Extraction and Transformation of knowledge from Object Oriented Representation, O-A-V Triplets and Use Cases into Knowledge Warehouse

Developed by

Ms. Zanib Bibi
332-FAS/MSCS/F06

Supervised by:

Prof. Dr. Maqbool Uddin Shaikh

Ms. Zakia Jalil

Department of Computer Science
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad
2010

International Islamic University, Islamabad
Department of Computer Science

Dated: 19-02-2011

FINAL APPROVAL

It is certified that we have read the thesis submitted by Ms. Zanib Bibi, Reg. No. 332-FAS/MSCS/F06 and it is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad for the MS Degree in Computer Science.

COMMITTEE

External Examiner

Dr. Abdus Sattar
Former D.G.
Pakistan Computer Bureau
Islamabad.

Internal Examiner
Dr. Ali Daud
Assistant Professor
Department of Computer Science,
International Islamic University,
Islamabad.

19-02-11

Supervisor
Prof. Dr. Maqbool-Uddin Shaikh
Department Computer Science
COMSATS Institute of Information Technology
Islamabad.

Co-Supervisor
Ms. Zakia Jalil
Lecturer,
Department of Computer Science
International Islamic University,
Islamabad.

A Dissertation Submitted to the
Department of Computer Science,
Faculty of Basic and Applied Sciences,
International Islamic University, Islamabad, Pakistan,
as a partial fulfillment of the requirements for the award of the degree
of MS in Computer Science

Dedication

DEDICATED

TO

MY FAMILY

DECLARATION

I hereby declare that this thesis, neither as a whole nor as a part there of has been copied from any source. It is further declared that I have developed this thesis entirely on the basis of my personal efforts made under the sincere guidance of my supervisors. No portion of the work presented in this report has been submitted in support of any application for any other degree or qualification of this or any other university or institute of learning.

ZANIB BIBI

332-FAS/MSCS/F06

Acknowledgment

First and foremost, I'd like to thank Almighty Allah for choosing me for this work.

Secondly I would like to thank my family who is always encouraging and supporting me to complete my thesis.

Then I would like to thank my supervisor Prof. Dr. Maqbool-ud-Din Shaikh.

Project In Brief

Project Title	**Extraction and Transformation of knowledge from Object Oriented Representation, O-A-V Triplets and Use Cases into Knowledge Warehouse**
Undertaken By	**Zanib Bibi**
Supervised By:	**Prof. Dr. Maqbool-Uddin Shaikh** **Ms.Zakia Jalil**
Start Date	**10th Feb ,2010**
Completion Date	**18 Aug,2010**
Tools & Technologies	**Visual Studio 2008, Visual C# and sql server**
Documentation Tools	**Microsoft Word 2003/2007**
Operating System:	**Windows XP**

Abstract

Data warehouses are currently used as an organizational repository to support top management in business decision making. Data warehouses extract data from multiple heterogeneous and distributed data sources and store into a central repository to provide organizational view of data. The limitation of data warehouse is, it can only store and manipulate data but not knowledge. So there is a need of a knowledge warehouse that can store and manipulate knowledge instead of data. Different knowledge warehouse architectures have been proposed and discussed in literature. But none of them focus on the quality of knowledge stored in knowledge warehouse.

In this research work, an enhanced architecture of knowledge warehouse is proposed and discussed. The proposed architecture introduced a new layer called correction factor layer that maintain the quality of knowledge in knowledge warehouse. This layer performs four important functions to enhance the quality of knowledge. These functions are knowledge refinement, profiling, standardization, and matching. The author also proposed and implements three algorithms for transformation of Object oriented representations; O-A-V Triplets and use cases into directed graph.

Table of Contents

List of Figures

CHAPTER NO.1 INTRODUCTION

1.1 Motivations and Challenges

The success of every organization is based on the decisions made by decision maker. Decision maker must make effective decision on time, for survival and to get competitive advantages over competitors. Decision Support Systems (DSS) are computer systems that help organization in making decision.

The researcher acknowledges the need of automatic system that assist decision maker in solving semi-structured problems in early 1970s. Late 1970s the DSS movement started focusing on "interactive computer-based systems which help decision-makers utilize data bases and models to solve ill-structured problems"[1]. Gory and Scott-Morton introduced the phrase 'DSS' for the first time in 1971. Decision support system becomes common in 1990s. In 1980s, advancements in information technology resulted in expert systems. Expert systems replicate expert knowledge to solve problem related to narrow domain. In mid 1980s executive information system (EIS) emerged to fulfill the information need of executives. Executive information system provides executives, critical and timely information to help them in the process of decision making.

Data warehouse is database maintained separately from organization's operational databases to help organization in corporate decision making process. Data warehouse is organizational data repository. "*A data warehouse combines various data sources into a single source for end user access. End user can perform ad hoc querying, reporting, analysis, data mining and visualization of warehouse information. The goal of data warehouse is to establish a data repository that makes operational data accessible in a form that is readily acceptable for decision support and other application*" [2].

''The data warehousing concept provide model for the flow of data from operational systems to decision support environments [3] . It is best solution for decision-making

process in a business organization. Although the main reason of data warehouses is to get knowledge but currently data warehouses store and process data.

After data warehouse, knowledge warehouse is considered a new way towards decision support systems. It combines and stores the knowledge from various heterogeneous data sources like data warehouse. The only difference between data warehouse and knowledge is that the knowledge warehouse stores knowledge while data warehouse stores data.

In literature, different knowledge representation techniques exit for example frames, semantic networks, knowledge map, production rules, decision table, decision trees, scripts, use cases etc [2]. "One knowledge representation scheme may be efficient than others depending on the nature and type of problem. Thus there is a need to map knowledge from one representation to another. This mapping may give faster response and reduces computation amount" [4]. Every technique has its own pros and cons.

Knowledge management systems are also used for the same purpose of decision making. Instead of extracting knowledge from data as in the case of a data warehouse, these systems use knowledge base to provide decision support. Knowledge Management System refers to a system for managing knowledge in organizations for supporting creation, capture, storage and dissemination of information [5].

Now a need is being considered for a new generation of decision support systems by integrating knowledge management, decision support, artificial intelligence and data warehousing. These knowledge-enabled systems will provide the infrastructure for capturing, cleansing, storing, organizing, leverage and disseminate not only data and information but knowledge as well [6].

In this work, an enhanced architecture for a knowledge warehouse with correction factor layer will be proposed. Three algorithms for knowledge extraction and transformation from object oriented representation, O-A-V Triplets and Use cases will be also proposed.

1.2 Background

Currently data warehousing system are used as decision support systems to help organizations in strategic business decision making. Data warehouse focuses on data storage. It is used for providing the basic infrastructure for decision making by extracting, cleansing and storing huge amount of data. The main purpose of data warehouses is to provide knowledge but they do not store or process knowledge directly. Instead, they are designed to store and process data.

"Knowledge is one the most precious resource of an organization. Within an organization knowledge is present in various form, may be in minds of workers or in documented form. Every organization wishes to preserve and fully utilize its knowledge for decision making "[4]. Knowledge is a necessary obligation for decision making. Knowledge enables the decision makers to answer questions like what to do? When to do and how to do?

A huge majority of knowledge exists in the minds of employees in the forms of procedures, best practices that are not appropriate for loading in a data warehouse. A vast quantity of knowledge may exist in the form of documents, audio and video libraries, web documents. So why not create such warehouses that enable it to store knowledge and process other types of knowledge.

''Knowledge management is a series of processes which include knowledge creation, organization, application and sharing etc. ultimate goal of knowledge management is to provide knowledge to the right person at the right time for correct decision making ''[15].

Knowledge has become the main factor of development in modern industry. Management in organizations wants important, significant and suitable information about their businesses. All the activities of business can not run without correct, proper, and solid

know how of knowledge. Knowledge is available in a vast variety of forms and a knowledge warehouse needs to store any type of knowledge.

Data, information and knowledge should be of high quality. They are of high quality "if they are fit for their intended uses in operations, decision making and planning" [16].Correct decisions depends upon the correct data and knowledge.

Some work on a conceptual level for implementing a knowledge warehouse is done by Humid R. Nemati [6]. He has proposed a hypothetical structure of the knowledge warehouse. Anthony Raymond has suggested that'' *the data warehouse model of Extract, Transform, and Load (ETL) has a parallel in knowledge warehouse* '' [7]. As data is extracted, transformed and loaded in a data warehouse, a knowledge warehouse can be developed on the same principles for knowledge. Hong Zhang and yin Liang proposed that an existing data warehouse can be extended to create a knowledge warehouse for knowledge management [8].

1.3 Research Domain

Various knowledge representation techniques are available e.g. Frames, Semantic Network, Decision Table Decision Tree, Use Cases, Production Rules, Knowledge Map, O-A-V Triplet, and Object Oriented Representation. To collect and store knowledge from a variety of heterogeneous sources and then to create a giant warehouse is a tedious task. High data quality is also required for accurate and timely decisions. A mechanism is required to collect knowledge from heterogeneous sources to load it into knowledge warehouse with enhanced quality of data.

1.4 Proposed Approach

In this research work, solution of the above mentioned research problem is provided by designing an enhanced architecture of a Knowledge Warehouse with the correction factor layer. The proposed infrastructure stores high quality of knowledge.

Generally, knowledge in the organizations is represented in the forms like decision trees / tables, frames, logics, semantic networks, O-A-V triplets, Object Oriented Representations, knowledge map, scripts and use cases etc. There is a need to convert knowledge from one knowledge representation scheme into another. For this purpose, three algorithms are proposed to convert knowledge from O-A-V Triplets, Object Oriented Representation and Use Cases into object oriented directed graph.

1.5 Thesis Outline

The thesis is organized into seven chapters. The list of chapters along with brief description is given in the following:

Chapter 1

First chapter is about the introduction and background of the problem. The proposed solution is also discussed here.

Chapter 2

The second chapter describes literature study. The related work and its relevance and further limitations are discussed.

Chapter 3

In this chapter, problem scenario and focus of the research is presented. This chapter also discusses the main points related to proposed work.

Chapter 4

Proposed solution for the problem is discussed in this chapter. Design requirement, reference architecture and research methodology are also part of this chapter.

Chapter 5

In this chapter, deployment environment of the application is discussed. Different UML diagrams i.e. Use case, sequence and flow chart diagram related to application is presented. Further, this chapter also describes the proposed algorithms.

Chapter 6

In this chapter, general testing process, types of test, and test cases for the research work are described.

Chapter 7

In this chapter, achievements, improvements and future work are presented.

CHAPTER NO.2 LITERATURE SURVEY

2.1 Introduction

In this chapter a detailed literature review is presented. Related terms and existing architectures and processes are the part of this chapter.

2.2 Data

"Data might be thought of as the atoms of knowledge. Data are typically what we attempt to gather and measure, such as age, size, or amount" [9] . Data need to be explain in the form of information because it themselves explain very little.

"*Data items about things, events, activities and transactions are recorded classified and stored but they are not organized in the proper fashion to convey any specific meaning. Data items can be numeric, alphanumeric, figures, sounds and images* "[2].

2.3 Information

"Information is data that have been organized so that they have meanings for the recipient. We need to process data items so that the results are meaningful for an intended action or decision" [2]. Data in an organized form tells us some information. *"Information is considered to be organized and sorted data. That can be used for answering a specific question"* [9].

2.4 Knowledge

"Knowledge is an asset, and it should be enhanced like any other asset". Knowledge is a very distinct form of data and information."Knowledge may be thought of as information in use, or the set of rules and relationships that enable value added, skilled performance" [7]. In summary processed data is called information and processed information is called knowledge.

2.5 Knowledge Management

"Knowledge management is the practice of adding actionable value to information by capturing tacit knowledge and converting it to explicit knowledge; by filtering, storing, retrieving and disseminating explicit knowledge; and by creating and testing new knowledge''.[6]

"It is the process that helps organizations identify, select ,organize ,disseminate and transfer important information and expertise that are part of organization's memory and that typically reside with in the organization in an unstructured manner. The structuring of knowledge enables effective and efficient problem solving, dynamic learning, strategic planning and decision-making'' [2].

The Knowledge Spiral" for creation of new knowledge for knowledge management is given in the Figure 2.1. This knowledge spiral and all its phases are discussed in detail by Nemati in [6] and Hong Zhang in [8] and are very useful for understanding the knowledge management process.

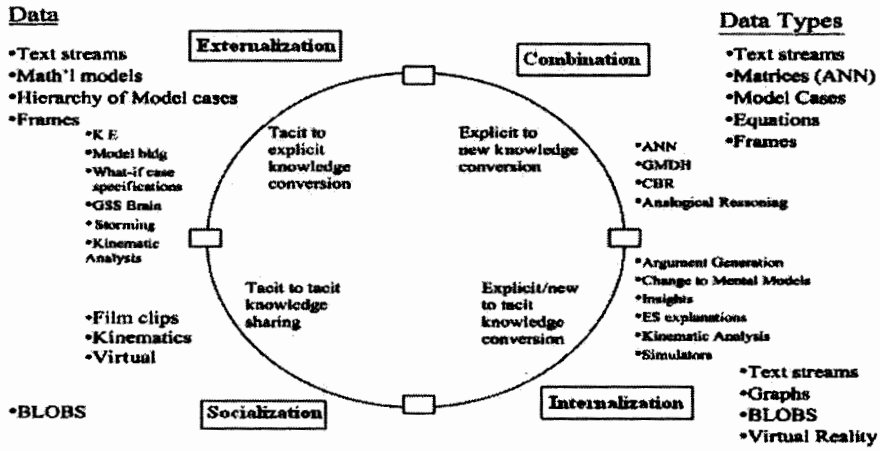

Figure 2.1: Knowledge Spinal [6]

The knowledge spiral consists of 4 steps.

A. Socialization

"*Socialization is the process of sharing the tacit knowledge such as experiences, technical skills, expertise, understanding, know-how, acumen, skill trade mental maps etc*"[10]. For example, a person can learn skill or expertise through the observation, imitating and practicing while working with his master. Information technology (IT) plays important role to share the tacit knowledge. For example, using IT a digitized film of a process once stored, can be made available on internet for any time. Now this digital film can be shared and access any time by the decision maker [6].

B. Externalization

This is the process of converting tacit knowledge to explicit knowledge. It is also called articulation [6].

"Articulation is the method of transforming tacit knowledge to explicit knowledge" [8].

The knowledge is codified into some useful form which can be processed by the computer.

C. Combination/Integration

"*In the process of integration several types of explicit knowledge are combined to create new patterns and relations*" [8]. The example of integration is given below.

"*Analysis of multiple and related 'what-if' cases of a mathematical model to find new relationships, or meta-model, that determine the key factors of the model to influence the decision can be one example of integration / combination* " [6].

D. Internalization

This is process of testing and validating the newly discovered knowledge in the proper context hence generating new tacit knowledge [6].Computers can be very useful at this stage as they can provide valuable aid for the knowledge workers to learn.

2.6 Knowledge Extraction, Transformation and Data Quality

Knowledge extraction is the discovery of the knowledge from huge amount of available data. It is concerned exploration of knowledge in the form of patterns hidden in the data base or data warehousing systems. This field is typically concerned with the data mining

process. Data mining is defined as a discovery of useful and meaningful information from huge amount of data. This process has following main tasks [11].

1. *Classification* – discovery of a predictive learning function that classifies a data item into one of several predefined classes.
2. *Regression* – discovery of a predictive learning function, which maps a data item to a real-value prediction variable.
3. *Clustering* – a common descriptive task in which one seeks to identify a finite set of categories or clusters to describe the data.
4. *Summarization* – an additional descriptive task that involves methods for finding a compact description for a set (or subset) of data.
5. *Dependency Modeling* – finding a local model that describes significant dependencies between variables or between the values of a feature in a data set or in a part of a data set.
6. *Change and Deviation Detection* – discovering the most significant changes in the data set.

"Knowledge warehouse can not only discover potential knowledge from a large quantity of information stored in data warehouse through using techniques such as data mining [12] *but also manage the knowledge assets of an enterprise'* '[8].The first part of an ETL process involves extracting the data from the heterogeneous source systems [10] , [13] .The second part of ETL process involves transforming the knowledge into a consistent format before loading [10], [13], [14].The data before loading should be free of dirty data. Transform data should not duplicate. So the data should be in high quality for better results [15] , [16].

Knowledge may acquired by fuzzy logic, decision trees [17], genetic algorithm [18] .Knowledge acquisition involves acquiring knowledge from different knowledge representations [20], [21].

2.7 Related Research Work

2.7.1. Data warehouse

Data warehouse is currently used as organizational data repository to support top management in business decision making. ''*Data warehouse provides an infrastructure that enables companies to extract, cleanse and store vast amount of corporate data from operation systems for efficient and accurate responses to user queries*" [8].

Bill Inmon defined data warehouse as "A Data Warehouse is a subject-oriented, integrated, time-variant, non volatile collection of data in support of management decisions" [27], [34]. Many papers defined data warehouse as "A set of materialized views over data sources" [29], [30], [39] .

The main goal of a knowledge warehouse is to provide an intelligent analysis platform to the decision makers that can be helpful in all stages of knowledge management [2]. H.R. Nemati has proposed architecture of the knowledge warehouse to perform all tasks of knowledge management. The purpose of this paper is to propose knowledge warehouse architecture (KW) that is an extension to the data warehouse model. This knowledge warehouse will facilitate the capturing, coding of knowledge and enhance the retrieval and sharing of knowledge across the organization. His architecture given in the Figure 2.2 consists of 6 main modules [6]:

1. knowledge acquisition module:
2. two feedback loops
3. knowledge extraction, transformation and loading module
4. knowledge storage module
5. analysis workbench
6. communication manager

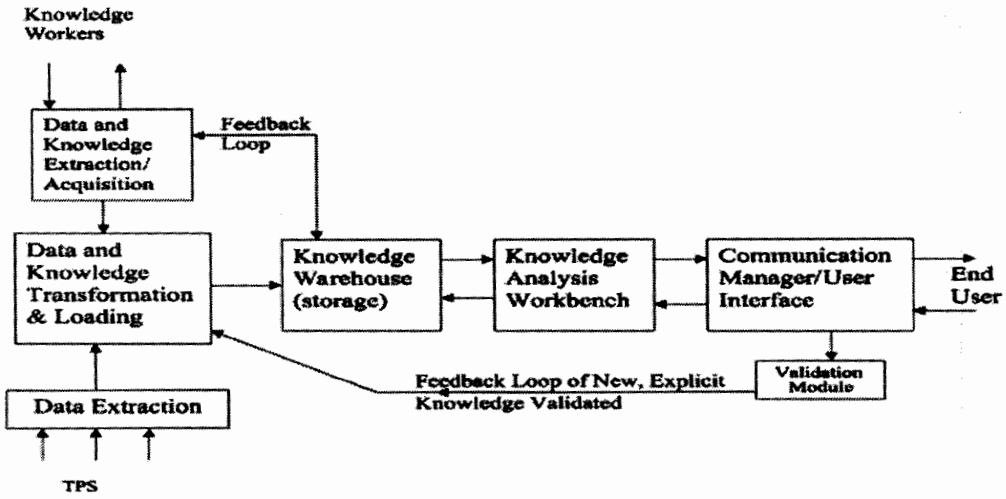

Figure 2.2: Knowledge warehouse architecture by Nemati [6]

Anthony Dymond proposed that an object oriented decision tree is the best form to store knowledge in a knowledge warehouse in [7]. He described a small example of a decision to cross the street in the form of an object oriented decision tree.

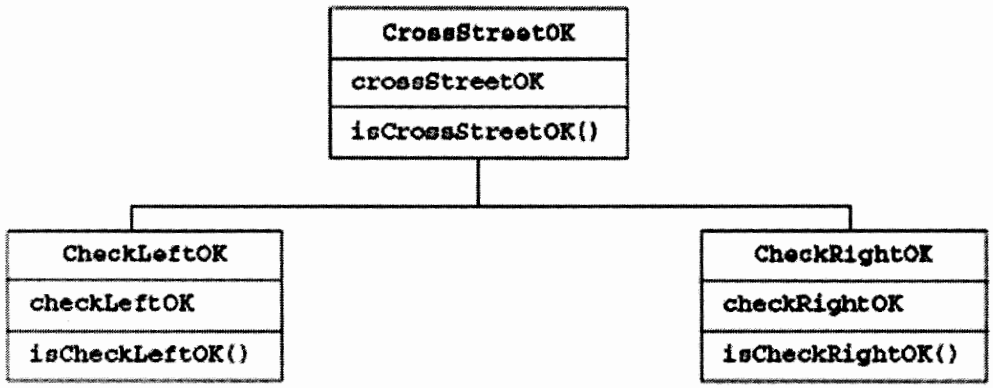

Figure 2.3: objects and tree for a knowledge base to cross the street [7]

He proposed that the methods will be activated by the way of tree search algorithms these algorithms are procedures to traverse the branches of tree [7]. The code for CrossStreeOK object will be implemented as

```
If CheckLeftOK = 'True' and CheckRightOK ='True' then
        CrossStreetOK = 'True'
End if
```

An algorithm touches the nodes and runs the methods as the "focus of attention" of the algorithm. For example, a depth first search algorithm will start at the root node and traverse the tree in the following order.

1. CrossStreetOK
2. CheckLeftOK
3. CrossStreetOK
4. CheckRightOK
5. CrossStreetOK

According to Raymond, this type of tree search supports a control structure. Assume that when the search reaches the node CheckLeftOK and it finds CheckLeftOK.CheckLeftOK = 'False'. The search next returns to CrossStreetOK where CrossStreetOK.CrossStreetOK can now be set to False. The search process can be stopped here as it is no longer necessary to check CheckRightOK. This type of executable code that will set the value of decision variable can be a nice way to implement this decision making [7].

A traditional data warehouse has the function of providing data to decision support systems (DSS) or online analytical processing systems (OLAP) but not knowledge .Hong Zhang and yin Liang in their paper proposes that an existing data warehouse can be extended to create a knowledge warehouse for knowledge management. Author described

the data warehouse architecture and knowledge warehouse architecture for ERP shown in Figure 2.3.

It consists of four major components:[8]

(1) Data sources

(2) Knowledge extraction, transformation and loading module,

(3) A knowledge warehouse module and

(4) User application module.

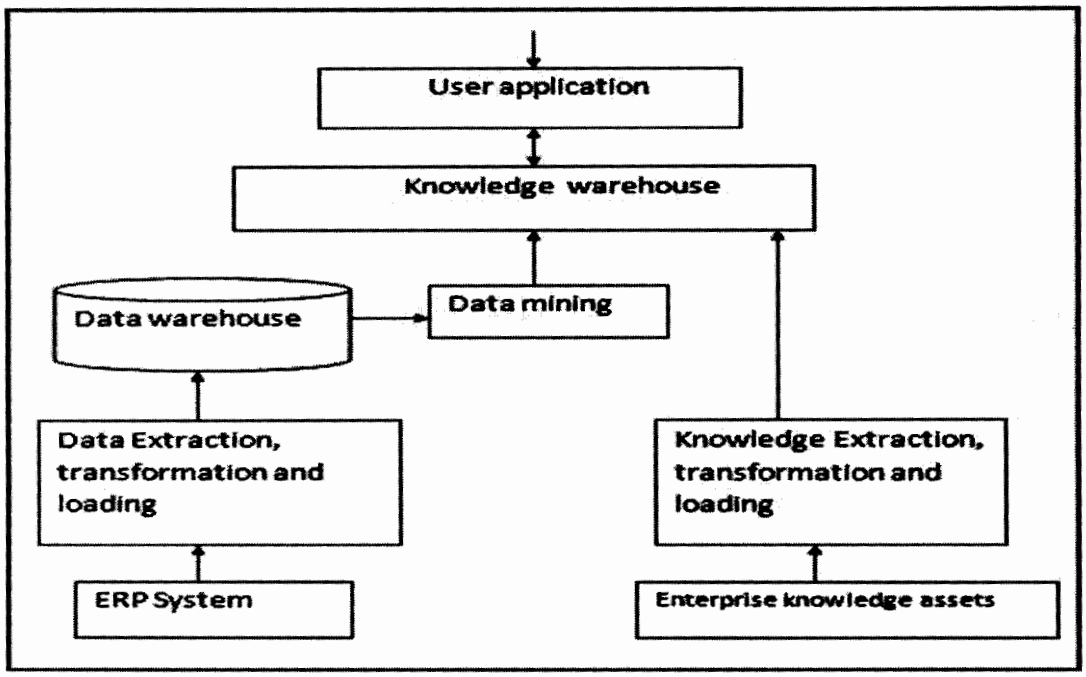

Figure 2.4: Knowledge warehouse architecture [8]

Micheal Yacci proposed another use of knowledge warehouse in the context of reusing knowledge components. This article discusses a Knowledge Warehouse, a conceptual solution to this problem. He suggests that this warehouse can be used for ad-hoc queries [9].

He proposed that knowledge can be stored in the form of knowledge components (KCs) that can be stored at different physical locations. System A and System B in the figure

show different physical storage locations. Different users can use different knowledge components in different views. He has proposed an abstract structure of the knowledge warehouse given in the Figure 2.4.

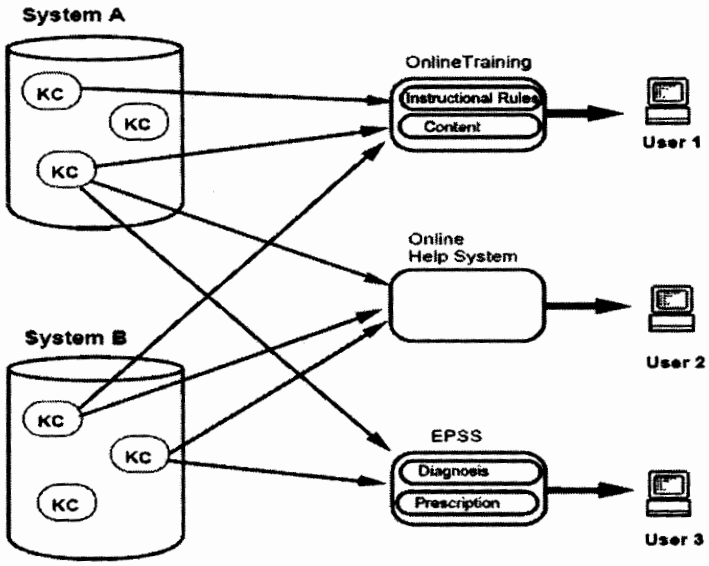

Figure 2.5: knowledge components shared across three different applications [9]

Executable code can also be very useful for performing some other tasks like taking an input from user at any stage of decision making. This technique has been used in this work during implementation of algorithms. This approach of using a decision variable and a decision function is found very helpful in this research.

Opim Salim and shahrul Azman Noah [14] proposed a design methodology for conceptual data warehouse design called the transformation orientation methodology .Which transform an Entity –Relationship (ER) model into a multidimensional model based on a series of transformation and analysis rule. This methodology is shown in Figure 2.7. A set of Synthesis and diagnosis rules will then gradually transform the problem domain model into the multidimensional model. A prototype KB tool called the DWDesigner have been developed to implement the aforementioned methodology.

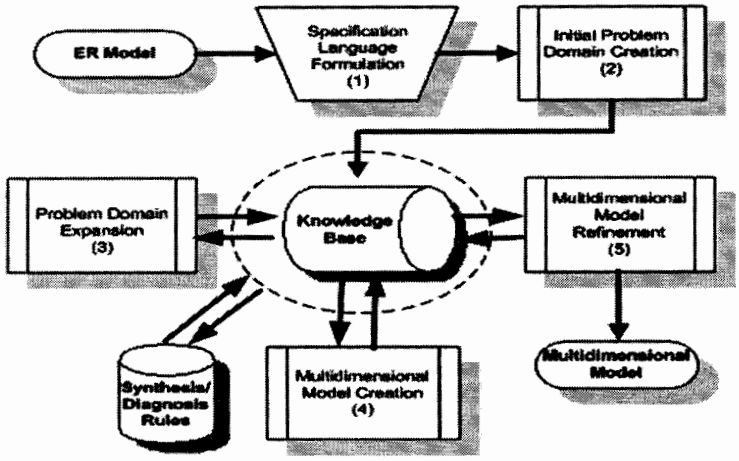

Figure 2.6: The five stages transformation oriented methodology [14]

A framework of knowledge warehouse is introduced by Tasleem Uddin [23] shown in Figure 2.7 which is enhanced form of data warehouse. It provides the platform / infrastructure to capture knowledge along with data to improve decision making in an organization. Two algorithms are also proposed .One will capture knowledge from two different sources i.e. semantic network and frames. Second algorithm will integrate and merge the acquired knowledge from two files into single object text file. Author also proposed a knowledge base storage using ES.

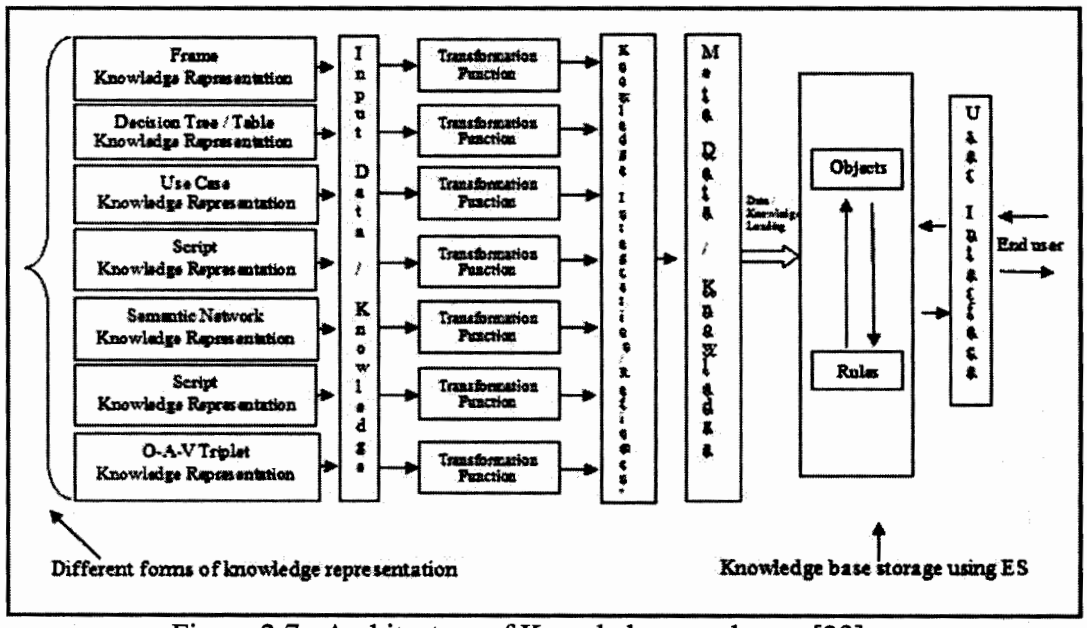

Figure 2.7: Architecture of Knowledge warehouse [23]

Ala'a H. And Soukaena Hssan [24] Proposed an alternative method to capturing and distributing of information and/or knowledge obtained from the DWH, so they called it as Knowledge warehouse (KWH). In this research the following suggestions have been proposed: grouping all the results obtained with warehouse then store and organize these results with new suggested database suitable for saving these different results. The database will be saved on new suggested server added to the traditional architecture of the warehouse to make the infrastructure of warehouse supporting the new suggested database. The suggested database will be knowledge base which stores all the results of SQL, OLAP and DM. The purpose of this suggestion is instead of extracting the results from warehouse databases by using an extraction tool (Data Mining, SQL, or OLAP), this research aims to save the time by searching the stored results of previous analysis to check if the desired analysis is extracted and stored previously.

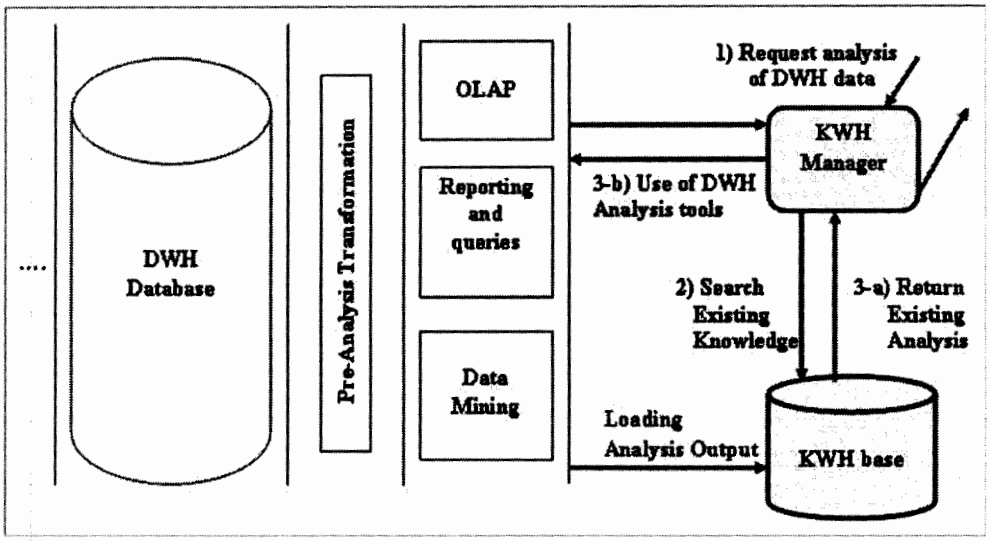

Figure 2.8: System Architecture [24]

Object Oriented Knowledge Warehouse is discussed that is inspired by the structure of human memory and data warehouse [25].

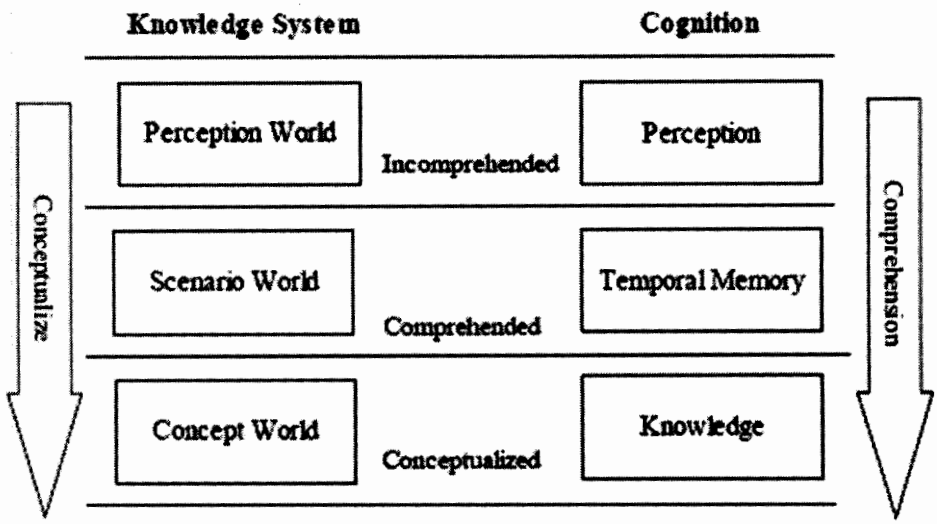

Figure 2.9: Knowledge in various stages of conceptualization and Human Comprehension [25]

2.7.2 Knowledge Extraction and Transformation for O-A-V Triplets

O-A-V Triplets provide suitable way to represent certain facts within a knowledge base and may be extended to provide the basis for the representation of heuristic rule. Each O-A-V triplets is concerned with some specific entity or object. During the literature review emphasis was on looking for existing algorithms to extract and transform data from O-A-V Triplets and store them in knowledge form. The two most relevant research works are discussed below.

Dmitry and Ari Rappoport present a novel framework for automatic extraction and approximation of numerical object attributes such as height and weight from the web[31]. This framework makes use of relation defining patterns. It has three steps that are as follows.

1- First they obtain the list of terms similar to given objects from web.

2- Retrieve attributes values for each term in this list.

3- Combine the retrieve data from all terms from list.

At the extraction of comparison information, they use directed graph in which nodes are objects and edges are correspond to extracted co-appearance of objects inside the comparison pattern. The direction of edges is determined by the comparison sign.

Marcos, François and Monique proposed a new approach for video event learning. [32].The approach incrementally aggregates the attributes and reliability information of tracked objects to learn a hierarchy of state and event concept.

Different algorithms were studied to store the O-A-V Triplets in a form which is computer readable but none of them were suitable in a knowledge warehouse perspective.

2.7.3 Knowledge Extraction and Transformation for use case

Use cases are used to document the scenarios of system usage. They are developed in the system requirement analysis phase of software engineering to describe the system usage. Use cases describe how the system will behave in response to an event from outside the system. They describe the system from the user's point of view.

An algorithm is proposed in [53]for using use-case models for reliability assessment. Reliability of a system is defined as the probability for a failure free operation of the system for a specified time with a set of operating conditions [53].Manual approaches for reliability assessment were completely dependent of the analysts and knowledge workers and hence were not accurate. They had their drawbacks specially when dealing with unfamiliar systems or very large complex systems.
The proposed algorithm discussed in [53]has three basic steps:

1- Convert use case into an SSD (System Sequence Diagram)
2- Convert SSD into a UCG (Use Case Graph)
3- Determine the reliability metric

A novel approach for prioritizing use case scenarios is proposed in [54]This prioritizing is suitable for project planning at an early phase of software development .This approach provides an analytical solution.The algorithm proposed in [54]consists of following steps.

1- Converting a use case scenario into a system sequence diagram.
2- Converting a set of *SSD* into a graphical representation.
3- Prioritizing the use case scenarios.

The graph representation in above cases is not a decision graph and discussed the graph as a sequence of messages passes which makes the nodes as states for further conversion.

2.7.4 Knowledge Extraction and transformation for OOR

Object-oriented Representation (OOR) is a paradigm that uses "objects" consisting of data fields and methods together with their interactions .It include features such as data abstraction, encapsulation, modularity, polymorphism, and inheritance.

The goal of the paper [55] is to explore automated techniques to determine what are the implicit dependencies between the lists of refactoring. Refactoring is a widely accepted technique to improve the structure of object oriented software [55].Author used AGG Graph transformation tool for this purpose.

An object-oriented model is developed and described in Unified Modeling Language (UML)[56].This model is composed of three objects, one for the jaw's physiological apparatus, one for the properties defining the mastication process and foods being chewed, and a further one for the association of the properties. With the object model, the chewing behavior is further instantiated by discovering knowledge hidden in the chewing database by data mining for association rules.

2.8 Summary

In this chapter, a detailed literature review, different terminologies related to research domain are discussed. Further, different architecture of data warehouse and knowledge warehouse are discussed in detail.

CHPATER NO.3 REQUIREMENT ANALYSIS

3.1 Introduction

Organizations have the information, data / knowledge in different forms coming from different sources such as semantic networks, frames, knowledge map, production rules, decision table, decision trees, scripts, use cases etc. Theses knowledge representation techniques have advantage and disadvantages. For example semantic networks are easy to follow hierarchy and easy to trace associations, flexile but at the same time have disadvantage like node might be ambiguous, difficult to program. Similarly production rules have simple syntax, easy to understand, flexible and highly modular but are inefficient for the large systems. It is also impossible to represent all the knowledge of organization in form of rules. Further, if a given rule is changed then there may be ripple effect.

3.2 Problem Scenarios

Data warehouse is currently used as organizational data repository. It combines and stores data that is coming from a variety of heterogeneous data sources for effective decision making. Data warehouse does not store any knowledge. The current state of data warehouse raises the issues of storing knowledge for high capability decision support [10] .

After data warehouse knowledge warehouse is considered a new way towards decision support systems. It combines and stores the knowledge from various heterogeneous data sources for effective decision making. The difference between data warehouse and knowledge is the knowledge warehouse store knowledge while data warehouse store data.

Nemati et. al. proposed an architecture of the knowledge warehouse [6]. This research work serves as a foundation for all researchers working on the knowledge warehouse. This architecture is in abstract form. Nemati et. al. explores the various concepts of a

knowledge warehouse but it is a conceptual work. This conceptual work needs to be investigated further for practical implementation. Zhang et. al. in [8] propose that a data warehouse can be used to create a knowledge warehouse . A framework of knowledge warehouse is introduced by Tasleem Uddin in [23]. Two algorithms are also proposed by Tasleem Uddin to transform and merge knowledge from semantic network and frames and store it in the knowledge base.

Different knowledge representation technique exit for example semantic networks, frames, knowledge map, production rules, decision table, decision trees, scripts, use cases etc[2]. Every technique has its own pros and cons. Organization stores their knowledge in different knowledge representation forms. Collecting and storing the enterprise knowledge from these heterogeneous representation that is coming form different sources into knowledge warehouse is a tedious task. Extract Transform and load (ETL) process is necessary to extract, transform and load knowledge into knowledge warehouse.

3.3 Focus of Research

In this research work, we are focusing on following works.

- High data quality is also required for accurate and timely decisions. The first focus is to design architecture of knowledge warehouse that stores high quality of knowledge to enhance decision making process. Design of knowledge warehouse that maintain high quality of data is the first focus of this work. For this purpose a correction factor layer will be added to the architecture of knowledge warehouse. The function of this layer is to insure the quality of knowledge before loading it into knowledge warehouse.

- The second focus is to design transformation algorithms one for Object oriented representations; one for O-A-V Triplets and one for use cases. These transformation algorithms convert O-A-V Triplets, Object oriented

representations and use cases into a unified format that is direct graph. This transformation of knowledge into a unified representation solves the problem of knowledge heterogeneity. These algorithms will serve as the guideline for knowledge ETL operations.

- Implementation of transformation algorithm is the third main focus. Implementation of transformation algorithms validates these algorithms.

3.3.1. Data Warehouse

Data warehouse currently store data. Connolly et. al. proposed three tier architecture of a data warehouse [27]. First tier consists of data warehouse and archive / backup data. Second tier consists of different data marts. Reporting, OLAP and data mining tools make third tier of the architecture. Hoffer et .al. presented generic two level architecture of data warehouse [34]. Some data warehouse architectures are discussed in [35], [36]. . Kimball et. al. have defined, "*A data warehouse is a copy of transaction data specially structured for query and analysis*" [38].

3.3.2. Extract, Transform, and Load (ETL) Process

The extraction, transformation and loading of data carried out through ETL process. ETL processes extract knowledge from heterogeneous knowledge sources, transformation of knowledge and loading it into the knowledge warehouse. The first part of an ETL process involves extracting the data from the heterogeneous source systems [41].The second part of ETL process involves transforming the knowledge into a unified format before loading [41].The data before loading should of high quality to make effective decisions [42].

3.3.3. Knowledge Warehouse

Knowledge warehouse discover potential knowledge from a large quantity of information stored in data warehouse through using techniques such as data mining [11]. It also

manages the knowledge assets of an enterprise [43]. Knowledge warehouse is a information repository which can store a variety of different views of knowledge coming from different sources, in different formats [44]. It combines knowledge from a variety of dissimilar sources to support decision maker in decision making process.

3.4 Summary

For the success of every organization updated and accurate knowledge is required. Data warehouse provides updated data to the top management for effective decision making. The next step in data warehouse is the knowledge warehouse. It is an emerging concept gaining popularity among different researchers. The basic difference between data warehouse and knowledge warehouse is that data warehouse store data and knowledge warehouse store knowledge. Different architecture of knowledge warehouse has been proposed by different researcher. But this area still needs to be mature. The focus of this research work is to design architecture of knowledge warehouse with enhanced quality of data and to transform O-A-V triple, object oriented representation and user cases into object oriented directed graph.

CHAPTER NO.4 SYSTEM DESIGN

4.1 Introduction

In this chapter we will propose and discuss an enhanced architecture of a knowledge warehouse. This architecture is based on the architecture of knowledge warehouse proposed by Hamid R. Nemati in [6] and Tasleem Uddin in [23]. A brief introduction of these reference architectures presented in this chapter. The proposed transformation functions for conversion of O-A-V Triplets, Object oriented representations and use cases into object oriented directed graph are also discussed in this chapter. Further, the research methodology used in this research work is also presented in this chapter.

4.2 Design Requirements

The following are the design requirement for the proposed architecture.

- Integrate knowledge from heterogeneous knowledge sources
- Store knowledge in consistent form
- Maintain High quality of knowledge
- Facilitate analysis
- Provide knowledge in appropriate format.
- Transform O-A-V Triplets into object oriented directed graph
- Transform Object oriented representations into object oriented directed Graph.
- Transform use cases into object oriented directed graph

4.3 Reference Architecture

The proposed architecture based on the architecture of knowledge warehouse proposed by Hamid R. Nemati in [6] and Tasleem Uddin in [23].

4.3.1. Reference Architecture 1

H.R. Nemati has proposed architecture of the knowledge warehouse in [6]. This architecture is an extension to the data warehouse architecture. Capturing knowledge, coding of knowledge and enhanced retrieval and sharing of knowledge across the organization are the major activities of this architecture. This architecture is shown in Figure 4.1. It consists of 6 main modules knowledge acquisition module, feedback loops, knowledge extraction, transformation and loading module, knowledge storage module, analysis workbench, and communication manager.

Following are the limitations of this architecture:

- It is very generic and abstract level work
- Data sources are only transaction processing system.
- How knowledge will be captured, codified, and cataloged from different sources like frames, semantic nets, decision tables etc?
- How the captured knowledge will be merged and integrated?
- Knowledge refinement process did not discuss in it.
- How knowledge will be transformed into unified format?
- What will be the knowledge storage structure?

4.3.2. Reference Architecture 2

Tasleem Uddin introduced a framework of knowledge warehouse in [23]. He proposed two algorithms, one for capture knowledge from semantic network and frames and second algorithm for the integration and merging of the acquired knowledge. Acquired knowledge is store into a knowledge base that is used by expert system for decision making.

Following are the limitation of this architecture:

- Transforms only two knowledge representation techniques.
- Load function is missing in proposed architecture.
- It does not show different transformation functions for different knowledge representation techniques.
- Knowledge Refinement Process is not discussed in detail.

4.4 Proposed Architecture

In this research work, author has proposed an enhanced architecture of knowledge warehouse with correction factor layer. This architecture is presented in layered approach. It consists of seven layers. Each layer is described below. A general view of knowledge warehouse is depicted in Figure 4.3 and detail architecture is shown in Figure 4.4.

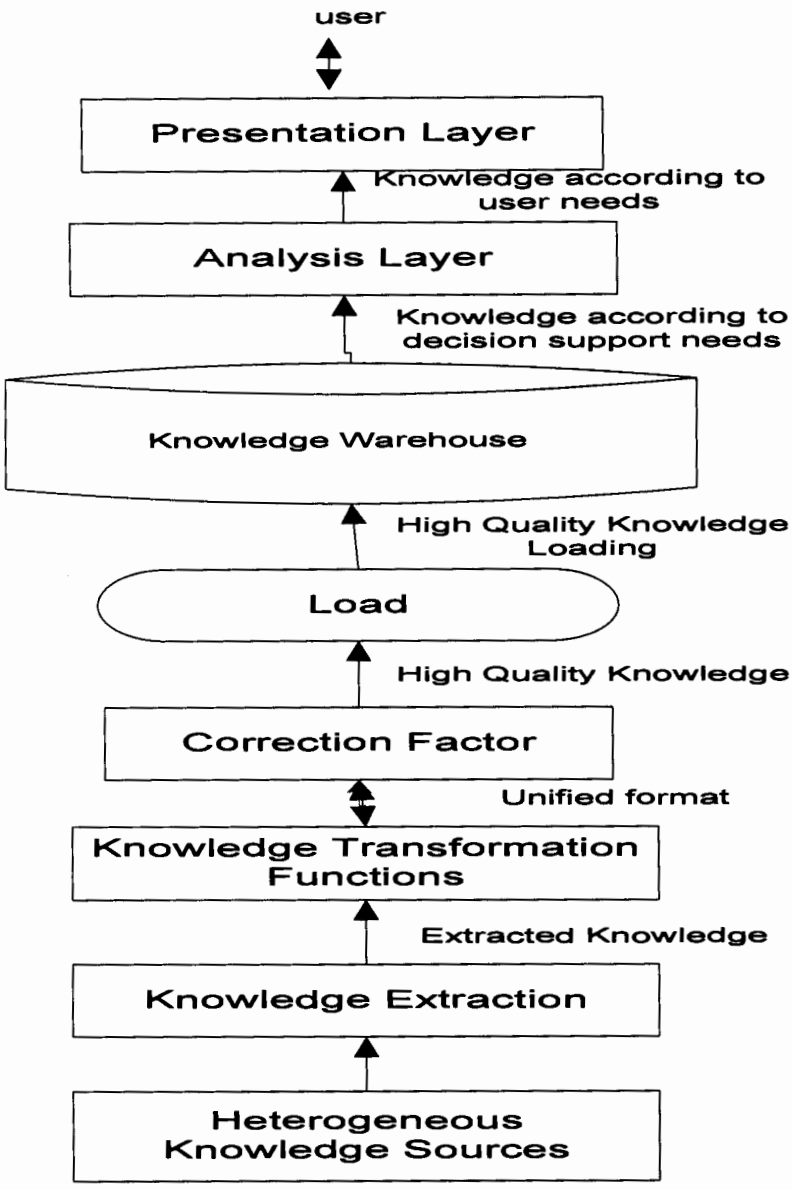

Figure 4.3: General View of Proposed Architecture

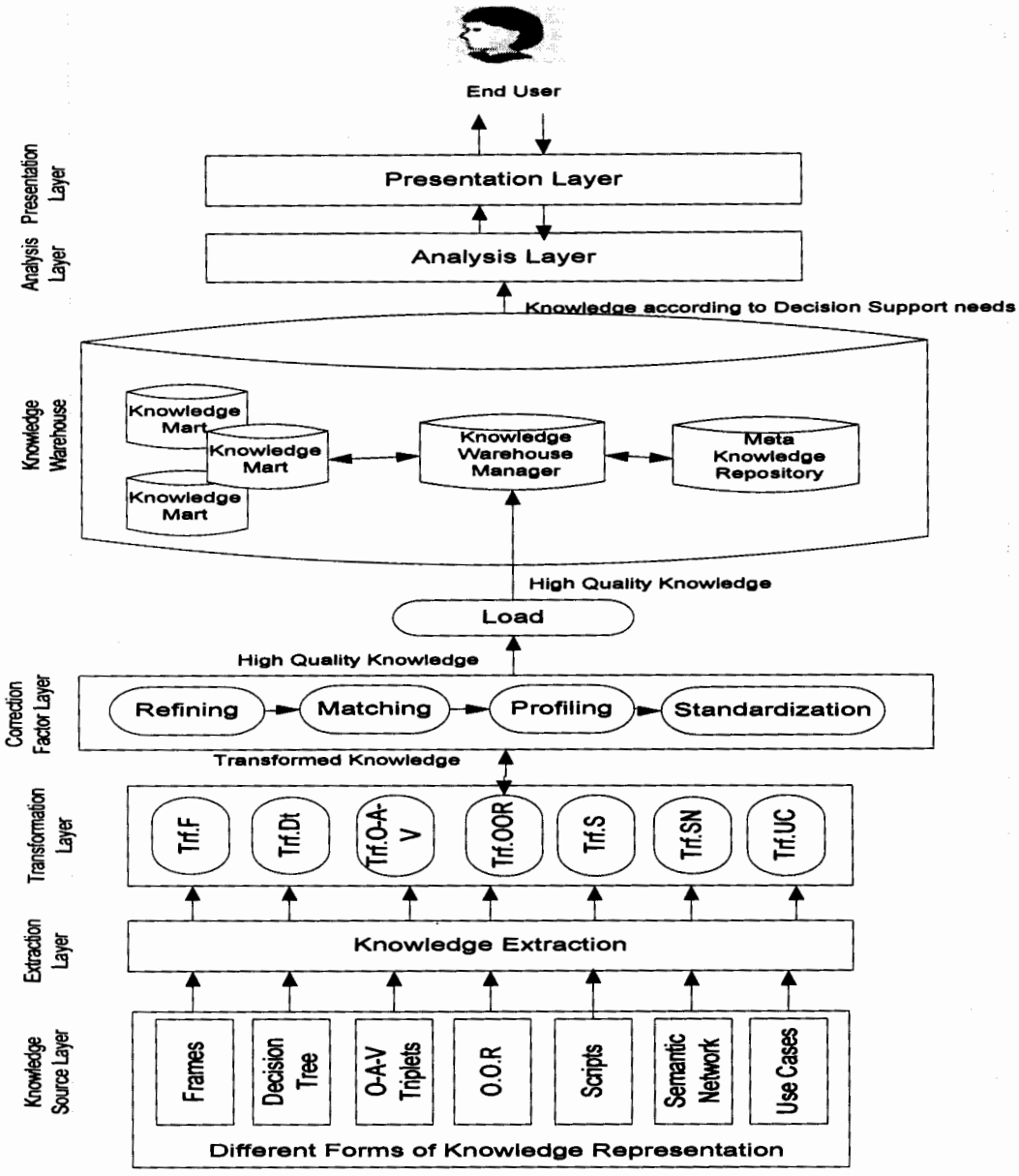

Figure 4.4: Proposed Architecture of Knowledge warehouse

4.4.1. Knowledge Source Layer

This layer consists of numerous heterogeneous knowledge sources. Knowledge extracted from these knowledge sources for loading it into a knowledge warehouse. These knowledge sources can be store knowledge in tacit or explicit form.

4.4.2. Extraction Layer

This module is responsible for extracting the knowledge from different sources. It provides the extracted knowledge to transformation layer to convert it into a unified format.

4.4.3. Transformation Layer

The basic purpose of this module is to get knowledge from extraction layer and convert it into a unified format by using different transformation function. It consists of seven transformation function, one for each knowledge representation technique.

4.4.4. Correction Factor Layer

Accurate and timely decision depends on updated and high quality knowledge. The author introduced this layer into the knowledge warehouse architecture to maintain high quality data. This layer will perform four processes before loading the knowledge into warehouse. Each process is discussed below.

- **Refinement**

First process of this layer is refinement that refines the extracted and transformed knowledge. Knowledge refinement means identifying and removing errors, inconsistencies, impurities, and duplicate knowledge. Knowledge refinement improves the quality of knowledge.

- **Matching**

This process matches or compares the knowledge with knowledge presented in the knowledge source to insure its correctness.

- **Profiling**

After refining and matching, knowledge profiling is performed of the layer. This process is used for knowledge analysis. It analyzes knowledge and collects statistics and information about that knowledge.

- **Standardization**

Last process of the layer is standardization. The process of achieving agreement on common knowledge definitions, representation, and structures to which all data layers must conform is called standardization. The job of this process is knowledge standardization.

4.4.5. Loader

Loader takes high quality of knowledge from correction factor layer and load in into the warehouse.

4.4.6. Knowledge warehouse

Knowledge warehouse is the repository of high quality knowledge that is used for decision making purpose. Knowledge warehouse consists of following parts.

a. Knowledge Marts

Knowledge related to specific business domain (area), is stored in knowledge marts. e. g. customer, products etc

b. Meta Knowledge Repository

Meta knowledge is the knowledge about knowledge.

c. Knowledge Warehouse Manager

Knowledge warehouse manager is responsible for the management of knowledge warehouse.

4.4.7. Analysis Layer

Analysis layer performs the knowledge analysis tasks. It gets user query from presentation layer and forwards it to knowledge warehouse to take the answer of user query.

4.4.8. Presentation Layer

This layer gets and forwards user query to analysis layer. It also collects result of user query from analysis layer and provides the result in suitable format.

An object oriented directed graph structure is also proposed. We will propose three algorithms, one for Object oriented representations; one for O-A-V Triplets, and one for use cases. These algorithms will extract and transform the knowledge from these three representations and make them in a consistent form before loading into a knowledge warehouse.

4.5 Knowledge Storage Structure

Storage module is a primary component of a knowledge warehouse. We are proposing object oriented directed graph structure with production rules. This knowledge base will be able to handle a wide variety of knowledge objects like data streams, relations, movie clips, animations, sounds etc. meta knowledge handling capability is also very important.

Main advantages of storing the knowledge in object form are

1. Meta knowledge can be stored with the knowledge. Knowledge source information and its analysis techniques can be stored together in an object providing encapsulation.
2. Executable routines for processing the knowledge can also be stored in the same object.

3 Object oriented concepts like inheritance, polymorphism, method overloading etc can be used to efficiently store and process the knowledge in the form of super-class and sub-class relationship.

4.6 Research Methodology

"Research is an organized, systematic, data based, critical, objective, scientific inquiry or investigation into a specific problem undertaken with the purpose of finding answers or solutions to it" [45]. The purpose of the research is to find out some fact, improvement of some practice, explaining why and how an event occurred, making some prediction, and looking for patterns [46]. There are many approaches or research methods to dealing with research depending on the size and nature of the problem. Basic categories of research approaches are quantitative and qualitative.

4.6.1 Qualitative Approach

Qualitative approach is exploratory in nature. "Qualitative researchers study things in their natural settings, attempting to make sense of or interpret phenomenon in terms of the meanings people bring to them" [47]. Qualitative research involves interpretative and naturalistic approach to its subject matter [48]. Figure 4.5 shows types of research methods.

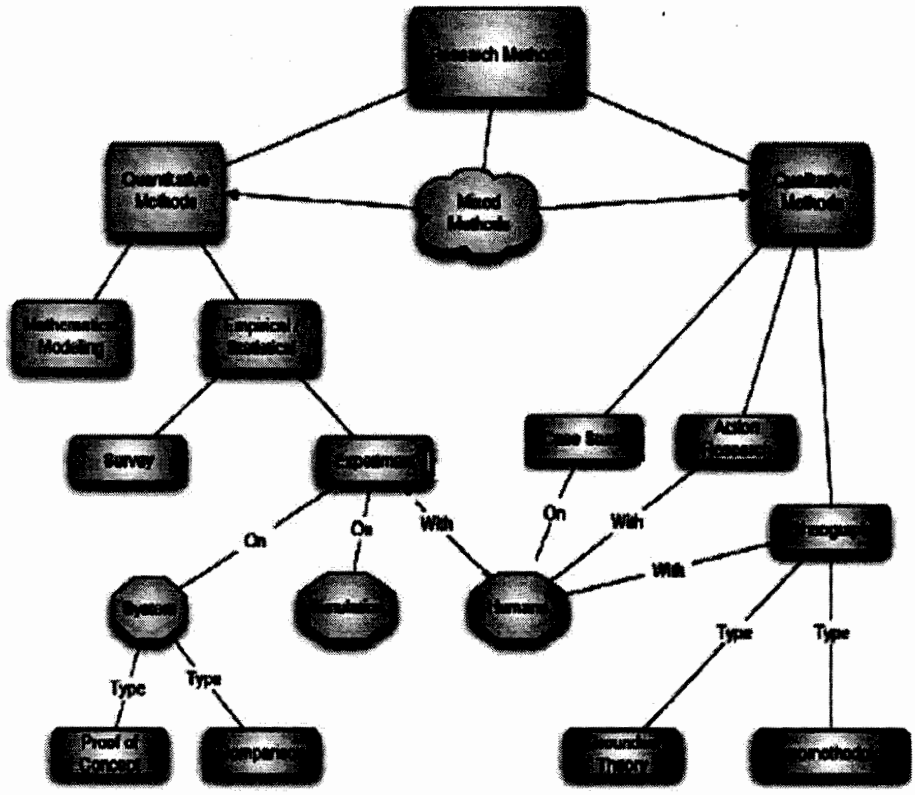

Figure 4.5: Types of Research Methods [49]

4.6.2 Quantitative Approach

A quantitative approach is one in which the investigator primarily uses post positivist claims for developing knowledge (i.e. cause and effect thinking, reduction to specific variables and hypotheses and questions, use of measurement and observation, and the test of theories) [49]. It involves collection and analyzing data that can be measured. Quantitative approach is descriptive in nature.

4.6.3 Research Approach Used

Qualitative research approach is used in this research work. A number of articles from different conferences and journals, many books and websites are consulted to understand

problem areas. A detailed literature review has been done. After detailed literature review architecture of a knowledge warehouse with improved knowledge quality is proposed and designed in this research work.

4.7 Summary

In this chapter, a new architecture of knowledge warehouse has been proposed and designed. This architecture is presented in layered approach. Knowledge from heterogeneous knowledge sources extracted transformed and load into the knowledge warehouse. A correction factor layer is introduced that maintain the quality of knowledge in knowledge warehouse. Three algorithms are proposed that transforms OOR, O-A-V and use case into object oriented directed graph. This chapter also discussed the research methodology used in the research work.

CHAPTER NO.5 IMPLEMENTATION

5.1. Introduction

In this chapter, deployment environment of the application is discussed. Different UML diagrams i.e. Use case, sequence and flow chart diagram related to application are presented. Further, this chapter also describes the proposed algorithms.

5.2. Deployment/Environment

5.2.1. Implementation Tools

The following tools are used to develop the application.

- **Visual Studio 2008**

In computing, Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It can be used to develop console and graphical user interface applications along with Windows Forms applications, web sites, web applications, and web services in both native code together with managed code for all platforms supported by Microsoft Windows, Windows Mobile, Windows CE, .NET Framework, .NET Compact Framework and Microsoft Silverlight.

- **Visual C#**

C# (pronounced see sharp) is a multi-paradigm programming language encompassing imperative, declarative, functional, generic, object-oriented (class-based), and component-oriented programming disciplines. It was developed by Microsoft within the .NET initiative and later approved as a standard by Ecma (ECMA-334) and ISO (ISO/IEC 23270). C# is one of the programming languages designed for the Common Language Infrastructure. C# is intended to be a simple, modern, general-purpose, object-

oriented programming language. Its development team is led by Anders Hejlsberg. The most recent version is C# 4.0, which was released on April 12, 2010.

- **Microsoft SQL Server**

MS SQL Server is a relational model database server produced by Microsoft. Its primary query languages are T-SQL and ANSI SQL.

5.2.2. Operating System

The operating system for the application can be MS Windows 2000/XP.

5.2.3. Hardware Requirement

Pentium "III" and Pentium "IV" with 512 MB RAM and 2.8 MHz processor is the hardware required to run the application.

5.3. Sequence Diagram

The sequence diagram of the application is presented below in the Figure 5.1.

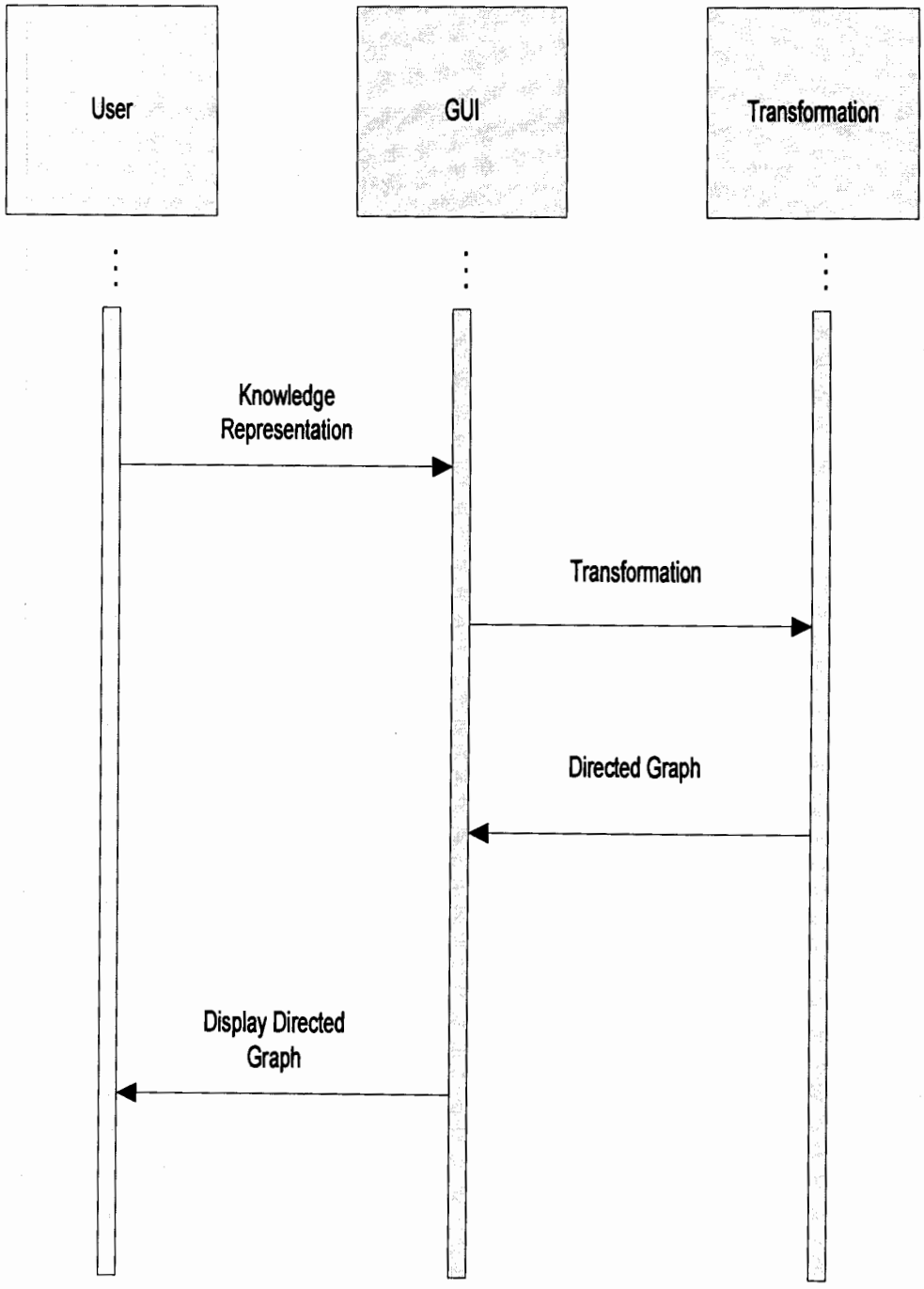

Figure 5.1: Sequence Diagram

5.4. Use Case Diagram

The use case diagram of the application is presented below in the Figure 5.2.

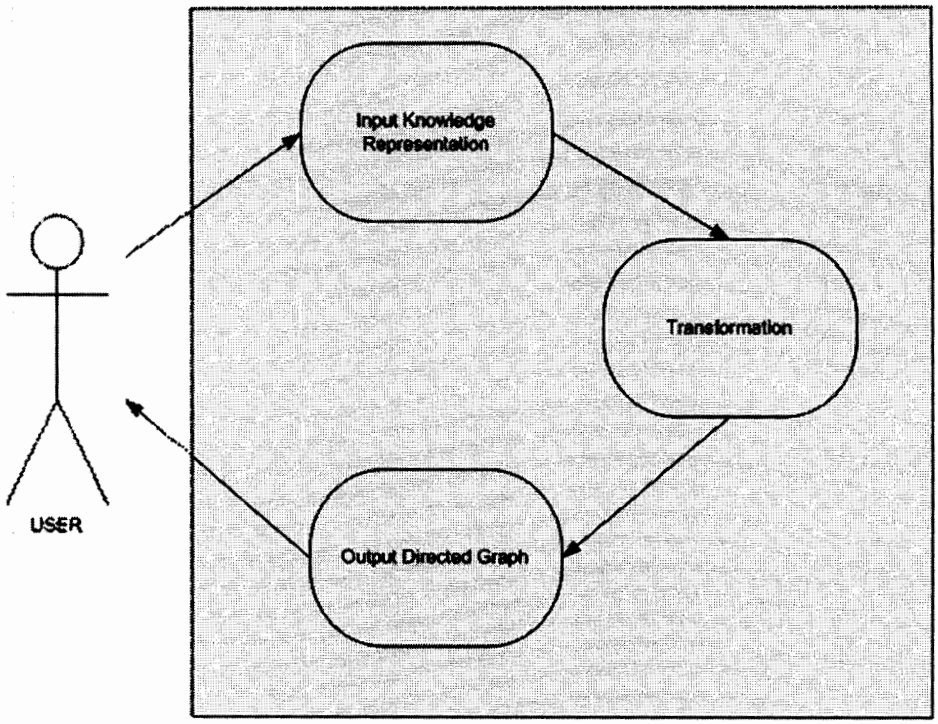

Figure 5.2: Use Case Diagram

5.5. Flowchart Diagram

The flowchart of the application is presented below in Figure

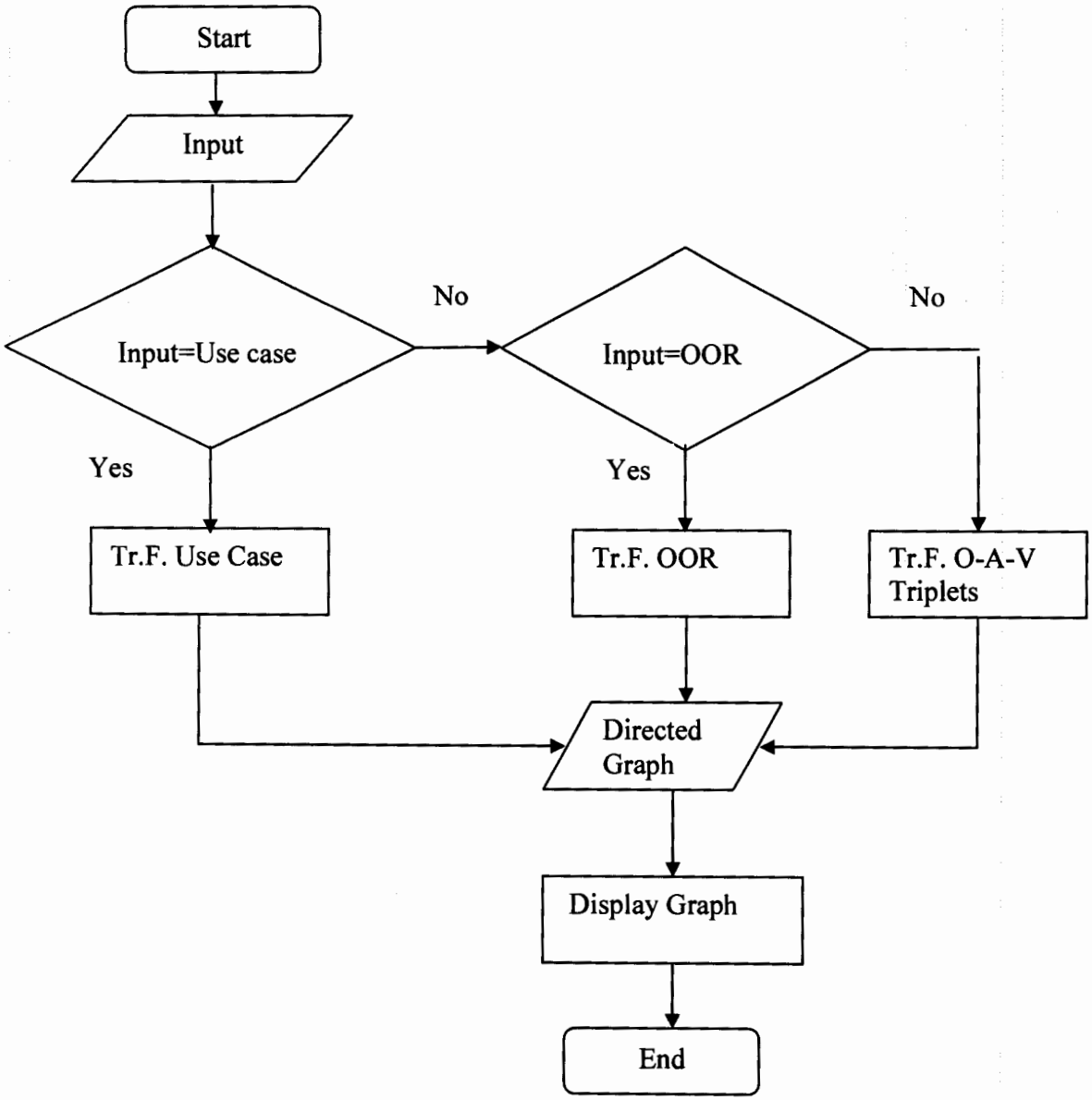

Figure 5.3: Flowchart

5.6.Proposed Algorithm

The following three algorithms are proposed to transform O-A-V Triplets, Object oriented representations, and use cases into object oriented directed graph.

5.6.1. Algorithm for O-O-R

The algorithm for transformation of objected oriented representation into directed graph is given below.

```
Create Graph graph
Int length;
For 0 to length
began
        Create edge (parentclass)
        Add label to edge (parentclass _attribute.name)
        Node c findNode(parentclass)
        c.addNode Attribute(name)
        create node c2 Node(sub_class,parentclass)
        Add label to node c2 kind of
        C2.addNodeAttribute(sub_class_AttributeName)
        C2.shapecircle
end
GraphView.addgraph(graph)
```

5.6.2. Algorithm for O-A-V Triplets

The algorithm for transformation of O-A-V Triplets into directed graph is given below.

```
Create Graph graph
Allclassess total number of classes
Tempcurrent_class;
Relation
For i=0 to allclassess
            began
            Creae node n  class_name[i]
            For j=o to all_attribute[class_name[i]]
                Began
                Create edge c create edge(class_name[i],class_attribute[j])
                c.LabelText consists;
```

```
        end
    relationdiscover_ralation(temp,class_name[i])
    create edge e create edge(temp,class_name[i])
    e.LabelText ralation;
    End
GraphView.addgraph(graph)
```

5.6.3. Algorithm for Use Case

The algorithm for transformation of Use Case Triplets into directed graph is given below.

```
Create Graph graph
 Int length total actors
    For 0 to length
    begain
    Create node actor actor_name[i]
    End
    For j 0 to total relations
    Began
    Node a=current_node
        For k=0; to current_node_all_relations; i++
        Began
        Create edge c(current_node,related_node)
        Add label to edge c  Label Text;
        End
    End

GraphView.addgraph(graph)
```

5.7. Implementation

Three different case studies are used for three different knowledge representations that are O-A-V Triplets , Use Case and OOR.These case studies are as follows.

5.7.1. O-A-V Triplets

The following O-A-V Triplets is the input to the application shown in table 5.1.

Object	Attribute	Values
Table	Color	Green
Chair	Leg	4
Bed	Leg	4

Table 5.1: O-A-V Representation (input of the application)

5.7.2 Object Oriented Representation

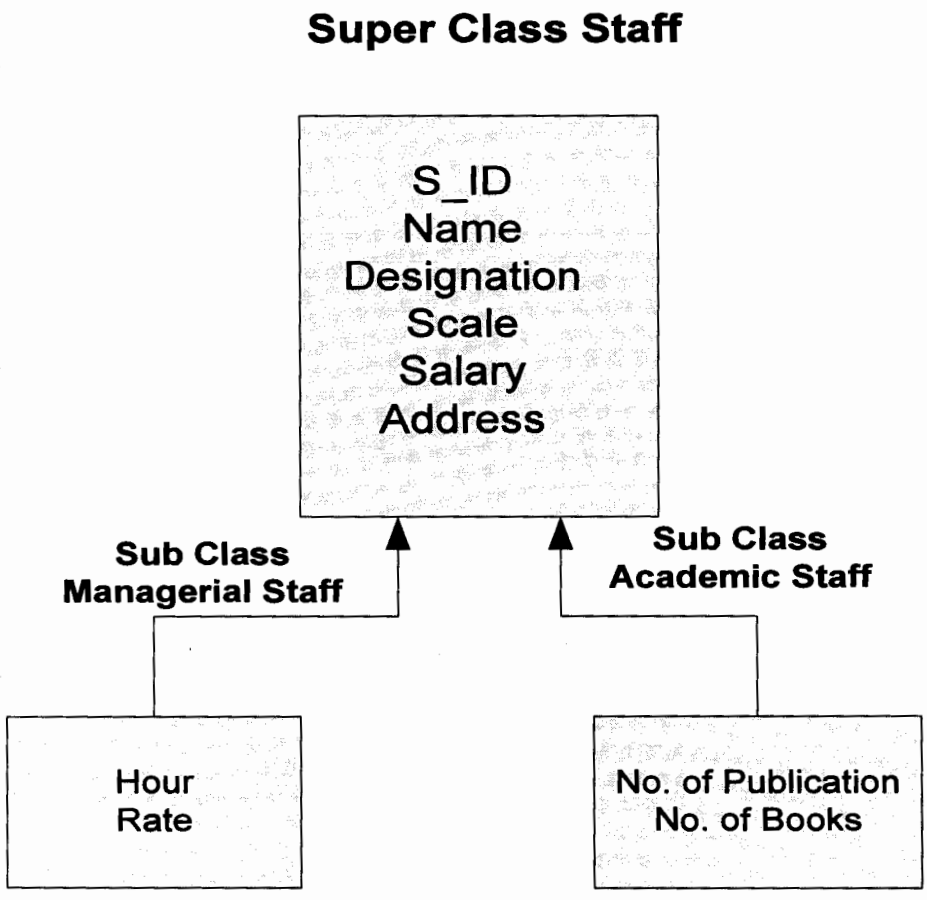

Figure 5.4:O.O.R (input of application)

5.7.3 Use Case

The following use case is input to the application .

Use case

Actors:
Customer, Sales Clerk, Inventory System, UPC reader

Pre Condition:
Customer wants to purchase some items.

Post Condition:
Customer purchased items.

Description:

Customer sets items. Sales clerk swipes UPC reader and check the prices of items in inventory. System generates bill.

1. Customer sets items
2. Sales clerk swipes UPC reader.
3. UPC reader checks the prices of items in inventory.
4. Inventory system generates the receipt.
5. Sales clerk provides receipt (bill) to customer.

5.8.Summary

This chapter presented sequence diagram, use case diagram and flow chart diagram along with three case studies, one for object oriented representation, one for O-A-V Triplets and one for use case. Three transformation algorithms, one for object oriented representations; one for O-A-V Triplets and one for use cases are also discussed in this chapter.

CHAPTER NO.6 TESTING AND PERFORMANCE

6.1. Introduction

The overall objective of the testing process is to identify the maximum number of errors in the code with minimum effort. The objective of this chapter is to describe test cases for the application to find potential bugs. This chapter covers general testing process, types of test, and test cases for the application.

6.2. Testing process

Testing of software consists of a number of test cases. These test cases test different aspect of software. Each test case includes what to do, what data to use, and what results to expect.

6.3. General types of Errors

The following are general types of errors:
Functional error (e.g. Application is not working properly)
Non functional error (e.g. performance is slow)
Logical error (e.g. errors in algorithms, use interface)

6.4. Types of Testing

Different types of testing strategies are discussed below:

6.4.1. Unit Testing

Unit testing tests the software components of module. Unit test is white box testing. In this testing important control transfers are tested to uncover errors within a module. This test can be performed in parallel to test multiple modules at a time. Unit testing checked that all the individual components were working properly, before integration of the entire components. Unit testing is essential because it gives a confidence that all components individually are working fine and ready to be integrated with other ones.

6.4.2. Integration Testing

Integration testing is a systematic technique for constructing the program structure while at the same time conduction test to uncover error associated with interfacing. The objective is to take unit tested components and build a program structure that has been dictated by design.

6.4.3. Regression Testing

In regression testing the software was tested against the boundary conditions. Various input fields were tested against abnormal values and it was tested that the software does not behave abnormally at any time.

6.4.4. Acceptance Testing

In acceptance testing the software was checked for completeness. Normally, the quality assurance department performs the acceptance testing to check is software ready and can be exported.

6.4.5. System Testing

System testing is a series of different test whose primary purpose is to fully exercise the computer based system. Although each test has a different purpose, all work to verify that system elements have been properly integrate and perform allocated functions.

6.5. Test Cases

Test Case No: TC- 01			Date 8th August , 2010	
Description: System takes O-A-V triple and save it.			Tester: Zanib	
Serial No.	Scenario	Expected Output	Actual Output	Comments

1.	User enter the object	Object saved	Object has been saved	Handled Properly
2.	User enter the attributes	Attributes saved	Attributes has been saved	Handled Properly
3.	User enter the value	Value saved	Value has been saved	Handled Properly

Test Case No: TC- 02			Date 8th August , 2010	
Description: deletion of O-A-V triple			Tester: Zanib	
Serial No.	Scenario	Expected Output	Actual Output	Comments
1.	User delete O-A-V triple	O-A-V triple deleted	O-A-V triple has been deleted	Handled Properly

Test Case No: TC- 03			Date 8th August , 2010	
Description: System takes OOR and save it.			Tester: Zanib	
Serial No.	Scenario	Expected Output	Actual Output	Comments
1.	User enter the attributes of parent class	Attributes saved	Attributes has been saved	Handled Properly

2.	User enter the attributes of sub class	Attributes saved	Attributes has been saved	Handled Properly

Test Case No: TC- 04			Date 8th August , 2010	
Description: deletion of OOR			Tester: Zanib	
Serial No.	Scenario	Expected Output	Actual Output	Comments
1.	User delete OOR	OOR triple deleted	OOR triple has been deleted	Handled Properly

Test Case No: TC- 05			Date 8th August , 2010	
Description: System takes use case and save it.			Tester: Zanib	
Serial No.	Scenario	Expected Output	Actual Output	Comments
1.	User actors	Actors saved	Actors has been saved	Handled Properly
2.	User enter relationships between actors	relationships between actors saved	relationships between actors has been saved	Handled Properly

Test Case No: TC- 06			Date 8th August , 2010	
Description: deletion of Use Case			Tester: Zanib	
Serial No.	Scenario	Expected Output	Actual Output	Comments
1.	User delete use case	use case deleted	use case has been deleted	Handled Properly

6.6. Performance and Evaluation

In this research work, an enhanced architecture of knowledge warehouse is proposed and discussed. The existing architectures did not maintain the high quality of knowledge in knowledge warehouse. The proposed architecture introduced a new layer called correction factor layer that maintain the quality of knowledge in knowledge warehouse. This layer performs four important functions to enhance the quality of knowledge. These functions are knowledge refinement, profiling, standardization, and matching. The author also proposed and implements three algorithms for transformation of Object oriented representations; O-A-V Triplets and use cases into directed graph.

We looked previously at the tree knowledge storage, which provides a useful way of storing knowledge for efficient searching. In a tree, each node can have up to two child nodes. All tree structures are hierarchical. This means that each node can only have one parent node. Trees can be used to store data which has a definite hierarchy e.g. a family tree or a computer files system. Some data need to have connections between items which do not fit into a hierarchy like this. Graph data structures can be useful in these situations. A graph consists of a number of data items, each of which is called a vertex. Any vertex may be connected to any other, and these connections are called edges. Graph knowledge storage structure has more understanding, greater representations and more flexible than binary tree. There are no stick rules of parent and child nodes. Iteration and loops make the sense more clear.

6.7. Summary

This chapter presented sequence diagram, use case diagram and flow chart diagram along with the deployment environment of the application. Three transformation algorithms, one for object oriented representations; one for O-A-V Triplets and one for use cases is also discussed in this chapter.

CHAPTER NO.7 CONCLUSION

7.1. Introduction

In these chapter achievements, improvements and future work is presented.

7.2. Achievements

The following are the achievements of this research work

- Integrate knowledge from heterogeneous knowledge sources
- Store knowledge in consistent form
- Maintain High quality of knowledge
- Facilitate analysis
- Provide knowledge in appropriate format.
- Transformation of O-A-V Triplets into object oriented directed graph
- Transformation of OOR into object oriented directed graph
- Transformation of use cases into object oriented directed graph
- Propose object oriented directed graph storage structure.

7.3. Improvements

Different architecture of knowledge warehouse has been proposed in literature by different researchers as discussed in chapter 2. Existing architecture of knowledge warehouse does not maintain quality of knowledge. In this research work, a new architecture of knowledge warehouse is proposed that maintain high quality of knowledge to support strategic business decisions. In addition, three algorithms have been proposed that transformed knowledge from O-A-V Triplets; object oriented representations and use cases into object oriented directed graph. The implementation of proposed algorithms is also done to validate them.Tree is less flexible in its structure.

Object oriented directed graph is more understandable and has greater flexibility than binary tree.

7.4. Future Recommendations

Many areas of the proposed knowledge warehouse architecture need to be explored in detail. Due to the time limitations, implementation of proposed architecture has not yet been carried out. During implementation many interesting topics will appear. Implementation of this architecture is the part of future work.

7.5. Summary

This chapter, achievement of the research work i.e. transformation of three knowledge scheme into object oriented directed graph is discussed. Improvements made in existing architecture and future work is also part of this chapter.

Appendices

Appendix A

User Manual

The following is the screen shot of first window of application.

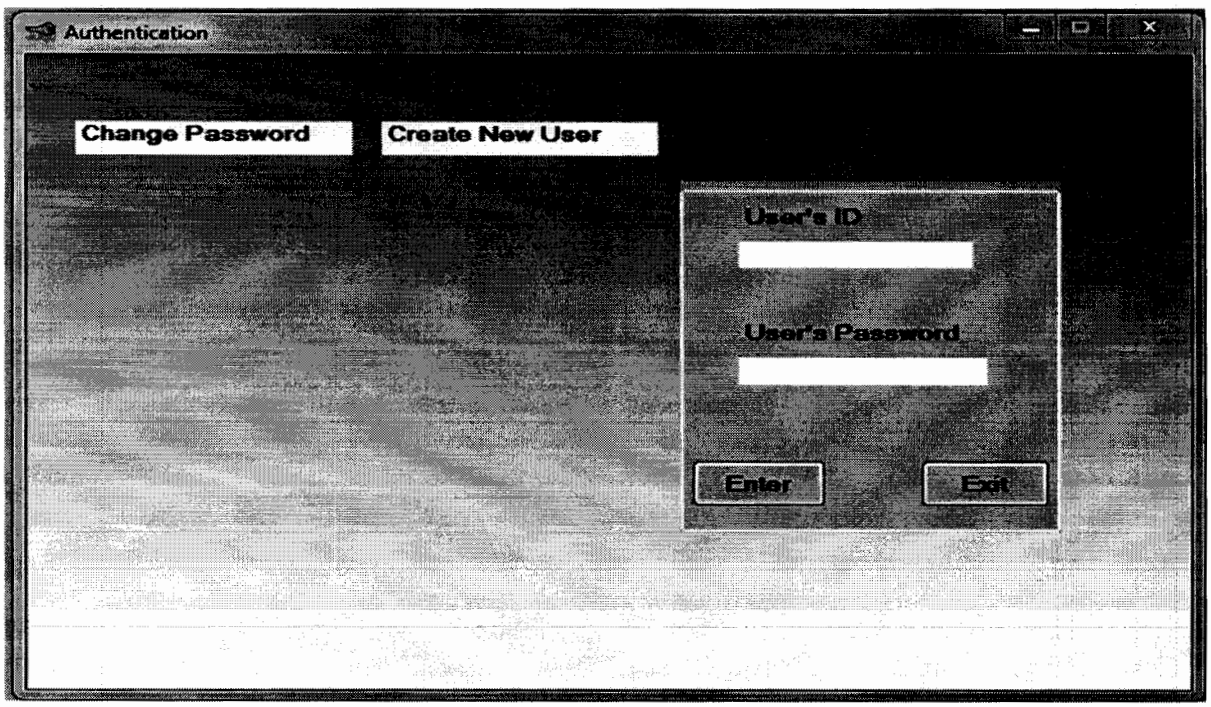

Figure 1: User Authentication Form

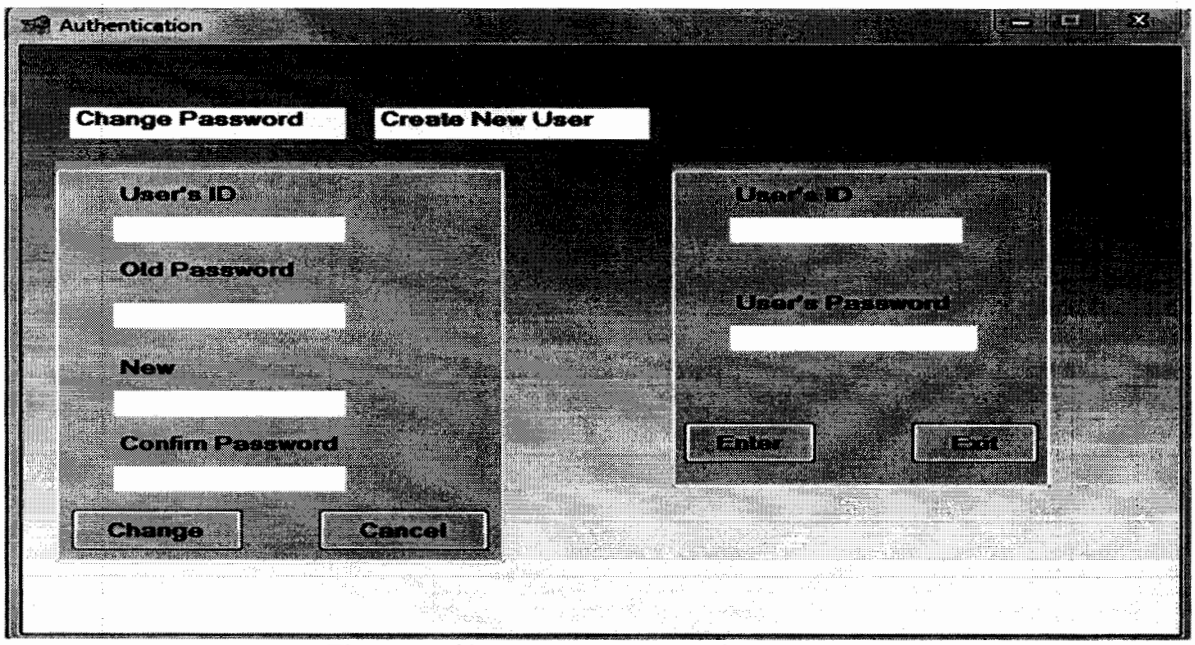

Figure 2: Change user Password Form

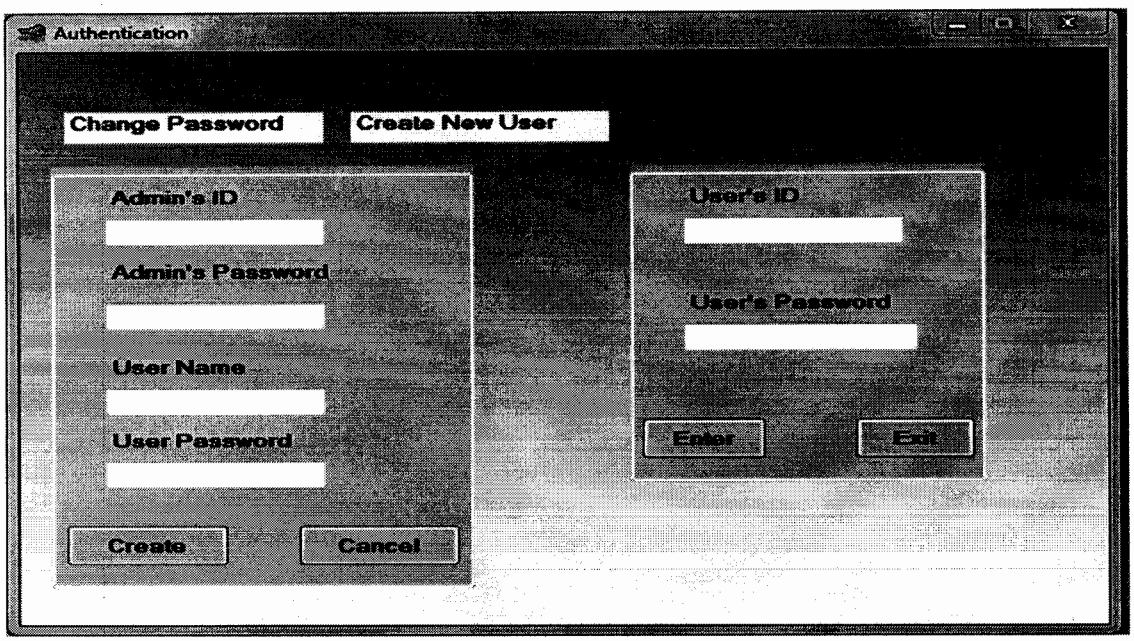

Figure 3: Create new user

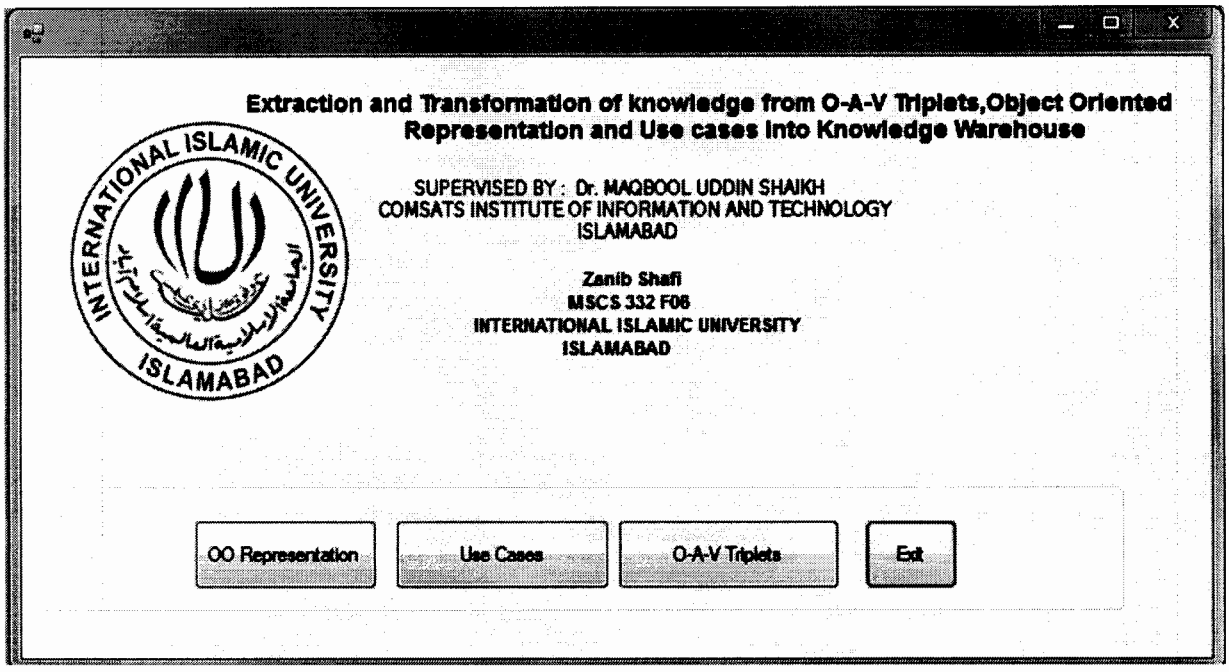

Figure 4: Main Page of the Application

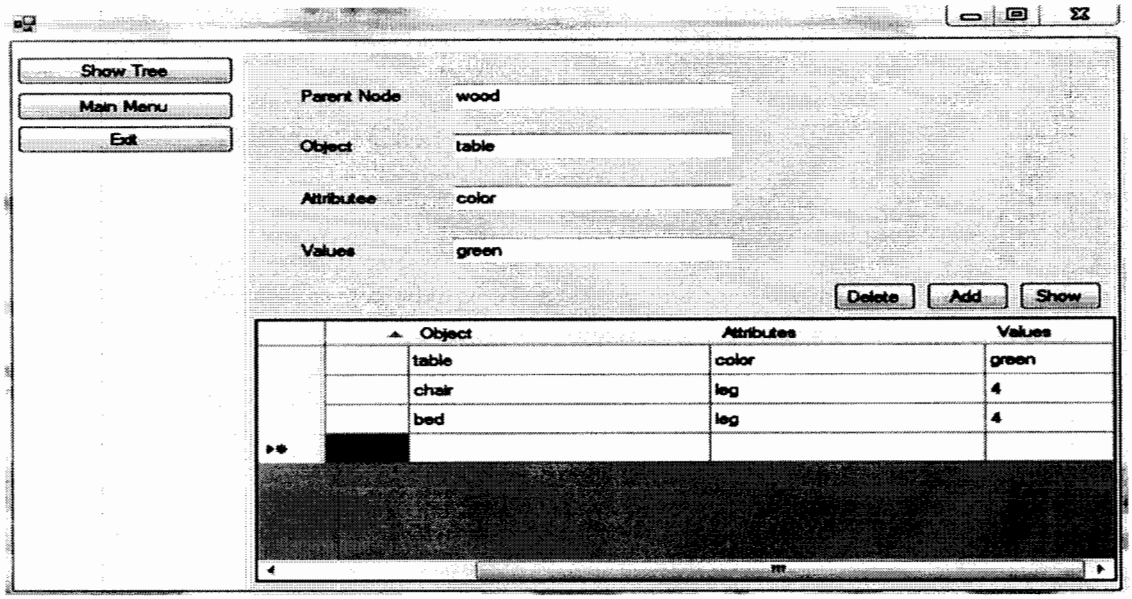

Figure 5: input of the application of O-A-V Representation

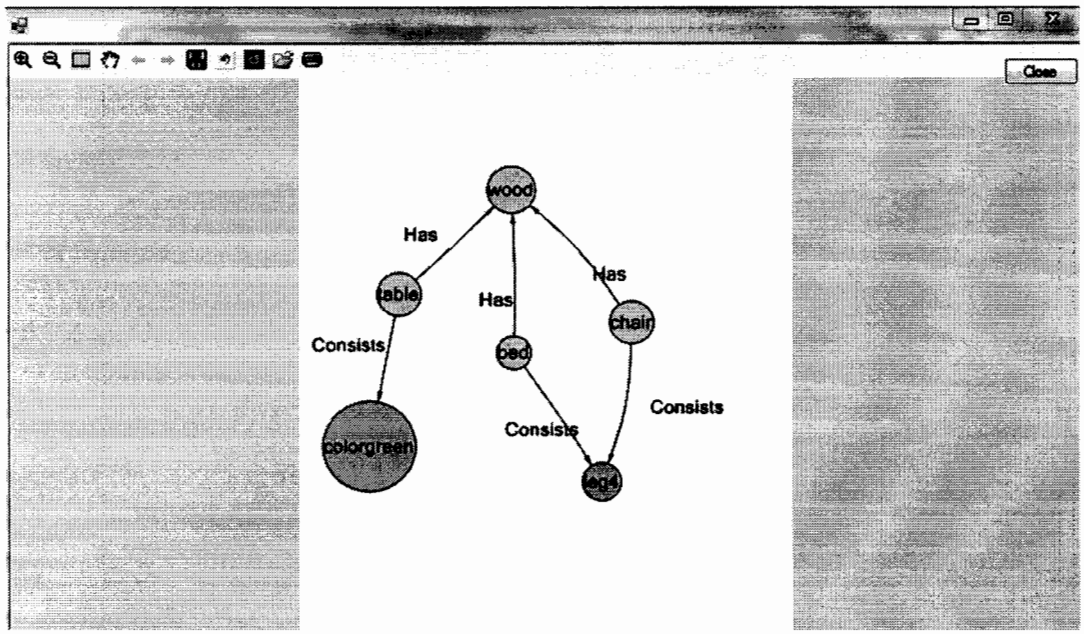

Figure 6: directed graph of O-A-V Representation

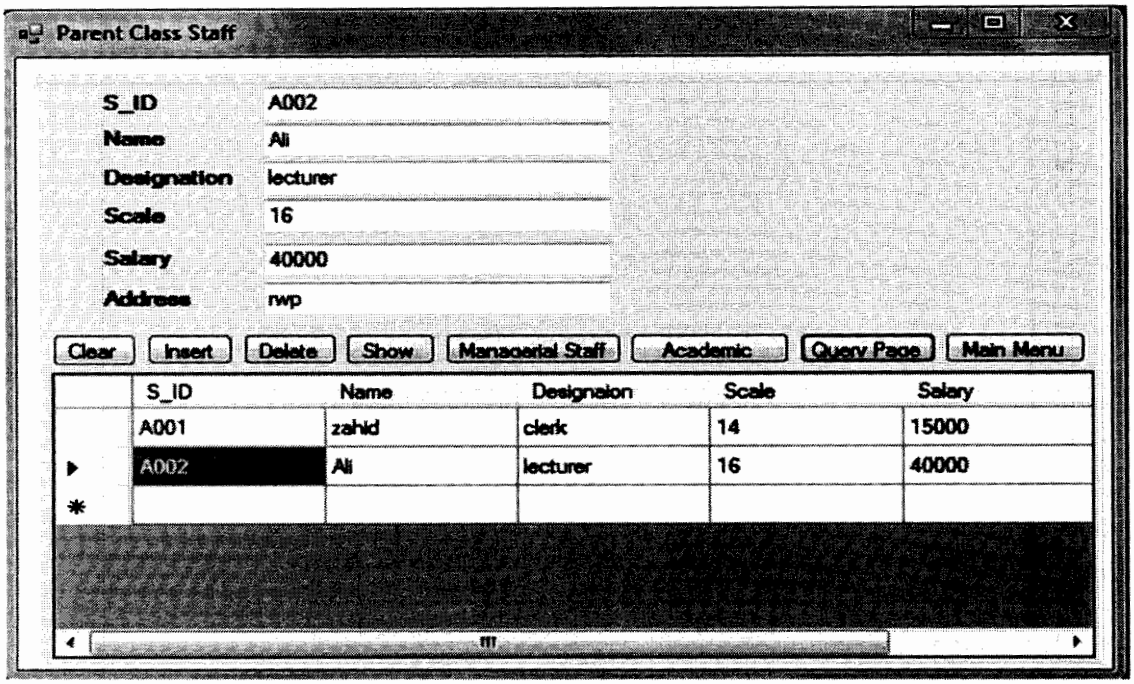

Figure 7: OOR input to the application

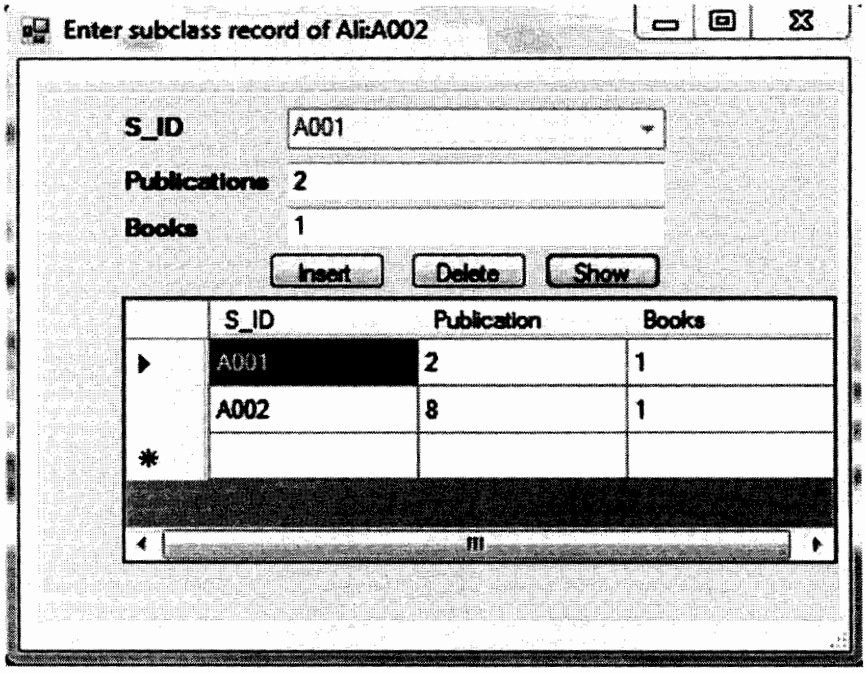

Figure 8: Sub-class input to application

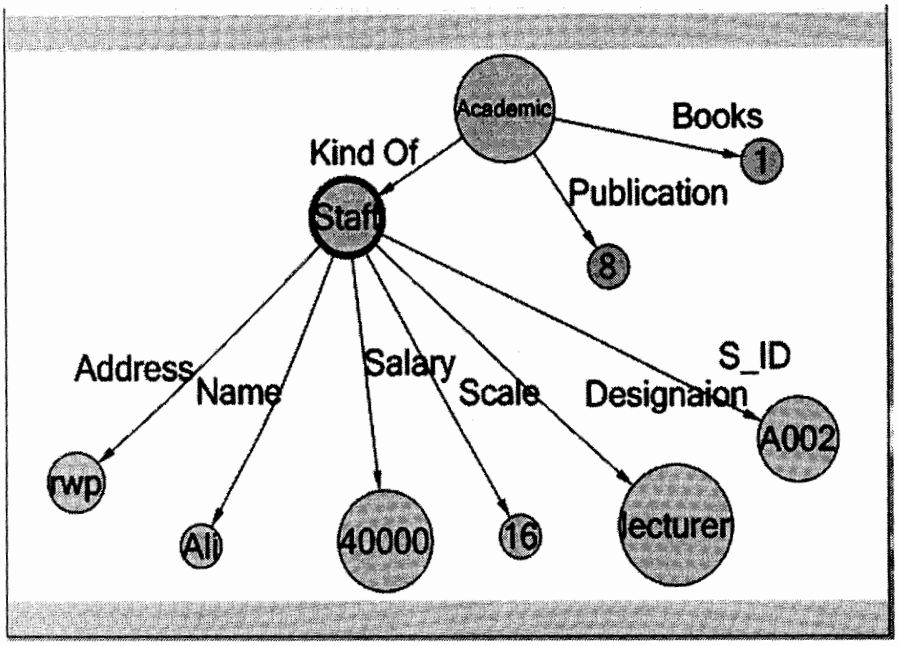

Figure 9: Directed graph of OOR

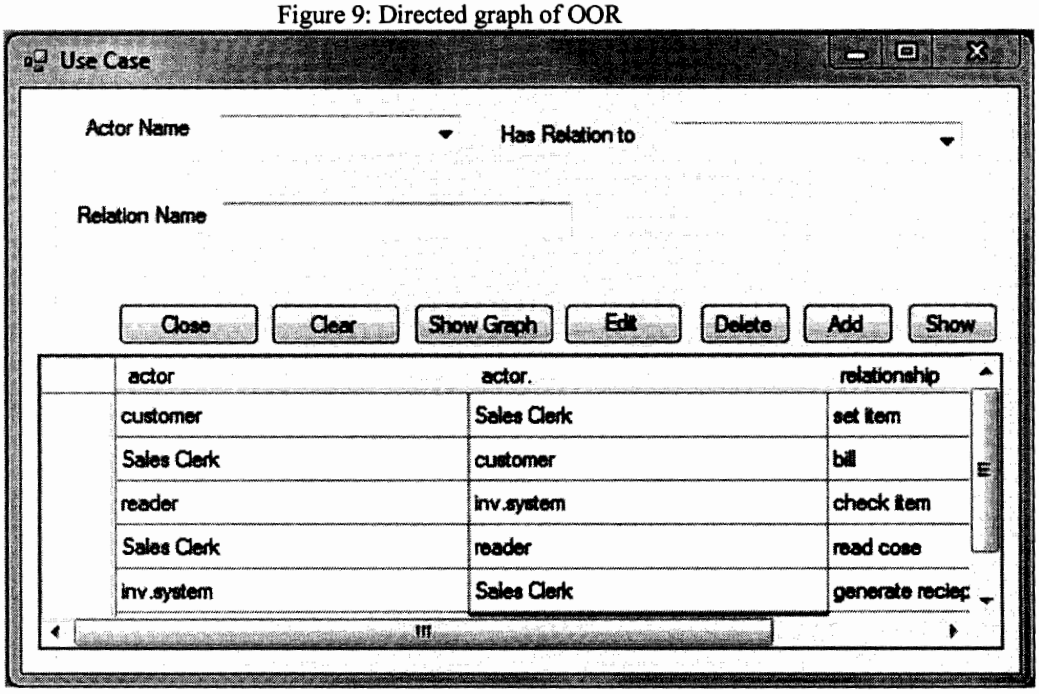

Figure 10: Use case input to application

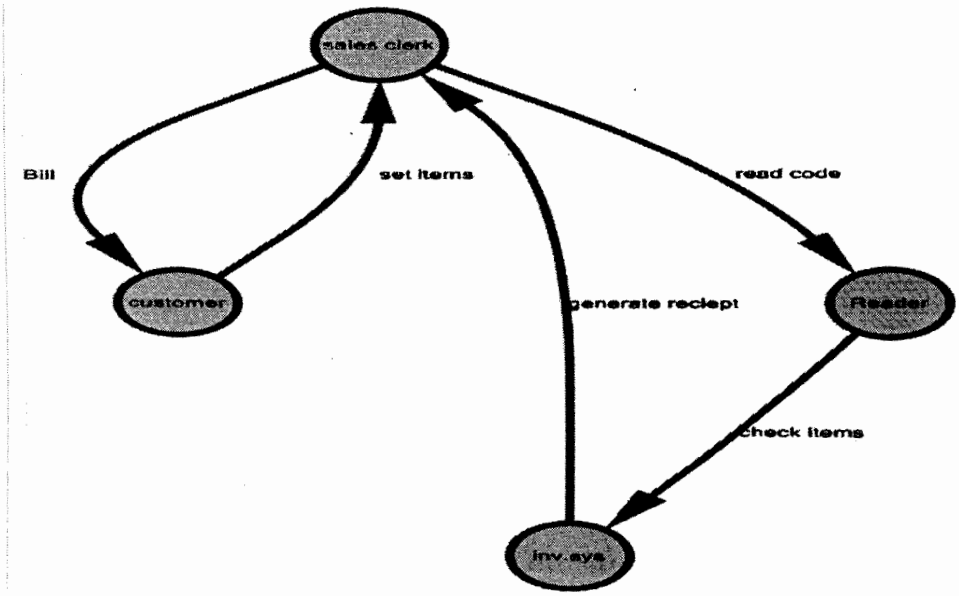

Figure 11: Directed Graph of Use Case

Appendix B

References

[1]. Decision Support System, http://en.wikipedia.org/wiki/Decision_support_system accessed on 20-5-2010

[2]. E Turban, Jay E. Aronson, "*Decision Support Systems and intelligent systems*", 6th edition, Prentice Hall of India, 2004

[3]. Data Warehouse'', http://en.wikipedia.org/wiki/Data_warehouse accessed on june 20 ,2010

[4]. Azra Shamim, Hameed Hussain, Maqbool Uddin Shaikh, A Framework for Generation of Rules from Decision Tree and Decision Table, International Conference on Information and Emerging Technologies , 2010

[5]. 'Knowledge Management System'' http://en.wikipedia.org/wiki/Knowledge_management_system accessed on 20-5-2010

[6]. Nemati H.R., David M, Laksmi S.Lyer.Herschel ''Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing'', Decision support Systems pp 33:143-161,2002

[7]. Anthony Dymond, "The Knowledge Warehouse: The Next Step beyond the Data Warehouse' *Data Warehousing and Enterprise Solutions*,2002

[8]. H. Zhang and Yin Liang's Knowledge Warehouse System for Enterprise Resource Planning Systems'' System Research and Behavioral Sciences, Syst.Res .pp 169-176,2006

[9]. Michael Yacci, "The Knowledge Warehouse: Reusing Knowledge Components pp 75-5416 , 2002

[10]. Kimball R. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Confirming and Delivering Data.* John Wiley & Sons, 2004

[11]. M. Kantardzic, ''*Data Mining: Concepts, Models, Methods, and Algorithms*'', Wiley-IEEE Press, 2002

[12]. Antony Brownea , Brian D. Hudsonb , David C. Whitleyb ,Martyn G. Fordb , Philip Pictonc ''Biological data mining with neural networks: Implementation and application of a flexible decision tree extraction algorithm to genomic problem domains''Neurocomputing, pp 275 – 293, 2004

[13]. Galhardas H, (2009) "ETL (Extract-Transform-Load", https://dspace.ist.utl.pt/bitstream/2295/294968/1/licao_6.pdf accessed on May 20 ,2010

[14]. Opim Salim and Shahrul Azman Noah' 'A Transformation Methodology to Knowledge-based Conceptual Data Warehouse Design'' *Journal of Computer Sciences,* pp 460-465 ,2006

[15]. Pedersen,T.B, (2006) -DWML course,Torben Bach Pedersen, Aalborg University http://www.cs.aau.dk/~tbp accessed on May 20, 2010

[16]. V. Raman, J. Hellerstein. Potter's Wheel: ''An Interactive Data Cleaning System.'' *Proceedings of 27th International Conference on Very Large Data Bases (VLDB),* pp 381-390, Italy, 2001.

[17]. Antony Brownea , Brian D. Hudsonb , David C. Whitleyb ,Martyn G. Fordb , Philip Pictonc ''Biological data mining with neural networks: Implementation and application of a flexible decision tree extraction algorithm to genomic problem domains''*Neurocomputing,* pp 275 – 293,2004

[18]. Ali Kamrani,Wang rong,Ricardo ''A genetic algorithm methodology for data mining and intelligent knowledge acquisition' *Computers and industrial engineering* pp 361-377 ,2001

[19]. Min-Yuh Day, Richard Tzong-Han Tsai, Cheng-Lung Sung,Chiu-Chen Hsieh, Cheng-Wei Lee , Shih-Hung Wu, Kun-Pin Wu,Chorng-Shyong Ong , Wen-Lian Hsu''Reference metadata extraction using a hierarchical knowledge, 2006

[20]. Omer Akgobek , Yavuz Selim Aydin , Ercan ztemel , Mehmet Sabih Aksoy.''A new algorithm for automatic knowledge acquisition in inductive learning''*Knowledge-Based System* , pp 388-395, 2006

[21]. John Cheng and Ray Emami, Larry Kerschberg Eugene Santos, Jr,. Qunhua Zhao,Hien Nguyen, and Hua Wang ,Michael Huhns,Marco Valtorta'' OmniSeer: A Cognitive Framework for User Modeling, Reuse of Prior and Tacit Knowledge, and Collaborative Knowledge Services'' *Proceedings of the 38th Hawaii International Conference on System Sciences*, 2005

[22]. J.Firestone "Knowledge Base Management Systems and the Knowledge Warehouse: A Strawman (2000)" In: Executive Information System

[23]. Tasleem uddin, "Knowledge Warehouse Framework: A new direction towards decision support system for executive management'' Thesis Report, International Islamic University Islamabad, *http://misqe.org/ojs2/index.php/misqe/author/index/completed* , 2008

[24]. Ala'a H., Soukaena Hssan''An Approach for Facilating Knowledge Data Warehouse ''*International Journal of Soft Computing applications ISSN, pp 35-40* ,2009

[25]. Xixu Fu, Hui Wei '' Cognition Inspired Object Oriented Knowledge Warehouse Architecture'', *Journal Of Software, Vol 5, No 9* ,2010

[26]. Bilal Ahmad, Knowledge Extraction Framework from Scripts, Use-cases, Knowledge Maps, Decision Tables and Decision Trees for Knowledge Warehouse, Thesis Report, International Islamic University Islamabad ,2008

[27]. Thomas Connolly, Carolyn Begg, Database Systems: A Practical Approach to Design, Implementation and Management, 4th Edition, Addison-Wesley, 2003

[28]. Data Warehouse Evolution Framework, http://syrcodis.citforum.ru/2007/13.pdf, Accessed Data: Oct 11, 2008

[29]. Z. Bellahsene, Schema, "Evolution in Data Warehouses", *Knowledge and Information Systems*, Springer-Verlag, pp 283-304, 2002

[30]. S. Chen, X. Zhang, E.A. Rundensteiner, "A Compensation-Based Approach for Materialized View Maintenance in Distributed Environments", In Computer Science Technical Report, Worcester Polytechnic Institute, Worcester, MA, USA, 2004

[31]. Dmitry Davidov,Ari Rappoport'' Extraction and approximation of Numerical attributes from the web ''*Proceedings of the 38 th Annual Meeting of the association for computational linguistics*, pp 1308-1317, 2010

[32]. Marcos Zuniga, Francois Bremond, Monique Thonnat''Incremental Video Event Learning''*Springer-Verlag Berlin Heidelberg* , pp 403-414, 2009

[33]. Data Warehousing and OLAP, www.cs.uh.edu/~ceick/6340/dw-olap.ppt , Accessed Data: Nov 25, 2008

[34]. Jeffrey A. Hoffer, Mary B. Prescott, Fred R. McFadden, Modern *database management*, Sixth Edition, Pearson Education Publishers, Singapore

[35]. Daniel L. Moody, Mark A.R. Kortink, "From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design", *In Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2000),* pp 55-66, 2000

[36]. Mohammad Rifaie, Erwin J. Blas, Abdel Rahman M. Muhsen, Terrance T. H. Mok, Keivan Kianmehr, Reda Alhajj, Mick J. Ridley, "Data warehouse Architecture for GIS Applications", In Proceedings of the 10th International Conference on Information Integration and Web-based Applications & Services (iiWAS '08) , 2008

[37]. William Inmon, *Building the Data Warehouse*, 2nd Edition, New York: Wiley publisher. Inc, 1996

[38]. Ralph Kimball, M. Joy and T. Warren, the Data warehouse Toolkit: with SQL server and Microsoft Business Intelligence Toolset, 2nd Edition, New York: Wiley publisher. Inc, 2006

[39]. E.A. Rundensteiner, A. Koeller, X. Zhang, "Maintaining Data Warehouses over Changing Information Sources", Communications of the ACM, Volume, 43, New York, NY, USA, pp 57-62, 2000

[40]. S. Chen, X. Zhang, E.A. Rundensteiner, "A Compensation-Based Approach for Materialized View Maintenance in Distributed Environments", In Computer Science Technical Report, Worcester Polytechnic Institute, Worcester, MA, USA, 2004

[41]. Kimball R,The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Confirming and Delivering Data. John Wiley & Sons, 2004

[42]. V. Raman, J. Hellerstein. Potter's Wheel: ''An Interactive Data Cleaning System.'' Proceedings of 27th International Conference on Very Large Data Bases (VLDB), pp 381-390, 2001

[43]. Hong Zhang and yin Liang''A Knowledge Warehouse System for Enterprise Resource Planning Systems''System Research and Behavioural Sciences, Syst.Res.23, pp 169-176,2006.

[44]. Rizwana Irfan ,Dr.Maqbool uddin Shaikh ''Enhance Knowledge Management Process for Group Decision Making'' in International Conference on Computer Engineering and Applications (ICCEA), 2010

[45]. Uma Sekaram, Research Methods for Business, 3rd Edition, United State of America, John Wiley and Sons Inc, 2000.

[46]. Qualitative and Quantities Research Methods, www.ics.heacademy.ac.uk/events/presentations/271_slides3Oct.ppt, Accessed Data: Nov 24, 2008

[47]. Norman K. Denzin, Yvonna S. Lincoln, The SAGE Handbook of Qualitative Research, Sage Publications, 2005

[48]. Qualitative Research An Introduction, www.staff.ncl.ac.uk/david.harvey/AEF801/MBQualMeth.ppt, Accessed Data: Nov 25, 2009

[49]. Qualitative Research Methodology https://idenet.bth.se/servlet/download/news/26148/winter_070329_qualitative_research_methodology.pdf, Accessed Data: Nov 25, 2009

[50]. Research Methodologies, www.cs.uvic.ca/~mstorey/teaching/infovis/course_notes/researchmethods.pdf, Accessed Data: Nov 25, 2009

[51]. Xixu Fu , Hui Wei, Cognition inspired object oriented knowledge warehouse architecture ,Journal of software ,Vol 5,No 9, 2009

[52]. Chenghua Dang et al, Wordnet-based summarization of unstructured document, , WSEAS Transactions on Computers, Volume 7 , Issue 9,pp:1467-1472 ,2008

[53]. Debasish Kundu, Debasis Samanta, ''An Approach for Assessment of Reliability of the System Using UseCase Model'' *10th International Conference on Information Technology,IEEE*,7695-3068, 2007

[54]. Debashish Kundu,Debasis Samanta,''A Novel Approach of Prioritizing Use Case Scenarios'',14th Asia-Pacific Software Engineering Conference,2007

[55]. Tom Mens,Gabriele Taentzer,Olga Runge ''Analyzing Refactoring Dependencies using Graph Transformation'',*Soft system Model,Springer-Verlag*, 2007

[56]. W.L. Xu_, L. Kuhnert1, K. Foster, J. Bronlund, J. Potgieter, O. Diegel '' Object-oriented knowledge representation and discovery of human chewing Behaviors'' Engineering Applications of Artificial Intelligence, 2007