand for the period of time of the future. For

example, through analyzing the borrowing type and the time to different major and different grade students, this is helpful to understand the student's study direction, study manner, and study interest to provide assistance for the educational reform.

Books purchase analysis. Through the DW and OLAP technology, the data can be collected from the various multi-channels, supply the scientific purchasing information and provide the policy-making service for book's purchasing. For example, through the reader borrowing analysis subject, we can analyze what kind of book's reader is most interested. Analyzing the relations between the reader and the books by using association rule, some interrelated rules are discovered assistance this can be used to provide the base rules for books purchasing and simultaneously provide help for readers choosing books.

Staff performance analysis. New books received by the Catalog Department are used to helping decision-makers to understand the staffs working performance and efficiency. Moreover, this can provide a basis for fixing position and other policy-making for the library staff's career.

C. UUM Library DW Model

DW model for UUM Library is developed according to Cataloging and Borrowing functionalities. Snowflake dimension modeling technique (Kimball, 1996) is used to present the facts, dimensions and measures for the requirements as defined in Section III (B).

Data Model for Borrowing. According to the requirements of borrowing, components structure of DW model are designed based on common dimensions that describe the borrowing fact, which contain measures for total borrowers of undergraduate and post graduate students. This DW model is shown in Table 3.

Table 3. Dimensions and fields for Borrowing

Dimensions	Fields
Book	book_ID, BIB_ID, Title, ISBN, Publisher, Author
Borrower	borrower_ID, first_name, last_name, contact, email, status_ID
Borrow	borrower_ID, student_ID, book_ID, borrow_date, return_date
Borrower status	status_ID, description (post-graduate, under-graduate, staff)
Time	time_ID, year, month, day

Attribute abbreviations:ISBN-International Standard Bibliographical, BIB_ID-System Supplied Identification Number

The DW model includes five dimensions such as time, book, borrower, borrower status, and borrows. This DW is presented in a snowflake dimension model as shown in Figure 4.
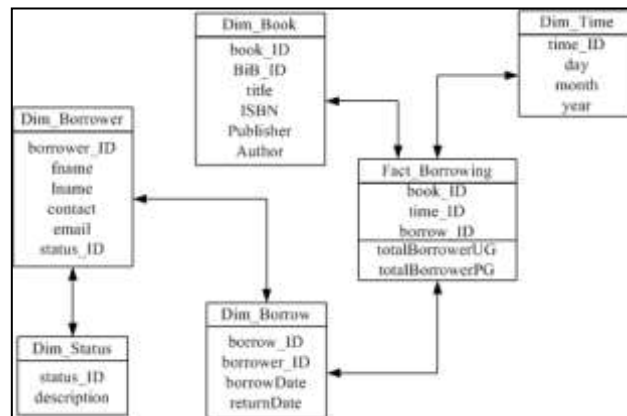

Figure 4.DW Model for Borrowing

Data Model for Cataloging. Catalog information is also built as a snowflake dimension model. The model includes four dimensions such as time, new books, type of book, and cataloger as determined in the Table 4.

Table4.Dimensions andFields for Cataloging

Dimensions	Fields
New Book	book_ID, attachment,type_ID, librarian_ID
Type_ID	type_ID, description
Librarian	librarian_ID, name, mobile, contact
Time	time_ID, year, month, day

Based on Table 4, the snowflake dimension model for Cataloging was designed as shown in Figure 5.
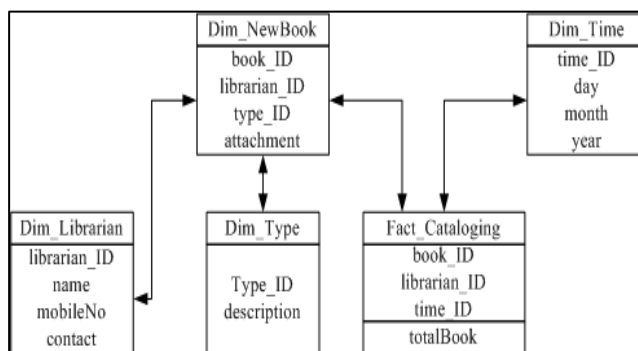

Figure 5.DW Model for Cataloging

D. Reporting Design

The design of various reporting formats for UUM Library DW is based on the DW model as discussed in Section III(C). The suggested reporting is used to validate the DW requirements to ensure the correctness of the requirements analysis and providing the relevant information as required by the users.

Report for lending book. This report will answer two questions. Firstly, how many volumes for each subject have been borrowed? Secondly, what is the special characteristic of each subject? The subject is defined as A for philosophy, C for history, and so on. Generally, the behavior of student's reading is

universality and particularity. The books should be bought based on public and specialized courses. Example of the report is shown in Table5.

Table5. Example - Report for Lending Book Categorize by Subject.

Subject Month	A	B	BE	C	E	G
January	143	207	221	894	89	562
February	351	567	456	675	948	431
March	110	84	344	311	504	112
April	0	43	79	83	220	21
May	40	23	53	45	76	27
June	17	34	0	16	54	11

Report for borrowing pattern. The borrowing pattern depends on student status (postgraduate and undergraduate). This report provides information about reader's behavior and what kinds of books are interesting. Example of the report is shown in Table6.

Table6. Example - Report for Borrowing Pattern.

Year	2009		2010		2011	
Borrower Category Subject	Post	Under	Post	Under	Post	Under
A	454	656	534	850	487	698
B	354	589	243	416	253	342
BE	123	654	564	634	324	732
C	78	374	342	453	232	653
E	134	935	745	654	521	709

Report for library utilization. The report consists of subject area, volume of books for each subject, borrowing volume, and percentage utilization. Utilization ratio of books was an important fact for acquiring what kinds of books and how many books need to be bought. Example of the report is shown in Table7.

Table7. Example - Report for Utilization of Book.

Subject	Library Collection(volume)	Borrowing (volume)	Utiliza- tion (%)
A	10234	1011	10%
B	20342	1123	18%
BE	6543	2030	3%
C	8076	220	36%
E	1239	987	1%
G	12345	1090	11%

Other reports are not presented due limited spaces. These reports are defined as *Borrowing pattern by specific book, The volume of new received books (purchases and gift),* and *Monthly cataloger performances.*

IV VALIDATION

As stated in Phase III of requirement analysis approach, the validation is performed in order to ensure the correctness of requirements according to user perspectives. The most important in validation is to examine the proposed DW models by the user who is familiar and authorized in library knowledge. The knowledge or corporate librarian should be a knowledge broker, which has enhanced the flow of knowledge in enterprise (Murray& Wheaton, 2009). Moreover, the requirements should be corrected and fulfilled the users demand at any levels. The validation process has been done with the chief of the librarian and the feedbacks are summarized in the following facts:

1- The lending status of each subject is categorized by year and subject.

2- The borrowing pattern is categorized by year, borrower category, and subject. This report supports the library management to understand the student's interest and study manner.

3- The purchasing is explains what kinds of books, how many books should be bought, and what is the utilization percentage of books.

4-The number of students (post or undergraduate) borrows a specific book.

5-The number of new books and what kind of book received (i.e., purchases, gift).

6-The performances and effectiveness of staffs are measured by number of new books cataloged by the staff.

The requirements for DW UUM Library have been successfully validated by the chief of the librarian. Adjustments have been made in the DW model based on the feedbacks given.

V RELATED WORKS

There are several benefits that can be achieved by developing a DW system in an academic environment. It provides a centralized source of information accessible across different academic units to quickly analyze problems and get acceptable solutions. Moreover, the necessary data is provided for developing the institution's strategic plan, and enabling the administrator to make better business decisions based on historical data available in legacy databases. Generally, a university DW is designed to supply information that satisfies the following needs (Carlo, Francesco, Ezio & Filippo, 2007):

1- A system for analyzing and reporting for supervisory staff, single organizational and administrative structures such as departments or secretariats for the students.

2- A system that supplies real-time data to information external agencies such as the university DW that structured on the typical multi-level layout.

Lei and Geng (2011) mentioned that building a library decision support system becomes very important with rapid development of DW system. The system has adopted three layered structure, namely data layer, DW layer and DW tool. By using this system, datasets will be analyzed to support decisions for gaining a competitive advantage. In another research, the objective of the decision system for the Kansas state University Library (KSUL)is to detect patterns for books used, and electronic services associated with costs. The management of KSUL must carefully select the books purchased, the subscription of serials, and the contracts with electronic vendor companies. An analysis of patterns can assist the KSUL management to allocate its finances more efficiently and to satisfy all the requests for books, journals and other periodicals from lecturers and students. DW requirements are the basis for all future venture activities and have a main impact on the achievement of the DW venture (Schiefer, List & Bruckner, 2002). Usually, DW is unsuccessful as an outcome of poor communication between IT and business professionals, as well as developers who have poor management skills(Kimball, Reeves, Ross & Thornthwaite, 1998). In order to achieve a successful DW system, a good of requirements engineering effort and planning is required. Therefore, the proper planning and management skill for data acquisition is necessary to gather the DW requirement successfully. A solid plan to implement the physical design of the DW should be easy to deploy and maintain in the future. However, little efforts have been made to study on requirements for library DW system. Therefore, this research was conducted to propose the solution for this issue.

VI   CONCLUSION

This research aims to propose a requirement analysis approach for DW system that specifically intends to provide requirement process guidelines for the library DW system development. Several limitations during the research have been faced such as limited time to the staff, which is difficult to arrange an appointment for eliciting the requirements properly. Moreover, more time is required to understand business process and policy of the library system. This research has adapted RMC methodology for requirement analysis model, which comprises of three phases: requirements management planning, requirements specification and requirements validating. Snowflake dimension model is used to model the DW structure and support for designing the various reporting. Finally, the requirements, DW structure, and reporting were validated by the authorized library users in order to

ensure the correctness of the DW requirement analysis approach.

REFERENCES

Carlo, A. D., Francesco, T. D., Ezio L, E., & Filippo, T. (2007). An Academic Data Warehouse. In Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications. USA:WSEAS

Fábio, S. R., Ana, C. E., & Jackson, D. C. (2008). Towards a Methodology for Requirements Analysis of Data Warehouse Systems. XVI Simpósio Brasileiro de Engenharia de Software , 150-155.

Giorgini, P., Rizzi, S., & Garzetti, M. (2008). GRAnD: A Goal-Oriented Approach to Requirement Analysis in Data Warehouses. Decision Support Systems, 45, 4-21.

Golfarelli, M., 2009. From User Requirements to Conceptual Design in Data Warehouse Design–a Survey. Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction, p.1–16.

Haiqing, L. (2010). Lending Statistics and Analysis of Readers in the Libraries of Updated Universities.In Proceedings of Information Science and Engineering (ICISE), 2010(pp. 315-319). Institute of Electrical and Electronics Engineers (IEEE), doi:10.1109/ICISE.2010.5691967

Inmon, W. H. (2002). Building the Data Warehouse - Third Edition: John Wiley & Sons, Inc.

Kimball, R. (1996). The Data Warehouse Toolkit - Practical Techniques for Building Dimensional Data Warehouses: John Wiley & Son.

Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W. (1998). Expert Methods for Designing,Developing, and Deploying Data Warehouses. canada: John Wiley & Sons.

Lei, S., & Geng, C. (2011). Design and implementation of library information decision support system. In Proceedings of Management and Service Science (MASS), 2011 International Conference (pp. 1-4). Institute of Electrical and Electronics Engineers (IEEE).Doi: 10.1109/ICMSS.2011.5998163

Mazon, J.-N., Pardillo, J., & Trujillo, J. (2007). A Model-Driven Goal-Oriented Requirement Engineering Approach for Data Warehouses. LECTURE NOTES IN COMPUTER SCIENCE(4802), 255–264.

Murray, A., & Wheaton, K. (2009). The future of the future: Rise of the knowledge librarian [Electronic Version]. KMWorld from http://www.kmworld.com/Articles/Column/Future-of-the-Future/The-future-of-the-future-Rise-of-the-knowledge-librarian--52362.aspx.

Niedrite, L., Solodovnikova, D., Treimanis, M., & Niedritis, A. (2007, February 16-19). Goal-Driven Design of a Data Warehouse-Based Business Process Analysis System Paper presented at the 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Corfu Island, Greece.

Romero, O., & Abelló, A. (2007). Automating Multidimensional Design from Ontologies. Paper presented at the DOLAP'07, Lisboa, Portugal.

Schiefer, J., List, B., & Bruckner, R. M. (2002). A holistic approach for managing requirements of data warehouse systems. In Proceedings of Eighth Americas Conference on Information Systems (pp. 77-87). Institute of Electrical and Electronics Engineers (IEEE).

Shahbani, M., & Shiratuddin, N. (2011). Conceptual Design Model Using Operational Data Store (CoDMODS) for Developing Business Intelligence Applications. IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.3, March 2011, Vol.11

Skoutas, D., & Simitsis, A. (2007). Ontology-Based Conceptual Design of ETL Processes for Both Structured and Semi-Structured Data. Semantic Web & Information Systems, 3(4), 1-24.

Ta'a, A., Abdullah, M. S., & Norwawi, N. M. (2010). RAMEPs: A Goal-Ontology Approach To Analyse The Requirements For Data Warehouse Systems. WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS, 7(2), 295-309.

Winter, R., & Strauch, B. (2004). Information Requirements Engineering for Data Warehouse Systems. Paper presented at the ACM Symposium on Applied Computing.