Knowledge Discovery for Moderating Collaborative Projects

by

Alok K Choudhary

A Doctoral Thesis

Submitted in partial fulfilment of the requirement for the award of

Doctor of Philosophy
of
Loughborough University

May 2009

Research Student Office, Academic Registry
Loughborough University Leicestershire LE11 3TU UK
Switchboard: +44 (0)1509 263171 Fax: +44 (0)1509 223938

Loughborough University

CERTIFICATE OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgments or in footnotes, and that neither the thesis nor the original work contained therein has been submitted to this or any other institution for a degree.

...Alok Choudhary.......... (Signed)

...26/05/09...... (Date)

Abstract

In today's global market environment, enterprises are increasingly turning towards collaboration in projects to leverage their resources, skills and expertise, and simultaneously address the challenges posed in diverse and competitive markets. Moderators, which are knowledge based systems have successfully been used to support collaborative teams by raising awareness of problems or conflicts. However, the functioning of a moderator is limited to the knowledge it has about the team members. Knowledge acquisition, learning and updating of knowledge are the major challenges for a Moderator's implementation. To address these challenges a **K**nowledge disc**O**very **A**nd da**T**a min**IN**g inte**G**rated (KOATING) framework is presented for Moderators to enable them to continuously learn from the operational databases of the company and semi-automatically update the corresponding expert module. The architecture for the Universal Knowledge Moderator (UKM) shows how the existing moderators can be extended to support global manufacturing.

A method for designing and developing the knowledge acquisition module of the Moderator for manual and semi-automatic update of knowledge is documented using the Unified Modelling Language (UML). UML has been used to explore the static structure and dynamic behaviour, and describe the system analysis, system design and system development aspects of the proposed KOATING framework.

The proof of design has been presented using a case study for a collaborative project in the form of construction project supply chain. It has been shown that Moderators can "learn" by extracting various kinds of knowledge from Post Project Reports (PPRs) using different types of text mining techniques. Furthermore, it also proposed that the knowledge discovery integrated moderators can be used to support and enhance collaboration by identifying appropriate business opportunities and identifying corresponding partners for creation of a virtual organization. A case study is presented in the context of a UK based SME. Finally, this thesis concludes by summarizing the thesis, outlining its novelties and contributions, and recommending future research.

Keywords: Knowledge Discovery and Data Mining, Moderator, Universal Knowledge Moderator, Unified Modelling Language, construction project supply chain, Post project reviews (PPRs), Text Mining, SMEs, and Virtual Organization.

Acknowledgements

Firstly, I would like to extend my most sincere thanks to Dr. Jenny Harding for her great support, motivation and extra ordinary guidance at various stages of my PhD journey. She has always been a great source of inspiration for me to pursue my research. Her constructive criticism and thought provoking discussion added a great value to the thesis. It is an honour for me to pursue my research under her learned supervision and I will be obliged throughout my life for her kindness and support.

My PhD journey could not have been possible without Prof. M. K. Tiwari; I am also thankful to Mrs. Usha Tiwari and lovely Manu and Ashu for their love and support throughout my PhD.

I also want to mention a great person, who is not with us, my dearest Rajeev Mama who was a great source of inspiration for a guy like me who hadn't even thought to go to an engineering college. It was his blessing and support which makes it possible not only to go to the engineering college but to achieve the highest degree in engineering.

My colleagues and friends Rahul, George, Srinivas, Bishnu, Nadeem, Nitesh and Claire need a special mention who always shared their knowledge and showed enthusiasm and provided supporting remarks to the research ideas.

My Friends Sanjeev, Prakash, Anand, Nagesh, Neeraj, Vikas, Raj, Akhilesh, Saurabh, Vikram, Pawan and Manish, have played a vital role in getting me through this difficult research journey and, also ensured that I enjoyed my life as well.

I also acknowledge the support of Innovative Manufacturing and Construction research centre in the form of research funding which made this research possible and provided an opportunity to realise my dream.

Finally, I would like to thank god for giving me such a good supervisor, mentor, parents, relatives and friends, without which it was impossible to reach this level.

Table of Content

List of Figures

List of Tables

Glossary of Terms

AI	artificial intelligence
AKTIRC	Advanced Knowledge Technologies Interdisciplinary Research Collaboration
ANN	Artificial Neural Network
API	Application Programming Interface
ARIMA	Autoregressive Integrated Moving Average
ARM	Association Rule Mining
CART	classification and regression tree
CN	Collaborative Network
CoAKTinG	Collaborative Advanced Technologies in the Grid
CPFR	Collaborative Planning Forecasting and Replenishment
CPSC	Construction project supply chain
CRISP-DM	Cross Industry Standard process for data mining
CRUD	create, read, update and delete
DM	Decision Maker
DM	Data Mining
DMASA	Data mining application system approach
DMM	Design Moderation Module
DMSTA	Data mining software tool approach
DSS	Decision Support System
ECOLEAD	European Collaborative networked Organizations LEADership initiative
EE	Extended Enterprise
EM	Engineering Moderator
EMs	Expert Modules
EPSRC	Engineering Physical Science Research Council
ERP	Enterprise Resource Planning
ES	expert systems
E-SCM	E-supply chain
GA	Genetic Algorithm
GATE	General Architecture for Text Engineering
ICT	Information and Communication Technology
IE	Information Extraction
IM	Information Manager
IR	Information Retrieval
KAM	Knowledge Acquisition Module
KBS	knowledge-based systems
KD/DM	Knowledge Discovery and Data mining
KDD	Knowledge Discovery in Databases
KDT	Knowledge Discovery in Text
KM	Knowledge Management

KMS	Knowledge Management systems
KRM	Knowledge Representation Model
LA	Ling Analysis
MISSION	Modelling and Simulation Environment For Design Planning and Operation of Globally Distributed Enterprises
MM	Moderation Module
MMP	MISSION Modelling Platform
MOSES	Model Oriented Simultaneous Engineering System,
MRCA	Multi-resolution collaborative architecture
MSE	Manufacturing System Engineering
MSEM	Manufacturing System Engineering Moderator
OLAP	Online Analytical Process
OMT	Object Modelling Technniqe
OODB	Object Oriented Data Base
OOSE	Object Oriented Software Engineering,
OWL	Web Ontology Language
PCB	Printed Circuit Board
PPR	Post Project Reviews
RDB	Ontology Acquisition Module
RDF	Resource Description Framework
RST	Rough Set Theory
SEC	Securities and Exchange Commission
SGML	Standard Generalized Markup Language
SMEs	Small and medium enterprises
SYNERGY	Supporting HighlY Adaptive Network Enterprise CollaboRation ThrouGh SemanticallY Enabled Knowledge Services
TA	Text Analysis
TeleDSS	web -based decision support systems
TM	Text Mining
UKM	Universal Knowledge Moderator
UMES	Universal Manufacturing Enterprise Schema
UML	Unified Modelling Language
URL	Universal Resource locators
VBE	VO Breeding environment
VE	Virtual Enterprise
VEN	Virtual Enterprise Network
VF	Virtual Factory
VICS	Voluntary Inter-Industry Standards
VM	Virtual Manufacturing
VO	Virtual Organization
VPN	Virtual Private Network
VS	Value System

1

Context

1.1 Background

Increasingly competitive market trends demand highly customized products with ever shortening production time and this trend is expected to accelerate. Consequently, modern business entities are challenged to identify effective means of reducing production cost, improving product and service quality, reducing time to market delivery, accelerating responses to customer requirement and bettering flexibility and system's reusability. Industries are striving to meet these challenges by focussing on their core competencies, integrating and collaborating intensively and migrating towards knowledge based manufacturing [1]. Collaboration varies from intra-organizations to inter-organizations through projects such as collaborative product development, collaborative supply chain, collaborative design, collaborative manufacturing, virtual enterprises (VE), extended enterprises (EE), and virtual organizations (VO) etc. These exploit the core competencies of all the enterprises concerned to form strategic alliances to enhance competitiveness by integrating value added activities, information, resources and knowledge between enterprises or teams.

The *ManuFuture 2005* Conference and its strategic research agenda "Assuring the Future of Manufacturing in Europe" clearly indicated that the future of manufacturing in Europe is focussed on knowledge based systems [2]. Nonaka [3] observed that, as the market shifts, technologies proliferate, competitors multiply, and products and services become rapidly obsolete, successful companies are characterized by their ability to consistently create new knowledge, quickly disseminate it, and embody it in new products and services. Therefore, it can be seen that in the knowledge based economy, where knowledge assets take the centre stage, whoever owns knowledge and can create useful knowledge from existing data, information and knowledge will enjoy absolute advantage over the competition. Knowledge exists in all business functions, from product conceptualization to its sale to and feedback from customers. However, knowledge can be notoriously difficult to identify, capture, manage and reuse. The proliferation of large masses of data has created many opportunities in diverse application areas including engineering and businesses. The opportunities are offered by the abundance and availability of data and at the same time the challenges are posed by the problems of how to organize, extract and retrieve useful and novel knowledge from this data. The field of knowledge discovery and data mining has emerged to address these new opportunities and challenges.

Globally distributed collaborative projects aim to enhance and facilitate engineering agility through collaboration, sharing information and exchanging knowledge seamlessly between partners. One of the major issues in multidiscipline collaborative projects is how best to share and simultaneously exploit different types of expertise, without duplicating efforts or inadvertently causing conflicts or loss of efficiency through misunderstandings of individual or shared goals. The concepts of Moderators to support collaboration and team working have been researched in major research projects [4-6]. The main function of a moderator is to support collaborative working teams by raising individual members' awareness of the needs and experiences of other team members and the concept has been successfully demonstrated in product design, manufacturing system design, extended enterprise and e-supply chain. Prototype Moderators have been demonstrated in the form of knowledge based software support systems consisting of a moderation module, multiple expert modules and a knowledge acquisition module. Until now, all knowledge acquisition for the prototype moderators has been done manually, based on human expertise and experience.

1.2 Research Aims and Objectives

The research presented in this thesis addresses the issues introduced in section 1.1 by enhancing the knowledge acquisition capability and functionality of Moderators. ***The Research Aim is:-***

To enhance the functionality and capability of Moderators through the integration of a knowledge discovery based semi-automatic knowledge acquisition framework which enables Moderators to "learn" and "update" their relevant expert modules from knowledge discovered in the existing operational databases of companies. The proposed Universal Knowledge Moderator, equipped with knowledge discovery capability should increase awareness within the project teams by highlighting potential problem areas (from previous project

experience) and should identify new collaborative business opportunities to initiate new projects

To satisfy this research aim, the following objectives have been addressed:-

1. To propose and develop a knowledge discovery and data mining integrated framework for Moderator services. This framework should enable semi-automatic knowledge acquisition and update of knowledge and learning by discovering knowledge from operational databases of industries.

2. To integrate this knowledge discovery and data mining integrated framework with a state of the art Moderator, called the Universal Knowledge Moderator (UKM) to enhance collaboration in e-manufacturing supply chains and illustrate this through an example.

3. To model the design of this framework using the Unified Modelling Language (UML).

4. To demonstrate the application of the framework by showing its use in extracting knowledge from Post Project Review (PPR) reports from construction project supply chains and showing how this can be used to update the expert modules of a Construction Project Moderator (CPM).

5. To demonstrate the application of the framework by showing its use in raising awareness of business opportunities and identifying collaborative SME partners for a virtual organisation (VO).

The knowledge discovery and data mining integrated framework that has been designed and developed during this research is called the KOATING framework (**K**nowledge disc**O**very **A**nd da**T**a m**IN**ing inte**G**rated framework), and it will be referred to as the KOATING framework throughout this thesis.

.

2

Scope of Research

This chapter defines the key research question, provides an overview of the scope of research and outlines the structure of the thesis. It identifies the main research areas and the main contribution areas.

Chapter Outline

2.1 The Research Question

Based on an initial literature review in the area of collaborative projects, knowledge management systems, knowledge discovery applications and industrial requirements to learn from past experiences by creating useful knowledge, the author devised the following research question.

> *"How to provide Ongoing Learning, (Semi)-Automated Knowledge Acquisition and Enhanced Awareness Capability through Knowledge Discovery Integrated Moderator Services for Collaborative Projects?"*

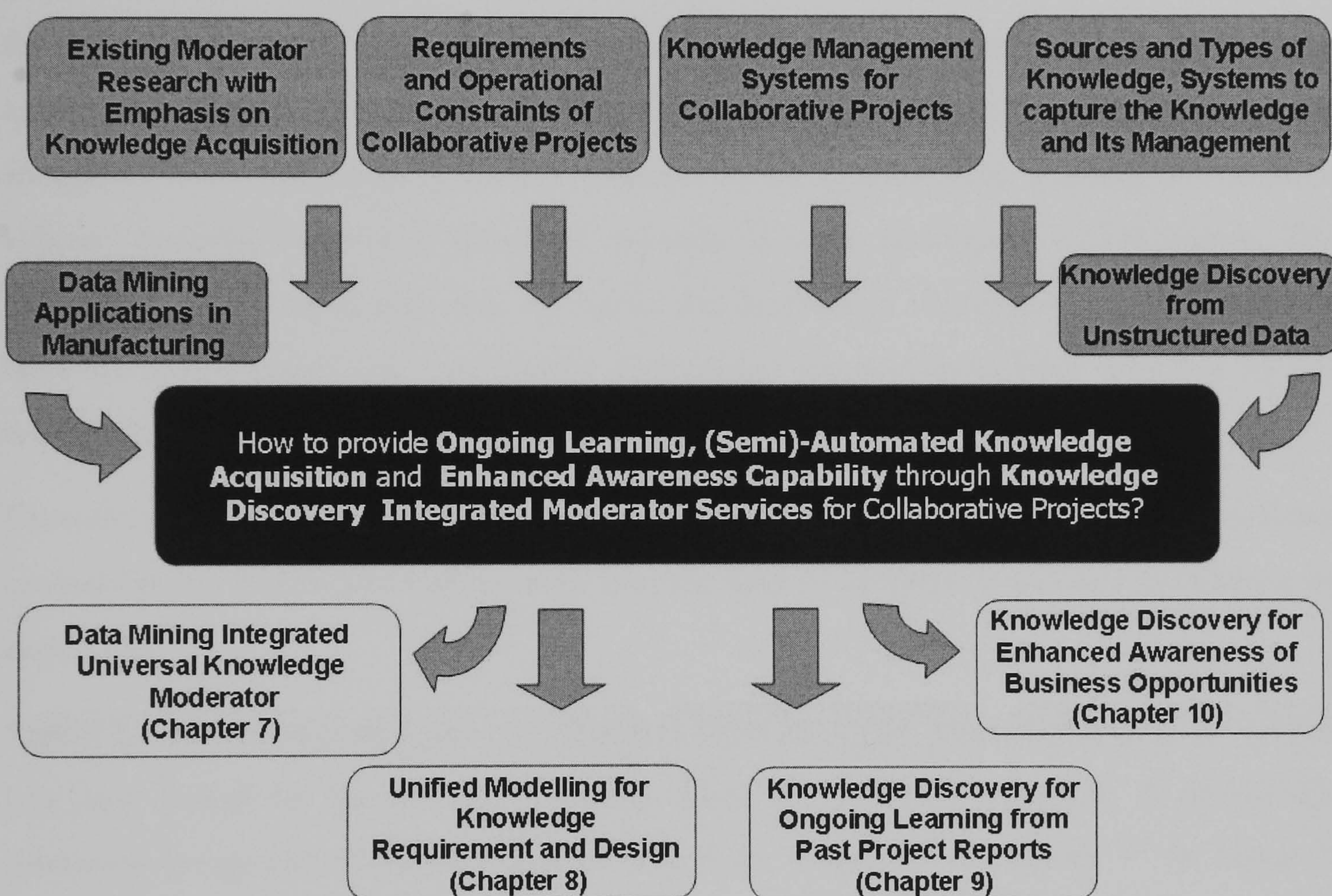


Figure 2-1: The research question

In Figure 2-1, the research question is shown in the central box, and the main research problem areas associated with this question are represented by shaded arrows going into the research question. This is multi-disciplinary research encompassing multiple domains as shown in the six shaded boxes above the research question box. Most collaborative projects face major challenges in identifying, acquiring, maintaining, evolving and sharing of knowledge. A better understanding of the requirements of virtual enterprise(VE), extended enterprise(EE) and virtual organization(VO), their operational constraints and functions are needed. Knowledge Management systems (KMS) have been developed for effective knowledge sharing, identifying problems and raising awareness between members. A thorough review and understanding of these systems was required to enable this research to make a novel contribution in the research areas. Moderators, which are knowledge based systems, promise to identify conflicts and raise awareness between project team members, but earlier work in Moderator technology is lacking in terms of capturing new knowledge and ongoing learning capability. Furthermore, Knowledge itself is notoriously difficult to capture and therefore an understanding of different types of knowledge, identification of sources of knowledge and various types of system for acquisition and management is needed. Knowledge discovery in databases and data mining have attracted the attention of researchers and practitioners interested in its application within the manufacturing domain. However, its application is not straightforward and major problems include its integration with existing information support systems, varieties of data, and diversity of tools, techniques and functions. The concept of text mining may also be useful for knowledge identification when available data are unstructured yet text mining approaches have seldom been adopted by the manufacturing and construction industries.

Considering all these requirements and challenges, there is a need for further analysis, evaluation and exploitation of research into the area of Moderator services to enhance its capability.

Figure 2-1 also shows the main contributions areas from this research beneath the central box and linked by the arrows emerging from the research question. A knowledge discovery integrated Moderator service has been proposed in chapter 7. In order to design the proposed system in chapter 8, Unified Modelling Language has been used to capture the design requirement and development of the system. It has been shown in chapter 9 that the proposed system can be used to discover and update knowledge of different kinds in the expert modules from post project reports of construction projects. Furthermore, the results of this research illustrate how the capability of existing moderator concepts can be enhanced by raising awareness of new business opportunities thereby enabling new or improved collaboration to be established in chapter 10.

2.2 Research Methodology

The research methodology adopted in this thesis is delineated in Figure 2-2. It shows that the first phase in the research methodology is about understanding the problem area and identifying knowledge discovery tools and techniques which can be used to provide a solution. The initial problem areas identified are shown as inputs in Figure 2-1 and are discussed in section 2.1. An extensive literature review has been carried out in these problem areas to identify the research gaps and determine the requirements and possible solutions for this research as detailed in chapters 3, 4 and 5. It has been found that knowledge discovery in databases (KDD) can be used to fulfil the aims and objectives. Therefore, a review of related tools and techniques related to KDD had to be carried out to identify appropriate tools and techniques which could be adopted for data mining in the KOATING framework.

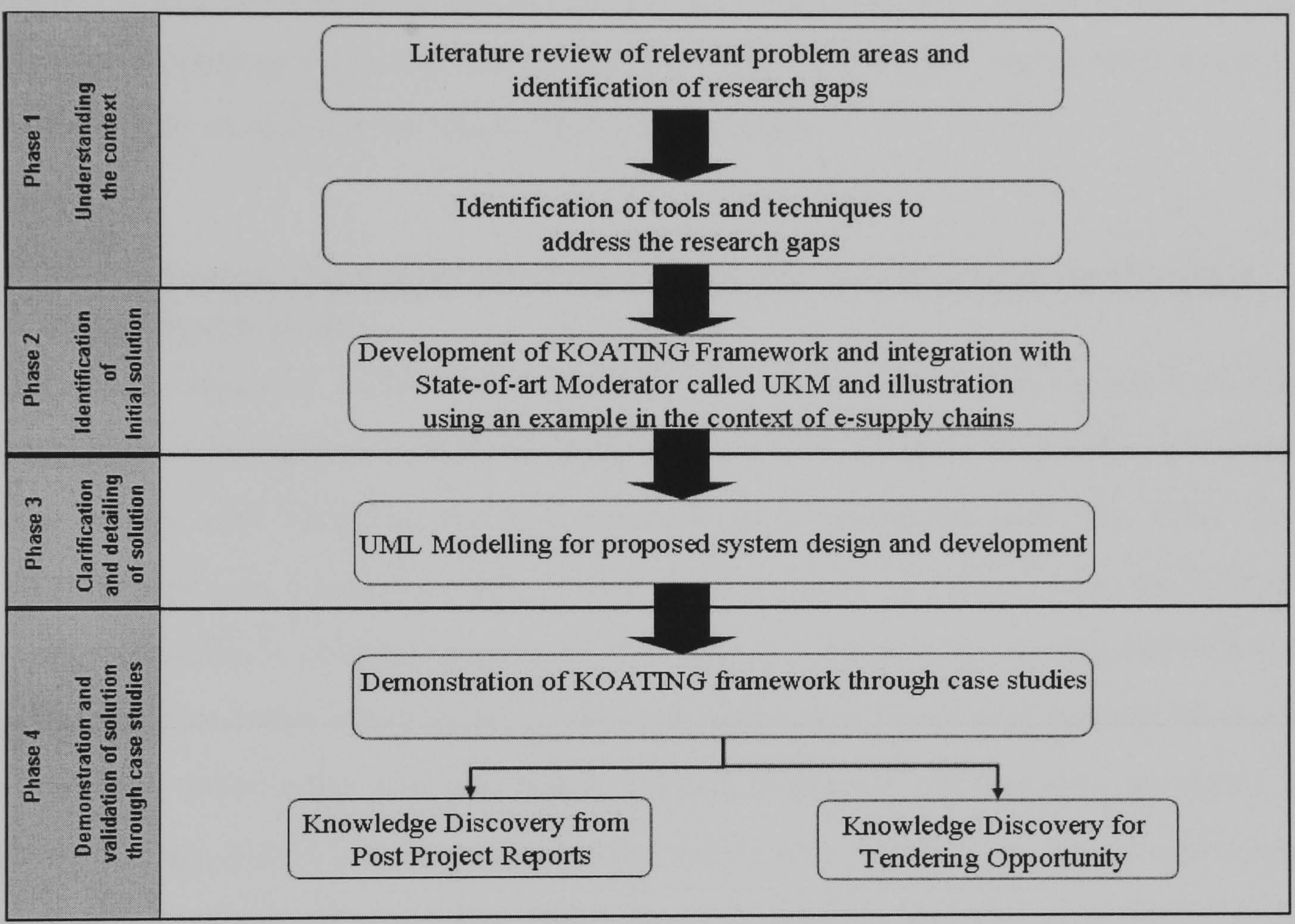


Figure 2-2: Research methodology adopted in this research.

To address the objective of embedding learning and semi-automatic knowledge updating capability, the second phase of the research methodology was to design a solution and carryout some initial validation by examining this initial solution in the context of a simple e-supply chain example. The KOATING framework was therefore developed and integrated with the existing Moderator services.

In phase 3, the KOATING framework was clarified and detailed further by modelling it using unified modelling language (UML).

Finally, in the 4th phase of the research the KOATING framework was tested by demonstrating its use in two different case studies.

2.3 Scope of Research

To address the challenges posed in section 2.1, it is important to thoroughly review the current knowledge management systems for collaborative projects and identify the technologies, which are most likely to underpin the advances in this research area. Taking account of the time and resources available in this research, it is therefore necessary to focus the research effort into those areas where clear benefits and contributions can be identified. Priority has therefore been given to the framework development and system design using modelling tools, and demonstration using case studies, rather than to try to implement a fully functional prototype Moderator system.

2.3.1 Knowledge Management Systems for Collaborative Project and Moderators

As previously explained collaborative projects of various kinds pose many complex challenges but offer many potential advantages. Therefore, Chapter 3 provides a detailed literature survey and identifies the key research gaps associated with this area. The chapter is divided into 3 major sections. Firstly, various types of collaborative projects are discussed. Secondly, a detailed review of the different knowledge management based architectures, knowledge based systems, decision support systems and frameworks used to resolve the knowledge and awareness related issues of collaborative projects, is presented. Finally, focus has been put on the Moderator technology, its key elements, structure and its journey through two major research projects MOSES and MISSION to its current application in the context of E-Supply Chains and their use on the semantic web. Based on this review and learning about different types of system, key challenges and research gaps have been identified. These, along with recommendations identified by researchers across the globe, have been used as a basis for the design of the proposed framework to specifically address shortcomings that has been reported in the published literature.

2.3.2 Knowledge; Types, System and Management

The identification and management of knowledge of different kinds is key to the success of the Moderator's function. Therefore, Chapter 4 consolidates the definitions and differentiations of data, information and knowledge. It classifies different types of knowledge, identifies the possible sources of knowledge and explores research into the acquisition and management of knowledge. Chapter 4 presents a detailed review of various systems to support the knowledge acquisition and sharing from an application point of view especially in the manufacturing domain. It includes knowledge management frameworks, knowledge based system, information and communication technology including semantic web and ontology, expert systems, database technology, modelling and simulation, and knowledge acquisition systems. Based on these application areas, research gaps have been identified primarily in the area of knowledge acquisition. During this research, efforts have primarily been focussed on understanding and exploiting the flexibility and benefits of existing knowledge based systems, whilst determining how to incorporate enhanced knowledge management and knowledge discovery functionality.

2.3.3 Knowledge Discovery and Data Mining

Knowledge Discovery in Databases (KDD) and Data Mining (DM) incorporate theories, algorithms and methods from the intersection of several research fields including database technology, machine learning, statistics, artificial intelligence, knowledge based systems and data visualization. The ultimate goal is to extract useful knowledge from large collections of data. The diversity of data mining tools, techniques and functionalities provides great opportunities, but the profusion of options leave open ended challenges. A detailed review of tools, techniques and functions has been carried out in chapter 5, with a focus on capturing different kinds of knowledge, and identifying appropriate tools and techniques. A novel semi-automated text mining approach has been adopted to automate the process of identifying the research gaps and determining the good practices in the area of data mining applications in manufacturing. In addition, application areas of text mining which are generally used for unstructured data have been reviewed and this review identified that manufacturing and construction industries have yet to exploit these techniques. A key aspect of the approach to KDD considered in this research is the requirement that knowledge should be semi-automatically discovered and reused to update the Moderator's knowledge. In this regard, chapter 9 and 10 use the case studies from the construction industry and SMEs respectively to capture the knowledge to enhance the Moderator's capability using Knowledge Discovery Modules.

2.4 Structure of Thesis

The thesis is divided into four parts as shown in Figure 2-3. The main background topics of research mentioned in above section contribute to the design of the proposed system as described in chapter 7.

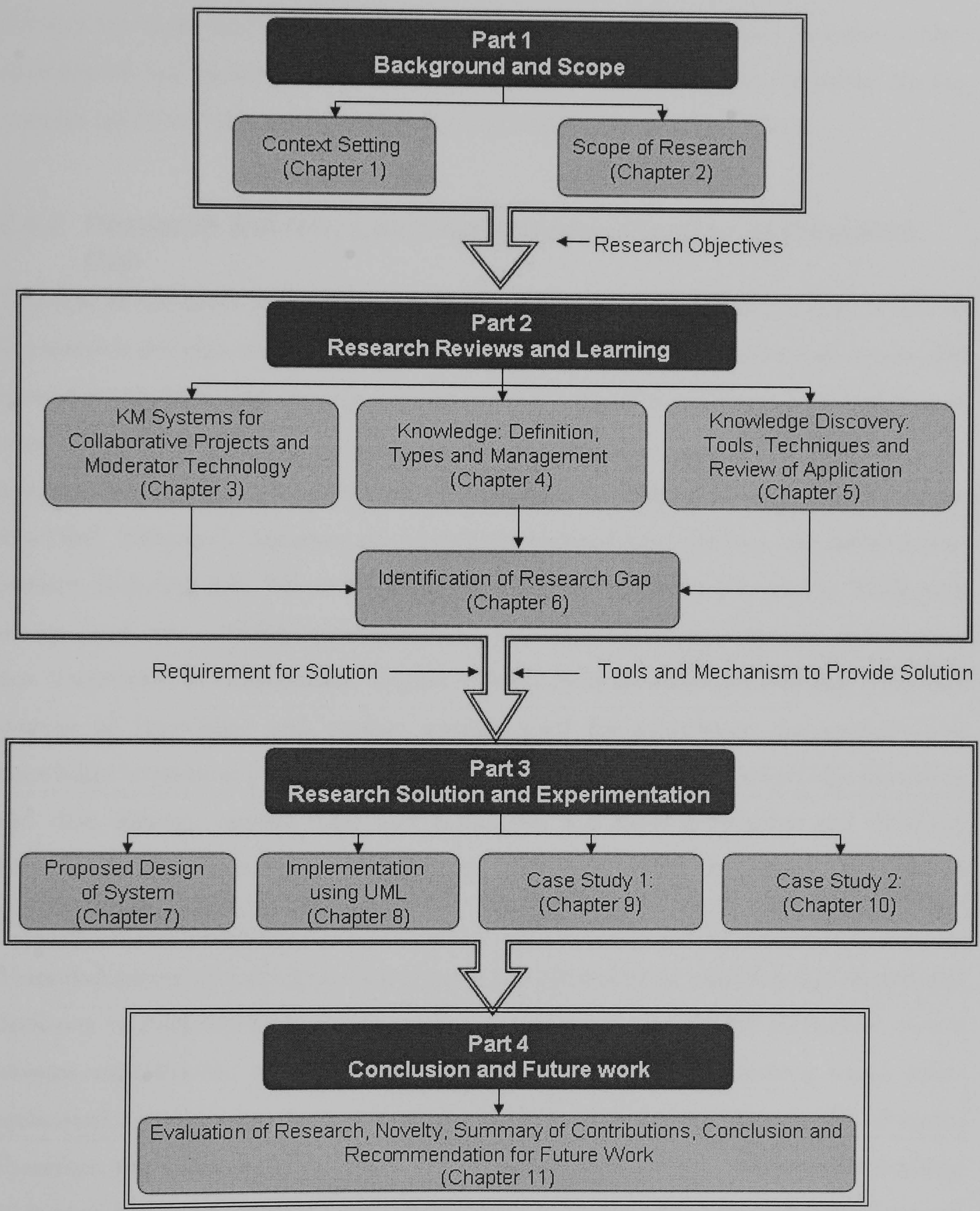


Figure 2-3: Structure of Thesis

2.4.1 Background and Scope

The background and scope of this thesis is comprised of two chapters (1) Context Setting and (2) Scope of Research, which establish the research domains and objectives. The first chapter sets the context and overviews the challenges faced by today's industries and identify the importance of Knowledge Discovery Integrated Moderator Services for quick and semi-automated knowledge acquisition. It also provides a brief objective of this research. Chapter 2 presents the scope of research including the key research questions with problem areas and contribution areas of this thesis.

2.4.2 Research Review, Learning and Identification of Research Gap

This part of the thesis consists of 4 chapters and presents a detailed literature review of the research domains mentioned in the above section. Based on the literature review and state-of-art findings, the existing research gaps are identified. Chapter 3, 4, and 5 each contribute to the thesis in a different way by providing the necessary understanding and learning that enable the requirements of the proposed original research solution to be identified. Chapter 3 discusses the knowledge management systems for collaborative projects including EE, VE and VO. A special focus has been put on the Moderator services and their need for ongoing learning, automatic knowledge updating and creating new knowledge. In consequence, chapter 4 defines and identifies the different types and sources of knowledge and various systems used for knowledge management and knowledge acquisition processes. Furthermore, chapter 5 describes knowledge discovery and data mining, various tools and techniques for their application and different functionality requirements for knowledge discovery from structured as well as unstructured data.

A detailed review of data mining applications in manufacturing identifies that operational databases of manufacturing can be used as a source to provide the moderator with a learning capability by discovering new knowledge. A novel text mining based semi-automated approach has been applied to identify the research gaps in this domain. Therefore, this chapter supports the validity of proposed system. Based on the findings of chapter 3, 4 and 5, chapter 6 identifies the research gaps and consolidates the focus of this research.

2.4.3 Research Solution and Experimentation

This part of the thesis consists of four chapters describing the framework, design, development and demonstration of the proposed system. Chapter 7 proposes KOATING framework for Moderators and its integration with the state-of-art moderators called Universal Knowledge Moderator. An illustrative example of a knowledge discovery process from structured data in the context of an e-supply chain has been presented. Chapter 8 discusses the modelling of the proposed system using use case diagrams, class diagram, sequence diagram, activity diagrams, component diagram and deployment diagram. Furthermore, chapters 9 and 10 discuss 2 industrial case studies illustrating how Knowledge Miners from the Knowledge Discovery Module can be used to extract knowledge from a variety of databases and sources. In chapter 9, the application of data mining and text mining on the Post Project Reports of construction projects has been demonstrated, whereas chapter 10 shows the enhanced capability of Moderators enabled by knowledge miners for awareness of business opportunities by discovering knowledge from invitation for tender projects.

2.4.4 Conclusion and Future Work

This part of the thesis consists of chapter 11, which evaluates the proposed research, discusses its novelty, summarizes the contribution, and concludes the thesis with recommendations for future work.

3

Knowledge Management System for Collaborative Projects and Moderators

During the exploratory period of this research, the author surveyed the literature in the area of various types of collaborative projects, architectures, decision support systems, frameworks and knowledge management systems to support collaborative projects. This chapter presents a summary of selected research in the problem domain areas. It also illustrates the evolution of Moderators from conception in 1996 to current state of the art.

Chapter Outline

3.1 Collaborative Projects: Types

Coming together is a beginning, Keeping together is progress, working together is success

Henry Ford

Collaboration between suppliers, manufacturers, and retailers is the key to success and can increase the number of satisfied customers by reducing lead times, improving service levels, decreasing cost and coping with unpredictable markets. Individual companies work together to form inter-enterprise networks across the product value chain. Collaboration between companies can facilitate both strategic and operational foci, allowing individual team members to exploit their core competencies, in turn strengthening the supply chain. Recent developments show that collaboration in manufacturing exists in many forms. Next, the author briefly describes some of the various kinds of collaborations such as collaborative product design, collaborative product development, CPFR, virtual factory, extended enterprises, virtual enterprise, and virtual organizations etc. that exists among enterprises.

In the 1990s, to meet the increased demand of mass customization, reduce product development time, and cutback product costs, the traditional approach of sequential product development was increasingly replaced by the collaborative product development approach. This approach comprises a multidisciplinary concurrent engineering approach in which organizations need to draw together representatives from all areas of the enterprise and to include suppliers who lie outside the enterprise. Some of the major issues faced by the companies include [35-37]: (1) understanding the customer's needs and meeting those needs by new product development (2) development of database models and intelligent problem solving systems (3) development of optimization strategies for evaluating manufacturing alternatives and tradeoffs for balanced designs (4) identification of problems and misunderstandings during design processes to avoid rework at later stages.

The success of the Collaborative Planning Forecasting and Replenishment process model (CPFR) developed by the Voluntary Inter-Industry Standards (VICS) association has already been proved by supply chain giants like Procter and Gamble, Wall-Mart and Kimberley Clark. CPFR is a type of collaboration where two or more parties in the supply chain jointly plan a number of promotional activities and work out synchronised forecasts, on the basis of which the production and replenishment processes are determined [7].

The concept of Virtual Factory (VF) and Virtual Manufacturing (VM) has emerged as a result of an increasing need for collaboration. In the late 1990s, the VF was a collaborative project (Virtuelle Fabrik) between 30 companies in Europe to develop new entrepreneurial solutions to gain competitive advantage for their industrial region through dynamic cooperation [182]. The basis for the creation of the VF was reliable behaviour among partners with "accepted rules of the game" and "defined roles and responsibilities" that individuals will play in the VF. Six different kinds of roles were assigned to the members such as broker, competence manager, project manager, in/outsourcing manager, network coach and auditors. In contrast, the target of the global VM project (TELEflow) in Europe was to develop a set of tools and methods to serve emerging market opportunities by dynamic configuration of resources and competencies in a global network [183]. Since European manufacturing companies are smaller in size in comparison to USA and Asia, they frequently have to choose a cooperative strategy to achieve global presence. The value system (VS) concept was used to represent the strategic network of globally distributed partners. The methods and tools developed for the design of VS encompass 4 views: namely, re-engineering the business system, information and communication technology, strategy based redesign of logistics, and organizational culture. The VS consists of 5 phases: (1) VS pre-phase, (2) VS Configuration, (3) VS Design, (4) VS Operation and (5) VS Disbandment. A detailed discussion about the functioning and concepts of VF and VM are detailed in [8].

Extended enterprise (EE) and virtual enterprise (VE) have evolved as advanced forms of collaboration to characterise the global supply chain.

Extended Enterprise: An EE can be viewed as a kind of "enterprise" which is represented by all those organizations or parts of organizations, customers, suppliers, retailers and sub-contractors, who contribute to the production and delivery of the product to the end user. It includes both the inbound supply chain and out bound logistics chain. The EE networks the activities of a number of entities to produce and market products and services. Therefore, it is important to have available advanced information and telecommunication mechanisms, such as network computing, to support the EE. The success of an EE largely depends on the efficiency and speed with which information is shared and managed between the partners. The major characteristics of an EE in the context of the coordination of relationships and communication between partners are delineated as follows [13]:

- The manufacturing enterprise focuses on its core business and technical activities and outsources non-core business activities to outside suppliers and other service providers. Outsourcing encourages both the manufacturer and its suppliers' competitive ability and enhances their mutual dependency.
- The manufacturer in a manufacturing centred EE develops long term relationships with key customers and treats them as important business partners beyond the product life cycle.
- Methods, business processes and technologies are available to support the partners for coordination and seamless information integration in the EE.

Virtual Enterprise: The VE may be defined as a formal but temporary collaboration between autonomous partners to conceive, design, market, manufacture, and deliver a specific product (or a range of products). Partners remain independent and possibly competitive in all other activities than those related to the product. In a VE, each intermediate stage is a supplier to its adjacent downstream stage and a customer to its upstream stage. Although the participants in the chain can play various roles, all their relationships are limited to supplier and customer roles. An open fast communication mechanism is essential for the companies entering into VE network activities, allowing members to jointly forecast, develop, produce, synchronize and deliver their product or services, and anticipate dynamic customer requirements. Some of the features for a VE are delineated as follows [9, 13]:

- Collaboration on the selected product is close, well coordinated and open so as to have the agility to maximize the exploitation of the product.
- Partners may or may not collaborate on any other products. Indeed, it may be possible that they will be competitors for different products or services. This implies a need for confidentiality, mitigating against the need for openness noted above and therefore, complicating collaboration issues. So not all the information can be shared.
- The life cycle of a VE is limited to that of the product or service only and partners have no obligation beyond that.
- Partners may join or leave the chain as the product life cycle progresses.
- The selection of VE partners including the initial design and formation of consortium is also an important issue.

Virtual Organization (VO): Virtual organization is a concept similar to a virtual enterprise, comprising a set of (Legally) independent organizations that share resources and skills to achieve the VO's mission/goal, providing to the outside world a set of services and functionality as if they were one organization, but a VO is not limited to an alliance of for profit making enterprises [184]. A VE is therefore a particular case of a VO. However, the terms VO and VE are generally used interchangeably by the research community. Similarly, VO Breeding environment (VBE) and Virtual Enterprise Network (VEN) are interchangeably used by the research community to represent an association (also known as cluster) or pool of organizations and their related supporting institutions that have both the potential and the will to cooperate with each other through the establishment of a "base" long term cooperation agreement and interoperable infrastructure. A VO may often be short-lived e.g., a team works together to complete a project taking a few weeks, whereas a VEN has a longer life probably stretching over several or more years. Several VOs can be formed from the pool of a VEN. Each VO can consist of a different groupings of members selected from a VEN. During the life span of a VEN, several VOs can be created or disbanded. These VOs may or may not follow a similar collaboration structure or pattern [30, 31, 180].

VO Lifecycle: from the knowledge requirement point of view, the life cycle of a VO can be divided into four distinct stages as: (1) VO Pre-creation, (2) VO Creation (3) VO Operation and (4) VO Termination. These stages are briefly discussed below:

- *VO Pre-Creation*: This stage involves the creation of a collaboration pool or VEN by prospective SMEs with the objective to collaborate with the members by offering its services when a new business opportunity arises. This stage also requires a cooperation agreement, establishing business rules, setting up an infrastructure and building trust between the members. During this period, a mechanism is required to raise the awareness of business opportunities. This might be on the basis of the core competencies of company and the complementary competencies of the collaborating companies.

- *VO Creation*: at this stage a VO is created by a strong partnership between a group of companies to respond and make decisions related to the business opportunity identified. At this stage, a knowledge management framework is required to raise the awareness of issues that may benefit or hinder the creation of the VO.

Table 3-1: Comparative study of various types of collaboration

	Virtual Factory	Virtual Manufacturing (Value System)	Extended Enterprise	Virtual Enterprise
Strategic issue	Stronger long term objective	Stronger short term objective	Stronger long term objective	Stronger short term objective
Partnerships purpose	Manufacturing System Design	Short term cooperation to deliver services across all phases of product life cycle	Long term business cooperation	Temporary working together to deliver product or services
Organizational stability	Stable organization.	Smaller companies with advanced competence management	Stable organization of companies across the product value chain	Dynamic organization of companies with core competencies.
Partner relationships	Temporary	Temporary and dynamic	Trust and mutual dependence for long term	Temporary and dynamic
Boundaries	Short term	Limited to product and services for short term.	The distinctions between responsibilities of the individual enterprises are fully blurred resulting in a single EE for long term.	The distinctions between responsibilities of the individual enterprises are partly blurred resulting in a single VE for short term.
Organization type	Manufacturing organization	Frequent project or niche market based	Product value chain based	Frequent project or niche market based
Co-ordination of partnerships	Main manufacturing company	6 defined roles and responsibilities	Usually manufacturer manages the partnerships	Frequently a broker manages the co-operation.
Information and communications technology (ICT)	Modelling and simulation tools	Sophisticated ICTs	Facilitated and enabled by ICT	Operation depends on sophisticated ICTs
Support during life cycle	Fully	Partly	Fully	Partly/fully

- *VO Operation*: this stage is often termed as the actual project period whereby the conceptualization to the delivery of the product or services to the customers is achieved within a promised time window. In terms of knowledge requirements, a set of functionalities need to be performed at this stage, including evaluating task status, indicating conflicts, identifying problem areas sharing knowledge between partners, addition or replacement of partners, evaluation of the project, improvement of project operations etc.

- *VO Termination*: This is the last stage of the VO lifecycle when it has completed it's task by delivering product/services to the customer and therefore it terminates. In terms of knowledge requirements, the "shared" knowledge is no longer needed in its shared form. Each company has gained knowledge through its experience or learning as part of the VO. There is a potential to learn from the operational databases of the collaborating firms to improve the process in future projects. In addition, mechanisms are required to capture the lessons learned knowledge.

Based on the aforementioned review of several types of collaborations and author's perception, a comparative study of various types of collaboration is presented in Table 3-1.

In this manner, one can see that the knowledge management frameworks are key to the success of collaborative projects of various forms.

3.2 Knowledge management Systems to Support Collaborative Projects.

In the era of knowledge economy and knowledge-based competition, an organization must be able to secure various types of knowledge assets and maximize their strategic values [3]. Collaborative projects within organizations or between organizations not only share the work based on each member's expertise, but also achieve a seamless information flow among the collaborative partners. Knowledge sharing is proven to improve the productivity and the decision quality of the participating organizations. Therefore, management of knowledge is essential and critical to achieve effective collaboration during a project [10].

Generally people from different disciplines, experiences and backgrounds try to work together in a collaborative project. There is a potential for misunderstanding or lack of awareness of the needs and interdependencies of each of the individual contributors. The importance of awareness and understanding of other partners' requirements in collaborative projects was highlighted in the mid 1990s. In this context, the development of Mediator was one of the earliest works by Gaines [11]. Mediator is an open architecture-based information and knowledge support system for geographically dispersed manufacturing processes from requirements through design, engineering, production, to maintenance and recycling. It has been implemented on the World-Wide

Web to provide a powerful tool for the support of the virtual manufacturing enterprises. At this time, the Engineering Physical Science Research Council (EPSRC) funded the MOSES project, in the UK which also introduced the concept of a specialist intelligent software system called a "Moderator". The "Moderator" system increases understanding and awareness in concurrent engineering teams [12].

Now a days, enterprises are moving from self-centered close enterprises to open networked enterprises such as VE and EE [13]. Frecon and Nou [14] developed a distributed virtual environment to support collaborative work in teams that are geographically scattered. They supported synchronous as well as asynchronous group collaboration. Zhou [15] presented a distributed information system architecture using CORBA and STEP standards to overcome the heterogeneity of partners and promote standardization respectively for VE. In a research project partly funded by the European commission, Slade [16] discussed the application of an extensible ontology as a principle for integrated information, knowledge management and knowledge sharing among geographically distributed collaborators of an EE. However, they mainly focussed on sharing, organization, interrelation, and visualization of documents for team members. Shafiei [17] proposed a multi-enterprise collaborative conceptual Enterprise Resource Planning (ERP)- Decision Support System (DSS) to maximize the intelligence density, improve the quality and visibility of information and to achieve the foundation for multi-enterprise collaboration. Panteli [18] developed a framework for understanding the dynamics of trust and conflict within the context of virtual inter-organizational arrangements. Ahn [19] presented a knowledge context model to facilitate the use of contextual information in virtual collaborative work. The benefits of this were suggested as evolutionary accumulation of knowledge aligned with collaborative activities, supporting the virtual team lifecycle, improving the understanding and searching the knowledge for similar context. A new software system called InteliTeam is developed based on a web-based collaborative system framework using a multiple perspective approach [20]. It consists of a group decision-making approach, many multiple criteria decision-making techniques, an intelligent system and advanced communication systems such as mobile e-services, wireless application protocol etc. Misono [21] proposed a distributed collaborative decision support system based on semantic web service technology to achieve the collaborative goal.

In the UK, the *Advanced Knowledge Technologies* Interdisciplinary Research Collaboration (AKTIRC) was £7.6 Million project funded by EPSRC to develop knowledge management technologies as part of the e-science initiative on grid computing. Collaborative Advanced Technologies in the Grid (CoAKTinG) aims to support distributed scientific collaboration by integrating meeting spaces, ontology annotated media streams from online meetings, decision rationale and group memory capture, issues handling, planning and coordination support, constraint satisfaction and rationale, and instant messaging presence [22].

Collaboration is the centrepiece of product development processes and involves multidisciplinary teams, functions and heterogeneous tools. Balasubraminiam [23] viewed new product development as a knowledge sensitive activity and identified problems associated with knowledge management issues of new product development by cross functional collaborative teams. They developed a prototype system that met the problem requirement and captured and managed tacit and explicit process knowledge and implemented it on a case study. Huang [24] proposed workflow management as a mechanism to facilitate team work in the collaborative product development environment where remote web-based decision support systems (TeleDSS) are extensively used by geographically distributed team members. Agents were delegated to manipulate the TeleDSS, share input and output data, and request remote services on the behalf of users. More recently, Mihaela [25] proposed a multi-resolution collaborative architecture (MRCA), based on a multi-agent co-ordination mechanism, as a web centric co-operative application in global manufacturing. The versatility of the proposed architecture and its recursive replication at all levels of resolution within the collaborative application were illustrated on a supply-chain example. Rodriguez and Al-Ashaab [26] proposed a web based knowledge driven collaborative product development system architecture to facilitate the provision of knowledge involved in product development. They detailed the research issues and industrial requirements for such a system. The implementation of the proposed system has been illustrated using a case study of an injection moulded product.

Recently, the Decision Support Systems (DSS) journal published a special issue dealing with knowledge management and collaborative work related issues in electronic business, such as organizational memory, virtual collaborative work, knowledge map, trust and conflict, intelligence of team, group support argumentation and distribution network optimization [10]. However, none of the work discussed above dealt with the knowledge discovery issues to aid the decision making process in a collaborative projects.

The concept of knowledge discovery based on semantic web technology has been applied in very few e-collaborative projects. Scotney [27] provided a flexible method of knowledge discovery from semantically heterogeneous data, based on the specification of ontology mapping. Wen [28] used web robots to discover the latest knowledge on the internet for better service of collaborative design. XML is used to make this system more efficient. Numata *et al.* [29] dealt with the knowledge conversion between and within tacit knowledge and explicit knowledge in new product development. They also provided an internet service for knowledge discovery and sharing for the development of an information system for knowledge amplification. However, these researches have focused on knowledge learned for distributed data processing and therefore do not cover the collaborative project issues and Moderator technology.

The European Collaborative networked Organizations LEADership initiative (ECOLEAD) was an "integrated project" funded by the European Commission within the 6th Framework Programme and involved 20 partners across 14 European countries. ECOLEAD aimed to create the strong foundations and mechanisms to establish the most advanced collaborative and network-based industry society in Europe. It suggested that in 10 years time, most enterprises will be part of some sustainable collaborative network that will act as a breeding environment for the formation of dynamic VOs capable of responding to fast changing market conditions. ECOLEAD mainly addressed the most fundamental and interrelated focus areas that form the basis of the dynamic sustainable networked organization, including the VBEs, dynamic VOs and the professional virtual community. ECOLEAD results claim to provide a set of tools to support various networked organizations. The set consists of a Dynamic VO creation assistance tool, a VO collaboration and performance measurement tool, a contract negotiation wizard tool, a VO management e-service tool, a collaborative problem solving support e-services tool, an advanced collaboration platform for professional virtual communities tools etc. [30, 31]. A detailed discussion has been published in research papers including [32, 33]. However, so far, none of the published literature as an output of this project deals with the knowledge discovery issues to support collaboration in an industrial context [34].

Extensive work has been carried out to provide solutions for collaborative and distributed product development. Li and Qiu [35] reviewed collaborative product development related works from three aspects (1) Visualization-based collaborative systems (2) Co-design Collaborative Systems and (3) Concurrent engineering-based collaborative systems. Based on the review of 130 papers, they concluded that the major issues for future collaborative system development are as follows:

- Integration of various collaborative manners and systems.
- Security and interoperability of collaborative systems.
- Efficient learning and sharing of knowledge and information for multiple application domains.

Current web technologies, such as the internet, intranets and extranets provide the platform independence needed for users to share data and information at any time within global network collaborations. However, the enormous amounts of heterogeneous data generated through globally distributed partners make it increasingly difficult to share and exchange information. In addition, none of the literature reviewed by Li and Qiu [35] is effectively used the "new knowledge" generated during the collaborative working. There is still a lack of support for the collaborative team members to analyze their operational data to generate potential knowledge for mutual support of the collaboration.

In addition, there is no doubt that considerable antagonism based on mutual distrust exists between the team members and a number of difficulties such as lack of awareness of each other's activities acts as a barrier to good collaboration between partners. The manifestation of distrust, lack of awareness, cooperative difficulties and distributed decision-making for own goal achievement, increase the risk of highly expensive conflicts.

The concept of Moderators was introduced to support collaboration and team working and to address the above mentioned issues. The main function of Moderators is to support the collaborative working team by raising the individual members' awareness of the needs and experiences of the other team members. Moderators have been successfully demonstrated in both the product design and manufacturing system design domains. The following section, describes the journey of Moderator from its first application supporting product design to the current state of the art supporting globally distributed partners on the semantic web.

3.3 Moderator Technology: History and Background

Moderators support global organizations and project teams working on a collaborative project to achieve their goals whilst reducing remoteness arising from distributing the team and improving the exchange of information between team members at different physical locations. Moderators support individuals to perform their individual roles from positions of strength and improve understanding through raised awareness of the needs of other contributors. A Moderator monitors an active project and alerts those collaborative team members who might be affected to changes in an object of interest. To date, all Moderators have been designed and implemented as modular specialist software systems, consisting of a moderation module, multiple expert modules and a knowledge acquisition module [36]. The development of Moderator Technology has been carried out in three phases.

3.3.1 Engineering Moderator (EM)

The concept of EM was first proposed in the MOSES (Model Oriented Simultaneous Engineering System, 1992-1995) research project as a support tool for design project teams [5]. It acted as coordinating software and addressed the fundamental requirements for the provision of support for a design team working in a concurrent engineering environment, by encouraging and facilitating communication between team members [12, 37, 43]. The MOSES architecture was based on the use of two information models, a product model and a manufacturing model, which could be accessed by an open set of application programs via an integration environment. More information about product model and Manufacturing model can be found in [38, 39] & [40] respectively. Design for Function and Design for Manufacturing were the application areas particularly studied during the MOSES project [41, 42].

The Engineering Moderator (EM) was included as a specialist manager or coordinating application, whose role was to drive concurrency within the MOSES system. To identify and signal conflicts in product design within the MOSES system, the EM needed to be capable of performing following operations [43].

- To promote communication and negotiation between design experts.
- To identify that a significant problem may have occurred in the design.
- To determine the course of action to be followed when a problem was identified.
- To maintain communication between interested design experts until the conflict of interest was resolved.

To raise awareness between members of the design team, the EM needed knowledge about the individuals within the team, for example, who they were, what elements of the design they were interested in and could contribute to. The EM stored knowledge in an adaptable format that could potentially work within different database applications.

As shown in Figure 3-1, the structure and content of all the three versions of Moderator mainly comprised of three main moderator's knowledge areas, Expert Module (EM), Knowledge Acquisition Module (KAM) and Moderation Module (MM). A flexible knowledge representation structure was therefore needed to support the implementation of these three knowledge types. The structure that was designed and developed was called the Knowledge Representation Model (KRM) [12, 37, 43]. Each element of knowledge from the simplest expression to the complex expert module shown in Figure 3-1, was designed as an object, which could interact with other knowledge objects. Storing the knowledge as persistent objects within the Object Oriented Data Base (OODB) enabled the Moderator to operate with a high level of flexibility in different computing environments. The details of these knowledge areas are discussed below:

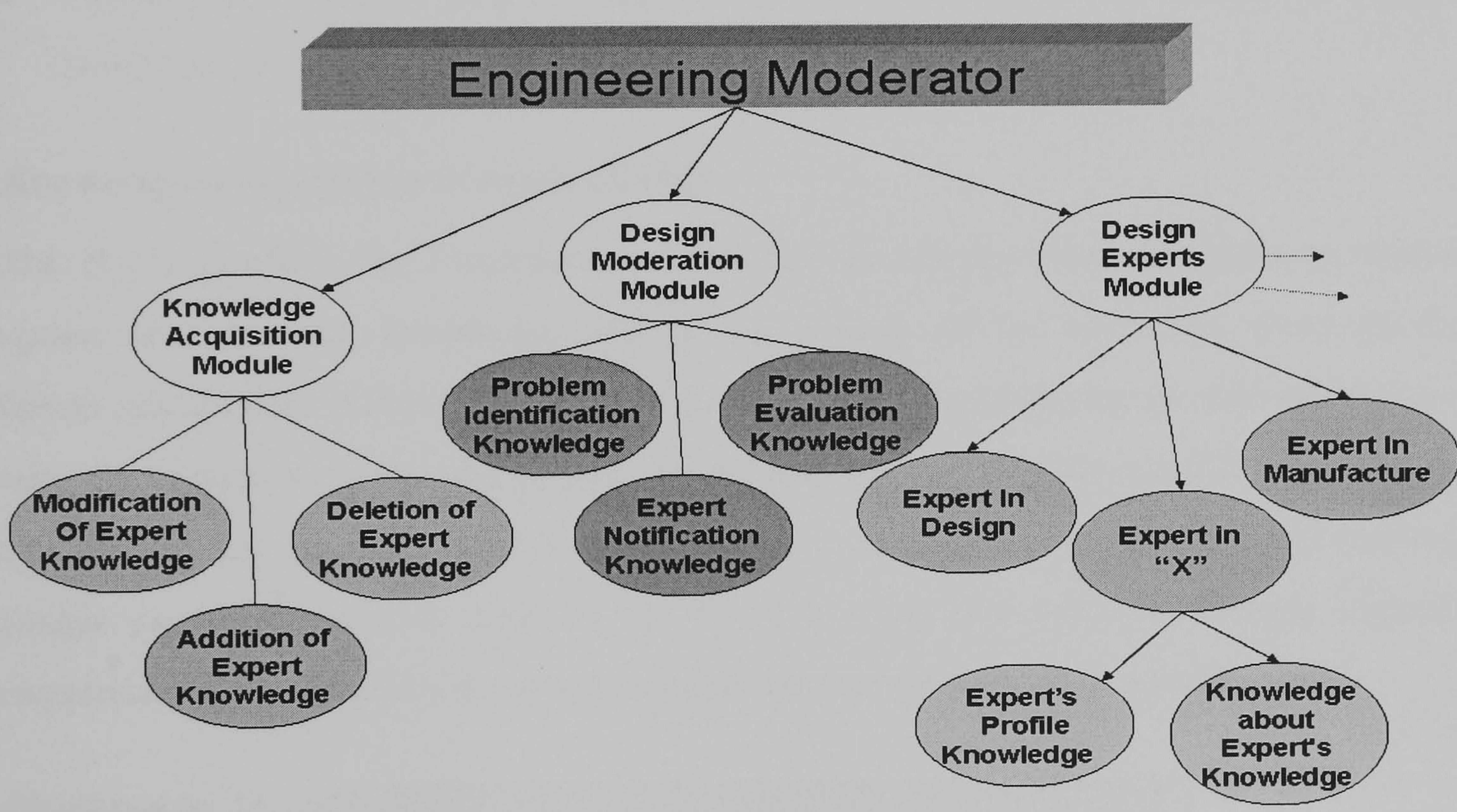


Figure 3-1: The knowledge areas and content of Engineering Moderator

Expert Modules (EMs)

Moderators consist of multiple expert modules where each expert module represents the expertise of an individual member of a team such as an expert in manufacture, an expert in design and so on. The key knowledge content of each expert module comprises of:

- *Personal Profile Detail:* This includes the name or identifier by which the design expert is known in the system, the expert's contact details, type of design expert, e.g. human or software, other information about role or purpose, etc.
- *Main Design Criteria:* This includes design items of particular interest. The main design criteria could also be thought of as which variables in the design are determined by, influence or constrain an individual design expert. This knowledge must be structured to enable the EM to decide whether or not design experts would be interested in the design step which has been taken and to assess the level of interest. It helps to decide whether a particular expert should be consulted and whether the design expert is likely to be able to identify any problems within the design, resulting from the change that has been made.
- *Communication methods:* This contains the information required to enable the EM to communicate with the design experts.

Knowledge Acquisition Module (KAM)

The *KAM* provides the Moderator with all the knowledge that it requires to create, update and remove knowledge during the course of its operation. Part of the functionality of the KAM is to add new design expert's modules to the EM whenever a new team member joins the design team. The content of this module can also be modified at any time as the project progresses. Also the knowledge required about a design expert should be captured in whatever way best suits the design expert's requirement, since the best approach depends on the specifics of the problem [44].

Moderation Module (MM)

In all the three phases of moderators' development, the major function of the MM is to identify the potential conflicts and to perform moderation activities. The MM operates continuously throughout the design project as changes to the design need to be examined to assess if any potential problem may be occurring. When a change is identified, the MM records it and does a quick check of the information stored for each EM to see if it is possible that the change may be significant to one or more of the team members. If the MM finds that the change is potentially significant, the moderator takes the further step of processing all the knowledge it has about the particular team member to which the change applies. Using all these sources of potential information, the Moderator will determine whether it is necessary to communicate with any of the other team members.

The MOSES design environment restricted the applicability of the Engineering Moderator by requiring the contributors to share a single object oriented database but did enable the feasibility of a Moderator to be demonstrated. The major sources of information provided for this section about MOSES project and EM are [5, 12, 36, 43].

3.3.2 Manufacturing Systems Engineering Moderator

The second phase of Moderator technology was started in the late 1990s during the IMS/ESPIRIT funded MISSION (Modelling and Simulation Environment For Design Planning and Operation of Globally Distributed Enterprises, 1998-2001) research project for manufacturing system design [4, 45]. The Manufacturing System Engineering Moderator (MSEM) was designed as a intelligent support system to monitor design decisions, evaluate their significance to individual project members and communicate them to any team members deemed necessary [51]. Methods for Manufacturing System (MS) design are not well understood or documented, although substantial amount of work exists in the parallel field of enterprise and manufacturing system modelling [41, 47-50].

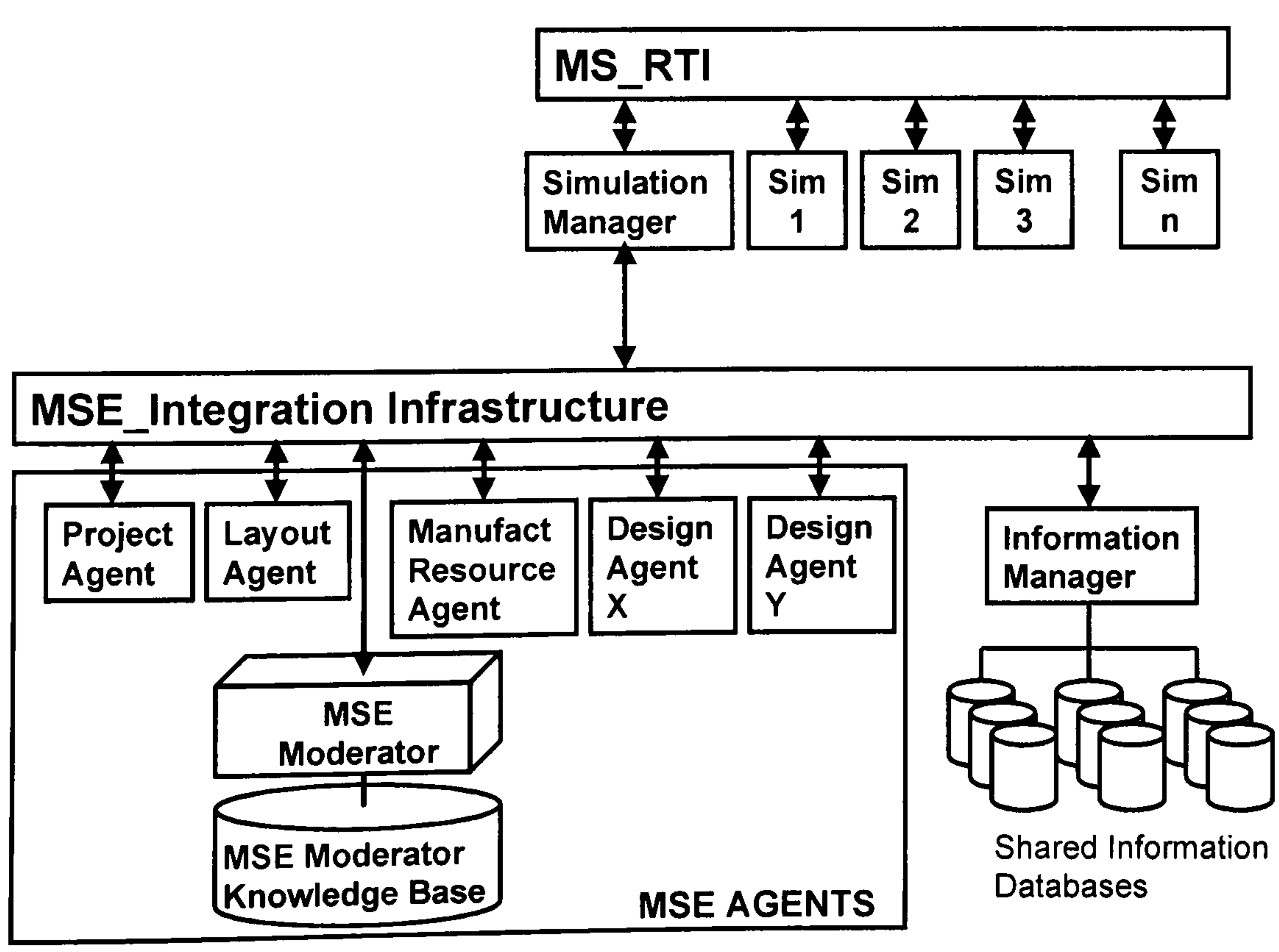


Figure 3-2: The Mission Modelling Platform

Designing a manufacturing system is a complex task and requires expertise from many different disciplines such as process selection, equipment selection, facility layout and many others to design or modify the system successfully [36].

The Mission project examined the process of designing a multi-site, globally distributed manufacturing system. The general goal of the research was to support the manufacturing system engineering (MSE) process by integrating the simulation and intelligent support system MSEM within manufacturing system design and operation as shown in Figure 3-2. The primary function of the MSEM was to support globally distributed MS design and enhance the degree of awareness, cooperation and coordination between members of the team within the MISSION environment. The basic functionality of the MSEM remained same as the EM. Additional necessary functionality for the MSE Moderator can be identified through examination of the activities in the moderation process. Like EM, MSEM mainly consists of three modules as briefly described below:

***MSE Design Agent Module*:** In the MISSION project, the term "MSE Agent" is used to refer to each combination of engineer(s) and supporting software performing an identifiable function in order to contribute to the developing MSE Design. Primarily, the knowledge of each MSE Agent is structured using the class Design Agent Module. Each DAM contains some static information about the MSE Agent, such as their name, contact details and the objects in which they seem to interest. The DAM is also linked to a knowledge base, which provides detailed knowledge for the MSE Moderator to use when determining whether an MSE Agent is interested in a particular change. Here it is important to mention that with advances in agent technology, the perception and definition of Agent changes. Therefore, the expert module has been used in recent versions of moderator to perform the same function. The content and structure of DAM remain same as the EM. The prototype implementation of MSE Moderator makes use of the flexible KRM concept [51], enabling knowledge about each MSE Agent's areas of interest.

The developing design is shared between Moderator and project contributors through the MSE_Intergration Infrastructure. As described in the MISSION project final report D24 [4], the Mission concept was embodied using the MISSION Modelling Platform (MMP). In the prototype MMP, information relating to the manufacturing system being designed was stored in an Oracle database. Access and sharing of the database was orchestrated and supported by an MSE agent called Information Manager (IM).

MSE Knowledge Acquisition Module: The content and functionality of KAM remain same as in EM. The MSEM must be able to acquire knowledge that is applicable to an enterprise, including the knowledge of all current MSE Design agents participating in the project. The processing of knowledge is achieved by message passing between instances of various classes, including Rule set, Rule, Condition and Action objects as shown in.Figure 3-3.

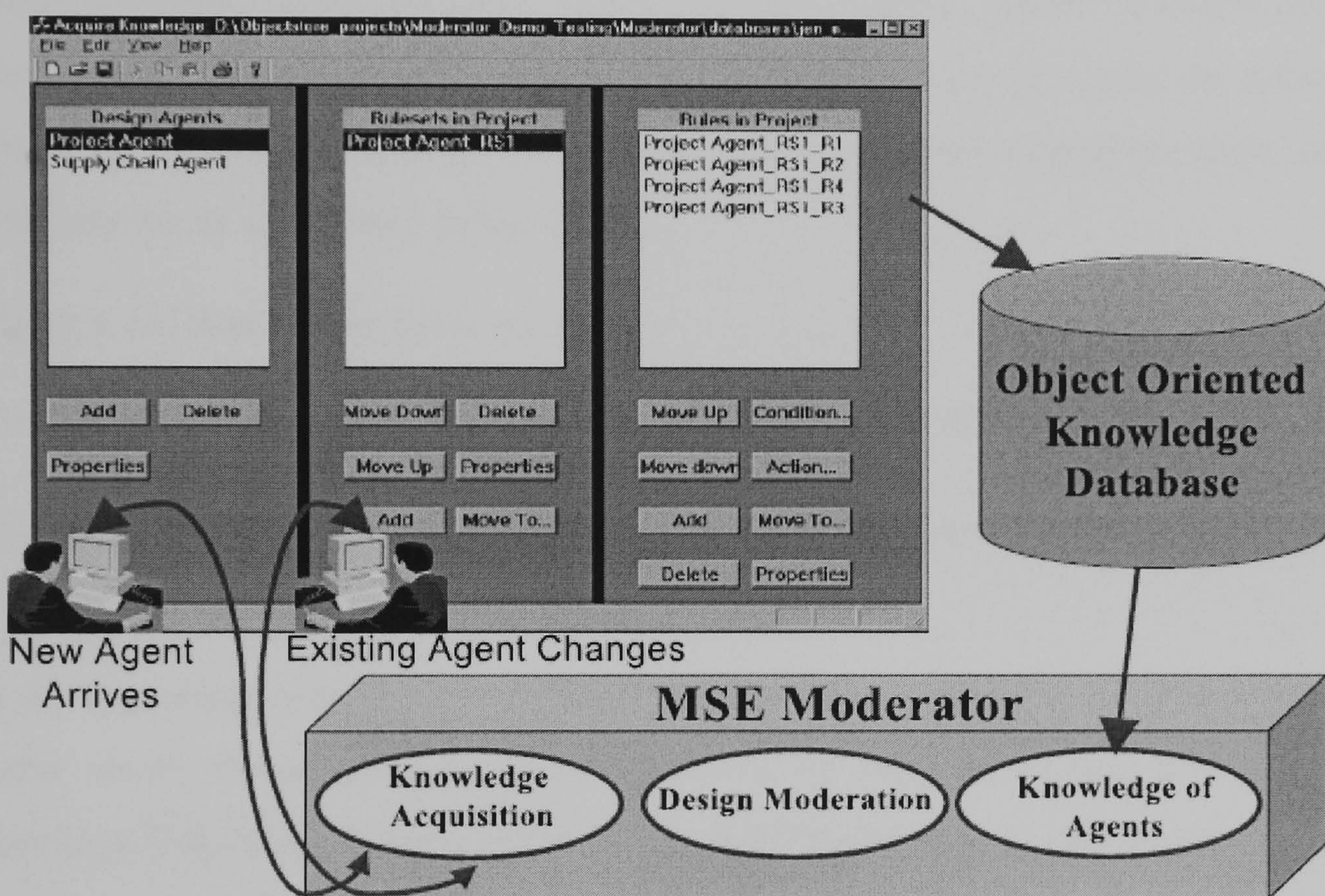


Figure 3-3: Knowledge Acquisition Module [51]

MSE Design Moderation Module: The moderation process was activated whenever a design decision is made. The making of a design decision was identified by a change being made to the design information within the shared database. In the prototype MISSION system whenever a change was made to the information held in the shared (Oracle) databases administrated by the Information Manager, the moderation process begins. The shared MISSION information model, managed by the IM, facilitated the communication activities in the MMP through the provision of communication message class objects which can be passed between the Moderator and MSE Agent through the MSE Infrastructure. The MSE Moderator also knew how to contact each MSE Agent. Hence the MSEM was able to communicate the detection of possible conflicts to all

MSE Design Agents who were needed to resolve the conflict. The MSEM remained in touch with the MSE Agent until the conflict is resolved.

More detailed information of MSEM can be found in [4, 36, 51, 52].

3.3.3 Extended Enterprise/E-Supply Chain Moderator

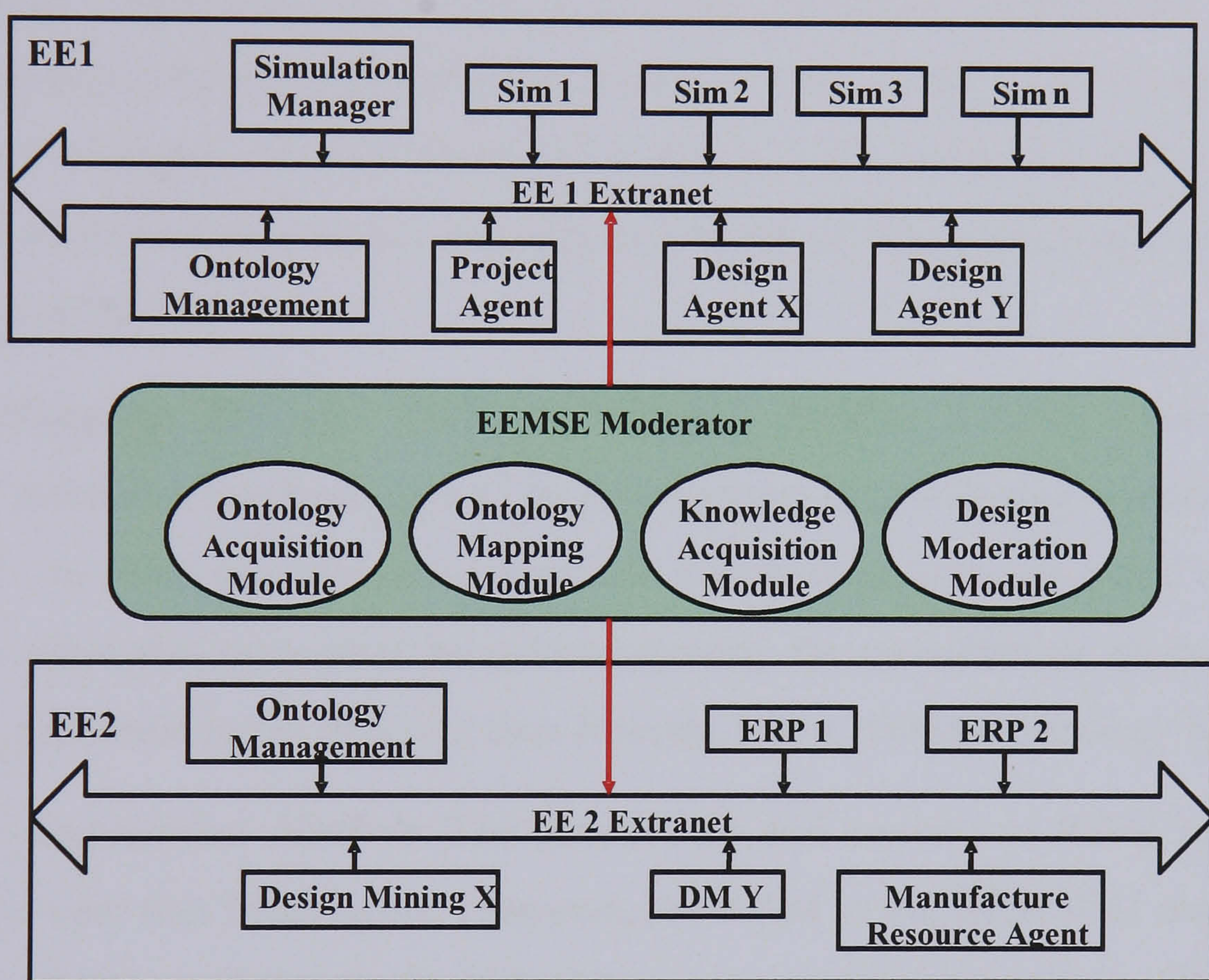
The third phase of the work extended the Moderator technology to make it applicable to extended enterprises, virtual enterprises and E-supply chain (E-SCM)[53] environments. It enhanced the semantic interoperability and reuse of knowledge resources within globally extended manufacturing teams or E- Supply Chain Management [55]. Heterogeneity of data causes a major problem in sharing and exchanging the information within a VE. The three most general heterogeneity problems among the multi database community are as mentioned below [36].

- Syntax: i.e. data format heterogeneity.
- Structure: e.g. schema heterogeneity in RDBs and OODBs.
- Semantic heterogeneity: e.g. differences or similarity in the meaning of data between the different component databases.

It is an important requirement to make design knowledge effectively accessible and sharable across virtual enterprise team members, by using an explicit and well defined terminology [54]. The adoption of "semantic web" technologies, like ontologies, content metadata and reasoning about conceptual knowledge, have been investigated by Lin *et al.* [55]to support a variety of the essential activities of evolving MSE Moderator knowledge management including knowledge retrieval, storage, sharing and moderation.

Lin and Harding [56] proposed an MSE Ontology Model to provide a common understanding of manufacturing related terms and therefore to enhance the semantic interoperability and reuse of knowledge resources within the global extended enterprises. They used protégé to automate the process of building domain specific knowledge acquisition and knowledge based systems. The MSE ontology defined in Lin's research has been converted into a formal ontology language, a resource description framework and a resource description framework schema, to serve as ontology metadata that may be used to create delete, modify and query the MSE ontology. A method of ontology inference was proposed by building sets of declarative mapping rules that could be applied to map all the shareable semantic metadata between the common MSE ontology and any manufacturing system models to enable semantic and syntax integration [54, 56, 57]. The MSE Ontology model has been captured in seven top level classes *(Project, Flow, Extended_Enterprise, Enterprise, Process, Resource, and Strategy)* using the knowledge and experience published in manufacturing system information models [41, 42, 58-60]. A detailed study about classes, sub-class hierarchy and properties of the MSE meta data can be found in [6].

The main function of the EEMSEM is to coordinate expertise and to support the role of concurrency within the engineering activities of an inter-enterprise environment. Two major differences that exist between MSEM and EEMSEM are delineated in Figure 3-4.



:

Figure 3-4: The architecture of EEMSE Moderator [55]

- Design Information changes (including addition or deletion) are expressed in different languages and terminologies.
- Information or knowledge of what team participants consider as important aspects of the design (e.g. key variables or values) is expressed in different languages and terminologies.

The first difference directly affects the EEMSEM's design moderation process and the second difference affects both the design moderation process and its knowledge acquisition process.

The EEMSE Moderator operates on an open extranet-based platform to support the execution of globally distributed MSE web applications on the WWW. It includes four major modules: Ontology Acquisition Module, Ontology Mapping Module, KAM and Design Moderation Module, as shown in Figure 3-4 [57].

The design of KAM and Design Moderation Module are broadly similar in implementation to their counterparts in MSE Moderator in the MISSION Project, however details of all four modules are given below:-

Ontology Acquisition Module: The first step in developing the Ontology Acquisition Module (OAM) is to acquire the common ontology and metadata created by a particular VE team group. Additionally the common ontology should be extensible so that it can be changed as necessary when the structure of project team in the VE's or supply chain's environment is changed. Further details are available in [6], where the model is used to illustrate the manufacturing system domain and cover all the terminology aspects and needs for an e-SCM.[36].

Ontology Mapping Module: Ontology Mapping Module (OMM) enables all the participants' individual terminologies to be translated to the mediating metadata created in the OAM. In order to perform the semantic match translation, the initial step of the OMM is to solve the syntactical level heterogeneity by transforming all participants' information, presented in the different data formats, into a standard ontology format.

Knowledge Acquisition Module: The functioning and content of KAM remains the same as in the very first Moderators. However, the KAM in the EEMSEM translates the knowledge into a neutral format to deal with any syntactic and semantic differences in the terminology that may be used by different project team members. This is achieved through the OAM and the OMM and then this knowledge about design experts can repose as mapped results in the Knowledge Rules Ontology Server, as shown in Figure 3-5 .

Design Moderation Module: The Design Moderation Module (DMM) is used to assist and keep track of changes made to the MSE design and identify whether any current design experts may be interested in the change. The DMM should be activated whenever a change is made to any information that may be related to interests recorded in any design expert's module. These changes can then be passed through the translation process, through the OAM and the OMM and into the VE Ontology Server as shown in Figure 3-5. If information changes in the VE Ontology Server have been identified, the

DMM will be notified of the change and also connected to the Knowledge Rule Ontology Server which is needed for the moderation process of conflict detection [36].

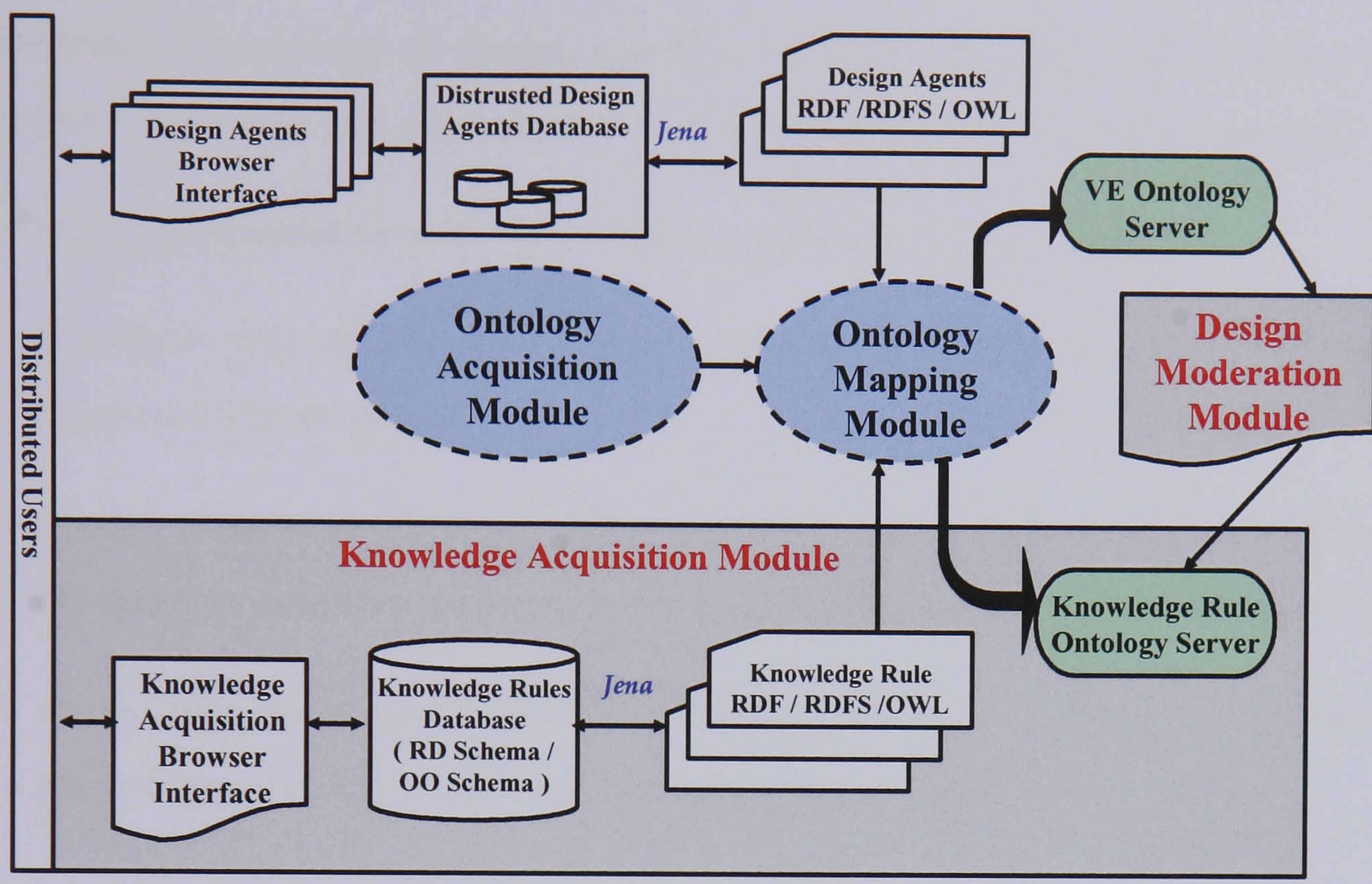


Figure 3-5: The structure of the KAM and DMM in the EEMSE Moderator

3.4 Ongoing Development: Collaboration Moderator Services.

The advancement of Moderator Technology continues with the recent seventh Framework Programme funded project "**S**upporting High**lY** Adaptive **N**etwork **E**nterprise Collabo**R**ation Throu**G**h Semantical**lY** Enabled Knowledge Services" (SYNERGY). SYNERGY is a multi-million euro project funded by European commission between 8 institutes and industries across Europe. The overall aim of SYNERGY is to enhance support of networked enterprises in the successful, timely creation of, and participation in, collaborative Virtual Organizations (VO). SYNERGY aims to provide an infrastructure and services to discover, capture, deliver and apply knowledge relevant to support collaboration throughout the life cycle of a Virtual Organization. The main objectives of this ongoing project are to [61]:

- provide semantically ontology based models of knowledge structures for collaborative working;
- develop a self-oriented self-adaptive SYNERGY holistic solution for knowledge-based collaboration services;

- facilitate the testing and evaluation of the effectiveness and efficiency of this solution on industrial case studies.

In order to contribute towards achieving the overall objective, advances are going on in Moderator Technology to design, specify and further develop Collaboration Moderator Services. The aims and objectives of collaboration Moderator services are [61]:

- raising awareness of potential business opportunities for user;
- alerting a user of short-falls in the competencies required to avail the user of the business opportunity;
- raising awareness if other users/partners are available with complementary or equivalent competencies to participate in exploiting the business opportunity;
- monitoring an active project and alerting changes in objects of interest to those project partners who might be affected by the changes.
- raising awareness of lessons learnt and updating knowledge, etc.

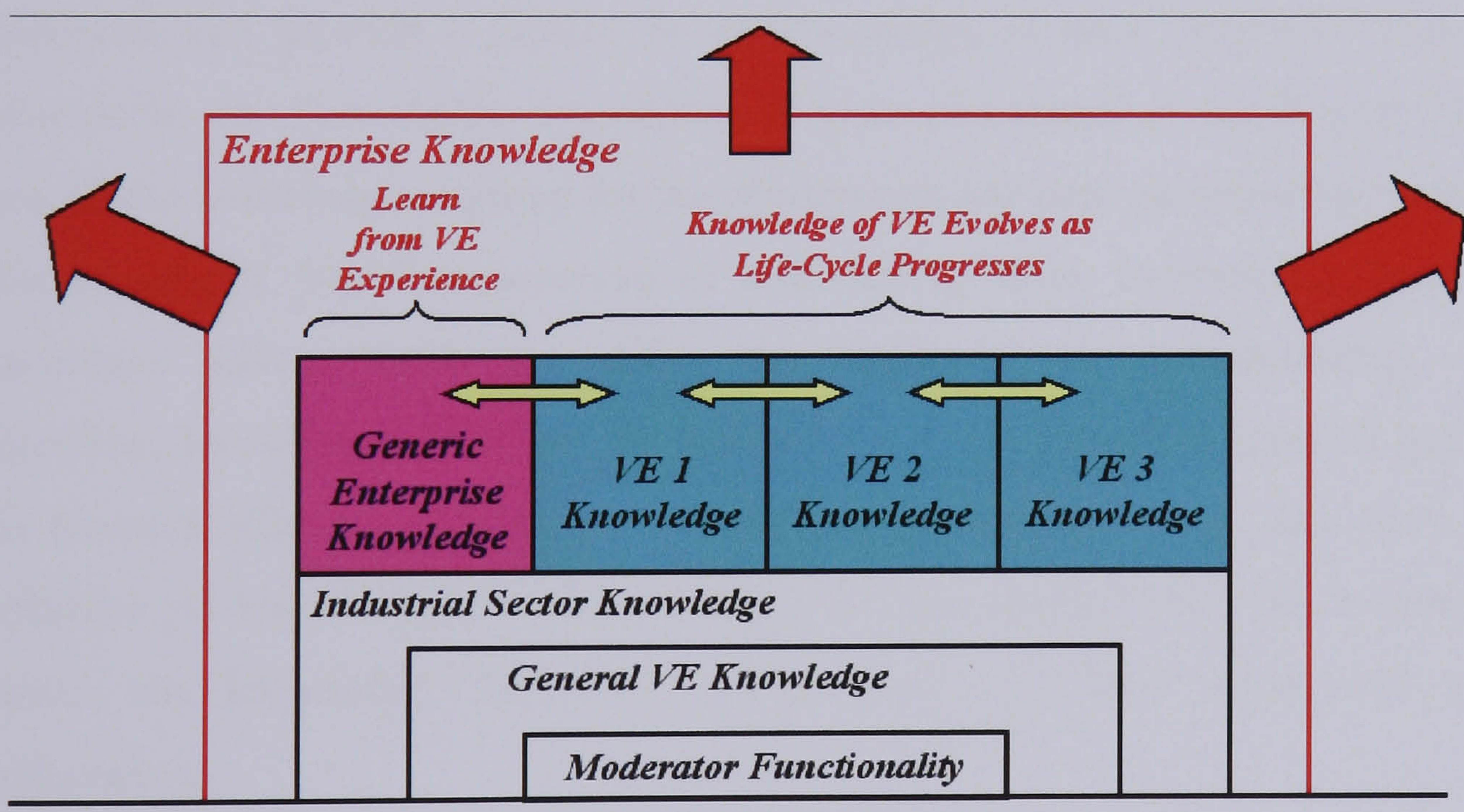


Figure 3-6: Evolving Moderator Knowledge

To achieve this, tools and methodologies are needed for acquiring, retaining and accessing the expanding range of knowledge within individual enterprises and in VOs to enhance the efficiency and productivity in collaboration and maximize the benefit of both long and short term knowledge sharing. Enterprise knowledge relevant to the Virtual organization life cycle of collaborative ventures will include, but is not limited to, (1) Enterprise Core competency, (2) process knowledge for VO formation, (3) Process knowledge for partner selection (4) VO operation management knowledge. Therefore, it is necessary to identify, share and apply different types of knowledge between partners and often across industrial sectors to gain the competitive advantage in the market.

3.5 Evolving Moderator Knowledge Areas

In all the versions of Moderators, more sophisticated knowledge about all the partners in the collaborating team is required to accomplish the Moderator's objective. Moderator knowledge structuring, sharing and evolution become even more important when teams come together from different companies and different locations to create, operate and manage the virtual enterprise. In such circumstances, the core functionality of the moderator remains the same but its knowledge must be extended to provide greater understanding of the extended environment in which it is operating. Figure 3-6 shows the schematic representation for evolving moderator knowledge. More details of the evolving knowledge moderator areas can be found in [36].

In this manner, it can be seen that management of different kinds of knowledge are key to the success of the Moderator's functions. The quality of the support that any Moderator can provide is limited by its knowledge of team members' knowledge, as collected by the Knowledge Acquisition Module and stored in the Expert Module. To date, all the knowledge acquired for the Moderators has been provided by human experts after looking at their best practices or interviewing them. However, huge amounts of experience and expertise lie within the databases' of manufacturing operations. Therefore, knowledge discovery for moderation as an identified research gap is key to this research [62]. To develop a framework to capture knowledge, one must know the definition of knowledge, its various types, its management and the systems used to capture the knowledge. The next chapter deals with these issues and reviews its applications.

4

Knowledge: Types, Management and Acquisition

During the initial phase of this research, the author explored issues related to knowledge, such as the definition of knowledge, the difference between data, information and knowledge, various types of knowledge, sources of knowledge, various frameworks and techniques related to knowledge management applications. The knowledge acquisition process and the need of automated knowledge acquisition methodology such as knowledge discovery to derive knowledge from data of various kinds are core issues in this research.

Chapter Outline

4.1 Towards an Understanding and Definition of Knowledge

Knowledge is Power

Francis Bacon

Knowledge and expertise are arguably the most valuable assets of any organization. The power of knowledge is a very important resource for preserving valuable heritage, learning new things, solving intricate problems, creating core competencies, and initiating new situations for both individuals and organizations now and in the future. Liao [63] categorized knowledge as explicit and tacit and recognizes the creation of knowledge as a spiraling process of interactions between explicit and tacit knowledge [64]. There are many challenges for researchers entering the areas of knowledge management, knowledge sharing and artificial intelligence. Indeed there is not even agreement between researchers on "What knowledge is?". Several authors have defined Knowledge from various perspectives ranging from the practical to conceptual to the philosophical and from narrow to broad scope [65]. The following definitions are from a knowledge management (KM) perspective.

- Knowledge is information that has been organized and analyzed to make it understandable and applicable to problem solving or decision making – [66]
- Knowledge encompasses the implicit and explicit restrictions placed upon objects, operation, and relationships along with general and specific heuristics and inference procedures involved in the situation being modeled –[67]
- Knowledge consists of truth and beliefs, perspectives and concepts, judgments and expectations, methodologies and know-how.- [68]
- Knowledge is the whole set of insights, experiences, and procedures that are considered correct and true and that therefore guide the thoughts, behaviors, and communication of people.—[69]
- Knowledge is reasoning about information and data to actively enable performance, problem solving, decision making, learning and teaching. –[70]

The next section differentiates between data, information, and knowledge

4.1.1 Data, Information, and Knowledge

Some people use the terms data, information and knowledge almost interchangeably, whilst others would consider data to provide less benefit and have less value than knowledge. Before proceeding to capture and manage knowledge, the difference between data, information and knowledge must be clarified. Spiegler [71] considered two models of knowledge, the first following a conventional hierarchy of data, information and knowledge and the second using a reverse hierarchy where knowledge precedes the data to information process. They defined a lifecycle of knowledge generation where a recursive and spiral model links the three. "Yesterday's data are today's information", and tomorrow's knowledge, which in turn recycles back through the value chain into information and then into data. They mentioned that the objective of data mining is to detect, interpret, and predict qualitative and quantitative patterns in *data*, leading to *information* and *knowledge*. Any definition of knowledge must start from data and information. A parallel inference has been made that if data becomes information when value is added, then information becomes knowledge when it adds insight, abstraction, and better understanding.

Several authors [65] draw distinctions between data, information and knowledge as follows:

1	Data	Facts, numbers, texts, images, or sounds (+ interpretation + meaning + organized =)
2	Information	Formatted, filtered, and summarized data (+ action + application =)
3	Knowledge	Instincts, ideas, rules, and procedures that guide action and decisions

Aamodt and Nygard [72] discussed that the role of knowledge is to play an active part in the process of transforming data into information, deriving other information and acquiring new knowledge. It can be summarized as follows:

- To transform data into information – referred to as data interpretation
- To derive new information from existing- referred to as elaboration
- To acquire new knowledge – referred to as learning

Researchers have also classified knowledge into knowing-what and knowing-how. In conclusion, it can be summarized that knowledge is the production of new facts, or even that the production of new knowledge, is a recursive or reflexive process that is , infact, infinite. Such a model may be a partial melding of two knowledge models: data → information→ knowledge, and the reverse hierarchy of knowledge preceding information and data.

4.2 Types of Knowledge

Generally knowledge can be categorized as follows [3, 73]:

- *Explicit knowledge.* This is formalized knowledge, easily expressed as words and numbers, shared in the form of data, principles, procedures, facts, figures, rules or scientific formulas. This kind of knowledge can be readily transmitted between individuals formally and systematically. As knowledge becomes more explicit, it becomes more stable, more routinized, and thus less complex in terms of its observed behaviour. Explicitness moves knowledge away from the knowledge pole and towards the data pole [74]. Generally, explicit knowledge is associated with data through business processes, and they may be implemented in an enterprise formation system, through primitive operations such as create, read, update and delete (CRUD). However, as operations become more complex and subject to interpretation, they move away from the data pole towards the knowledge pole, and include complex processing algorithms.
- *Tacit knowledge* includes experience, ideals, emotions, intuitions, and insights, and therefore is subjective and cannot be expressed. Tacit knowledge is demonstrated by the application of knowledge – the interrelationship between the content of knowledge and the associated behaviours, experiences, and feedback. Tacit knowledge has two dimensions: technical and cognitive. Technical knowledge is best described as a craft-like skill or "know-how". It is highly dependent on experience. Cognitive knowledge is composed of schema, values and beliefs. Tacit knowledge is the foundation of innovation and creativity – all knowledge originates as tacit knowledge in the heads of people. In terms of accessing the knowledge, Liebowitz [65] defined three stages of accessibility as follows:
    - *Tacit* (Human mind, Organization): Accessible indirectly only with difficulty through knowledge elicitation and observation of behaviour.

- *Implicit* (Human mind, organization): Accessible through querying and discussion, but informal knowledge must first be located and then communicated.
- *Explicit* (Documents, computer): readily accessible, as well as documented into formal knowledge sources that are often well organized.

- *Social and Individual knowledge.* The social knowledge is shared, and may be either explicit or tacit. The most obvious kind of shared explicit knowledge is referred to as *scientific knowledge*, which is generally available. But, knowledge may be both shared and tacit. This is known as *communal knowledge*; its social and tacit dimensions arise from the fact that it is taken for granted among members of an organization. Individual knowledge is always tacit. An individual may be aware of his or her knowledge, known as *conscious knowledge*, or take it for granted; in that case it is known as *automatic knowledge.* Spender [73] suggests that competitive advantage arises from the interaction of the four types of knowledge (scientific, communal, conscious, and automatic).

In recent decades, the conversion of data into information and knowledge, and managing knowledge has become important issues. The Knowledge Management (KM) community has developed several frameworks, architectures, methodologies, tools, functions, and real world implementations in terms of demonstrating KM technologies and applications.

4.3 Knowledge Management

KM involves the identification and analysis of required knowledge assets and knowledge-asset-related processes, subsequent planning, and finally, control of procedures, which are needed to develop both the assets and the processes in order to satisfy organizational goals [75]. KM can establish routines for identifying knowledge, as well as supporting the experts who have possession of the knowledge [76]. KM recognizes the process of mapping out where knowledge resides and identifying the conditions that foster its generation and re-use. The definition of KM, can be found in "Knowledge Management Handbook" [65] and Harvard Business Review on "Knowledge Management", as well as the recent review papers of [63, 77, 78, 79, 80, 81].

As a part of KM research, the next section will focus on surveying knowledge management framework developments through a literature review. KM technologies have been classified into seven categories. These are KM frameworks and applications, knowledge-based systems (KBS), information and communication technology (ICT), expert systems (ES) and applications, database technology and applications, modelling and applications and knowledge discovery in databases together with their applications on different research and problem domains.

4.3.1 Knowledge Management Frameworks and Applications

Recently, researchers have developed a set of management definitions, concept activities, stages, circulations, and procedures, all directed toward dealing with objects in order to describe the framework of knowledge management as a KM methodology [65][78]. A conceptual framework presents KM as consisting of a repertoire of methods, techniques, and tools with four activities such as Review, Conceptualize, Reflect and Act, performed sequentially. Furthermore, Rubenstein-Montano *et al.* [82] developed a systems thinking framework for KM, providing suggestions for what a general KM framework should include. Also, the emergence and future of KM and its link to artificial intelligence has been discussed in [83].

Several authors including [84], [83], [85] have implemented KM frameworks and methods for knowledge creation, knowledge assets, knowledge inertia methods and techniques, KM development and history, organization learning, organizational innovation, organization impact, intellectual capital, and strategy management.

4.3.2 Knowledge based Systems and Applications

The most common definition of a KBS is human-centred, highlighting the fact that KBS have their roots in the field of artificial intelligence (AI) and that they attempt to understand and implement human knowledge in computer systems [86]. The four main components of KBS are usually distinguished as: a knowledge base, an inference engine, a knowledge engineering tool and a specific user interface [87]. On the other hand, the term KBS includes all the organizational information technology applications that may be helpful to manage the knowledge assets of an organization, such as Expert Systems (ESs), rule-based systems, groupware, and database management systems (DBMS) [88]. In the context of manufacturing, KBS have been applied on problem domains as shown in Table 4-1. Most of the researchers, as mentioned above, have identified the major challenges as follows with regard to KBS:

- How knowledge is used and produced within an organization
- What must be done so that KBS can earn their place as tools for KM? What technology do KBS support?

- How to implement KBS in a specific problem domain.

Whilst the published literatures on knowledge based systems and applications is largely from academics sources, many of the projects represented in the following Table 4-1 are supported by industrial case studies and examples showing the research's industrial applicability.

Table 4-1: Knowledge based systems and their industrial applications

Knowledge based system and application	***Authors with year of publication***
Engineering failure analysis	[89]
Production Management,	[90]
Decision support system	[91]
Knowledge management	[92]
Knowledge representation,	[93]
Decision making and learning	[94]
Plant process control	[95]
Concurrent system design	[96]
Process Monitoring	[97]
Material Selection	[98]
Computer aided process planning	[99]
Customer support	[100]
Prediction	[101]
Mould design system	[102]
Concurrent engineering design	[103]
Project Management	[104]
Autonomous Vehicle planning	[105]

4.3.3 Information and Communication Technology

An information and communication technology (ICT) infrastructure provides a broad platform for exchanging data, coordinating activities, sharing information and supporting globalized commerce. The internet is a kind of ICT that can be combined with other network technologies and services such as Intranet, extranet, virtual private network (VPN), and wireless web, to construct a digital environment to consistently create new knowledge, quickly disseminate it and embody it in an organization/enterprise. [106] discussed the KM software tools of ICT in terms of their origin and applications.

ICT enables KM activities for collaborative decision support, information sharing, organizational learning, and organization memory [107]. Now a day, ontology is an integral part of ICT which enables other types of knowledge integration through different representations of the same type of knowledge at different levels of formalization.

4.3.4 Semantic Web Applications

The semantic web has added a new level to web services [7]. The concept of the semantic web is, without any doubt, gaining attention in both industry and academia. It is defined as the conceptual structuring of the semantics of data in a machine readable way that enables web entities to interoperate with each other, dynamically discover resources, extract knowledge, and solve problems [7]. The declaration of domain knowledge in a machine readable way enhances the understandability of disparate information. Although the semantic web has been proposed as the way to provide machine understandable and interpretable semantics for computers, it is unable to provide a suitable infrastructure for interaction and encapsulation of processes and skills. Therefore, web services are needed to address these requirements directly by providing an interface for encapsulating and unifying the technologies. The concept of semantic web services was proposed for enabling a system automatically to discover, invoke, compose, and monitor the associated processes and system. Recently, semantic web services have received a great deal of attention for their ability to make interactions among firms more flexible and automated. They have been used to annotate the information available on the web for automated processing and information integration. Semantic web services have been used in the following research directions [7]:

(a) Provide common syntax for machine understandable statements;

(b) Establish common vocabulary;

(c) Develop repository of computer manipulatable data having well defined semantics.

A detailed review of Ontology and Semantic web applications in the context of manufacturing enterprise has been carried out. This work is accepted for publication [7] and an abstract of the paper is attached as appendix 1.

4.3.5 Expert Systems and Applications

Expert Systems (ESs) are a branch of applied artificial intelligence. Expert systems are knowledge intensive computer programs that capture human expertise in a limited domain of knowledge [88]. Usually, ESs capture the human knowledge in the form of a set of rules. An ES can assist decision making by asking relevant questions and explaining the reasons for adopting certain actions. ESs for representing knowledge include knowledge based systems, rule based systems, knowledge frames, expert system shells, inference engines and case based reasoning [63, 108]. Sometimes, ESs are integrated with other AI methods such as neural network, fuzzy logic, genetic algorithm and intelligent agents using their functions of automated reasoning and machine learning. Object-Oriented (OO) technology provides an alternative approach to ESs that combines, through encapsulation, knowledge and procedures into a single object with specific procedures that operate on data, where the object combines data and program code. Object oriented programming leads the ES towards the fourth generation language, which provides a user friendly structure and environment.

4.3.6 Database technology and Applications

A database is a collection of data organized to efficiently serve many applications by centralizing the data and minimizing redundant data. A database management system may be defined as software that permits an organization to centralize the data, manage them efficiently, and provide access to the stored data by application programs. Large databases make knowledge computationally expensive. Therefore, modern database technologies need to process large volumes, multiple hierarchies and different data formats to discover in depth knowledge from large databases. These include multi-dimensional data analysis, online analytical processing, data warehouses, web and hypermedia databases. Recently, databases and architecture design have been used with other methodologies for implementing ontology creation heuristics, and intelligent agents into databases conceptual modelling and knowledge repository domains. Database technology has been applied in several domains including, hierarchical modelling, knowledge refinement, machine learning, error analysis, knowledge representation, knowledge discovery, database design, knowledge reuse, and web services [63].

4.3.7 Modelling and Applications

The modelling technology of KM includes the quantitative methods for exploring the issues of knowledge discovery, knowledge classification, knowledge acquisition, learning, pattern recognition, artificial intelligence algorithms, and decision support. Some methodologies are presented as examples of fuzzy logic, including process modelling, cognitive modelling, pattern language, system dynamics, decision trees, knowledge value modelling, genetic algorithm/programming, intangible asset modelling, and mathematical modelling etc., [63]. Modelling technology has been recognized as an interdisciplinary methodology of KM in order to build formal relationships with logical model design in different knowledge/problem domains. Modelling application areas of KM includes Knowledge discovery, knowledge classification, learning, business values, pattern languages, knowledge acquisition, cognitive modelling, value of knowledge, process reengineering, intellectual capital, intangible assets, and knowledge transforming etc.

UML is generally used to write software blueprints that are used to model, specify, and visualize various types of software-intensive systems or its artefacts. It provides the ability to capture the characteristics of a system by using notations and several diagrams. UML diagrams could be adapted for developing knowledge-based systems [109]. Afshar [110] suggested that until now, the process of knowledge acquisition remains a great barrier in knowledge engineering. Hakansson [111] suggested Unified Modelling Language (UML) to represent the expert's domain knowledge, knowledge acquisition and related case studies in this area are rare. In addition, none of work reviewed in the domain of UML applications have modelled the automated knowledge acquisition systems.

4.4 Knowledge Acquisition

Knowledge acquisition is a key component of the KMS architecture. Knowledge acquisition includes the elicitation, collection, analysis, modelling, and validation of knowledge for knowledge engineering and KM initiatives. Any application constructed will depend directly on the quality of the knowledge acquired. Therefore, it is vital to determine where in the organization the knowledge exists, how to capture it, and how to disseminate it throughout the enterprise. Knowledge acquisition is the most expensive and challenging task in building and maintaining the knowledge in a KMS due to elicitation of knowledge from human experts [112, 113]. Knowledge acquisition is the process of acquiring, organizing, and studying knowledge about a certain domain and transforming the knowledge into a computerized representation. An important part of the process concerning knowledge acquisition is identifying the sources of where to uncover and identify the rules, policies, procedures and practices applied to the domain. A detailed study of various knowledge acquisition technologies is given in [65].

4.4.1 Automated Knowledge Acquisition: Knowledge Discovery

In the digital world, Enterprises capture and store large volumes of detailed data related to business processes, operation, design, manufacturing, sales, marketing and inventories etc., in real time. However, these collected data are commonly not exploited to their full potential. This potentially valuable knowledge resource is generally not thoroughly understood, reused or exploited. The current limited use of the accumulated data has led to the rich data but poor information problem. Existing databases generally contain large numbers of records and attributes that need to be simultaneously explored because of the possible relationships and interrelationships that may exist. The volume and/or complexity of such data generated make manual analysis impossible and therefore automated knowledge acquisition methodologies are essential for knowledge to be extracted in a form that can benefit the business. Hence, knowledge discovery in databases and data mining will become extremely important tools for manufacturing enterprises which can use them to derive knowledge from data [114]. The next chapter discusses knowledge discovery in different types of databases.

5

Knowledge Discovery in Databases- Tools and Techniques

This chapter discusses knowledge discovery in databases (KDD) and differentiates it from data mining. It describes types of data mining, different functions, and tools and technique used for data mining. It reviews the existing literature in the domain of knowledge discovery applications in manufacturing, discusses text mining and reviews its application areas with a view to identifying the research gaps.

Chapter Outline

5.1 Motivation

Necessity is the Mother of Invention.

In recent years, knowledge discovery in database (KDD) and data mining (DM) have attracted a great deal of attention in manufacturing enterprises. The major reason for this is the wide availability of huge amounts of data and the need for turning these data into useful and novel information and knowledge. In the context of Moderator technology, the efficiency and effectiveness of decision making is largely dependent on the relevancy and accuracy of knowledge associated with it. One of the major problems identified in the earlier Moderator's research is the gathering of required expert knowledge manually by interviewing and inputting knowledge into a database to implement the Moderator system. The review in chapter 3 revealed that the Moderator system requires up to date knowledge and therefore [36] recommended that the Moderators need the capability of ongoing learning. KDD techniques can help in (semi-) automating the time consuming process of knowledge acquisition that is essential in the development of knowledge based system like Moderators. Automation would increase the speed and reduce the cost of development by decreasing the amount of time needed from experts and knowledge engineers. Implementation of KDD tools and techniques also has the potential to uncover knowledge that might otherwise be overlooked by those involved in the knowledge acquisition process. Therefore, as mentioned in section 1.2 the research aims and objectives and section 6.2 as focus of this work, this research proposes that the Moderator system should be integrated with a knowledge discovery capability to provide an ongoing learning mechanism and identify the new knowledge necessary to enhance the awareness within project teams. However, this is not a straightforward process due to the wide variety of tools, techniques and functions that are needed to perform KDD on a variety of structured and unstructured data.

5.2 Knowledge Discovery in Databases and Data Mining

5.2.1 KDD and Data Mining

In the past, deriving knowledge from data has been given a variety of names such as data mining, knowledge extraction, information discovery, information harvesting, data archaeology and data pattern processing. The term *data mining* has mostly been used by data analysts, statisticians and management information system communities. The term

Knowledge Discovery in Databases (KDD) was first coined at the first KDD workshop in 1989 [115] to emphasize the fact that knowledge is the end product of data driven discovery. KDD refers to the overall process of discovering useful knowledge from data and data mining refers to a particular step in this multi step process. Data mining is the application of specific algorithms for extracting patterns from data [116]. The KDD process includes several pre-processing methods aimed at facilitating the application of the chosen data mining algorithm and post processing methods aimed at refining and improving the discovered knowledge.

5.2.1.1 Interdisciplinary Nature of KDD

KDD is an interdisciplinary field, using methods from several research fields including machine learning, statistics, pattern recognition, databases technology, artificial intelligence, knowledge acquisition for expert systems, data visualization and high performance computing. The unifying goal is to extract high level knowledge from low level data from large data sets. The data mining component of KDD currently relies heavily on known techniques from machine learning, statistics and pattern recognition to find patterns from data in the data mining step of the KDD process. Therefore, the over all KDD process can be viewed as a multi-disciplinary activity that encompasses techniques beyond the scope of a single discipline such as machine learning or statistics. However, some authors [117, 118] mentioned that statistics in particular has much in common with KDD. Knowledge discovery from data is fundamentally a statistical endeavour. Statistics provide a language and framework for quantifying the uncertainty that results when one tries to infer general patterns from a particular sample of an overall population. Thus, data mining is a legitimate activity as long as one understands how to do it correctly; data mining carried out without paying attention to the statistical aspect of the problems needs to be avoided. KDD can also be viewed as incorporating a broader view of modelling than statistics. KDD aims to provide tools to (semi)-automate the entire process of data analysis and the statistician's "art" of hypothesis selection. Another related field called data warehousing helps set the stage for KDD by (1) collecting and cleaning the data and (2) providing access to the data. Another popular approach from data warehousing called Online Analytical Process (OLAP) focuses on simplifying and supporting multi dimensional data analysis, but the goal of KDD is to automate the process as much as possible.

5.2.1.2 Basic Definitions of KDD

KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [116]. Here, data are a set of facts F, and a pattern is an expression E in some language L describing a subset F_E of the data or model F. E is called a pattern if it is simpler than the enumeration of all facts in $F_{E.}$ The term process implies that KDD comprises many steps such as data preparation, search for patterns, knowledge evaluation and refinement, all repeated in multiple iterations. A measure of certainty, measuring the novelty of discovered patterns, is a function C mapping expression in L to a partially or totally ordered measure space M_C. An expression in L about a subset $F_E \subset F$ can be assigned a certainty measure c = C (E, F). Another term in the definition novelty of patterns can be measured by a function N (E, F) with respect to change in data or knowledge. Patterns found should potentially lead to some useful actions as measured by some utility function u = U (E, F) mapping expressions in L to a partially or totally measured space M_U. The unified goal of KDD is to make patterns understandable to humans. This is measured by a function s = S (E, F) mapping expression E in L to a partially or totally measured space M_S [119].

5.2.1.3 Desirable Properties of Discovered Knowledge

Freitas [120] mentioned that the knowledge derived from the KDD process must satisfy three general properties namely, accurate, comprehensible and interesting. The discovered knowledge should therefore have a high predictive accuracy rate. In addition, discovered knowledge should be comprehensible for the user. Knowledge comprehensibility can be achieved by using high level knowledge representation. In the context of data mining, a popular representation is a set of IF-THEN rules. In the context of prediction, each rule is of the form:

IF (some_conditions_are_satisfied)

THEN (predict_some_value_for_an_attribute)

The third and important property called *interestingness* is usually taken as an overall measure of pattern value, combining validity, novelty, usefulness, simplicity and understandability, and can be expressed as $i = I\ (E, F, C, N, U, S)$, which maps expressions in L to a measure space M_I. Given these notions, a pattern $E \in L$ can be considered as knowledge if it exceeds some user specified threshold $I \in M_I$, i.e., $I\ (E, F, C, N, U, S) > I$. A user can select some threshold $c \in M_C$, and $s \in M_S$ and $u \in M_U$ and term a pattern E knowledge iff $C(E, F) > c$, and $S(E, F) > s$, and $U(E, F) > u$. As a matter of fact, knowledge in this definition is purely user oriented and domain specific and is determined by whatever function and threshold the user chooses. The role of interestingness is to threshold the huge number of discovered patterns and reports only those which may be of some use [116, 119].

Data mining is a step in the KDD process that consists of applying data analysis and discovery algorithms that under acceptable computational efficiency limitations produce a particular enumeration of patterns (or models) over the data.

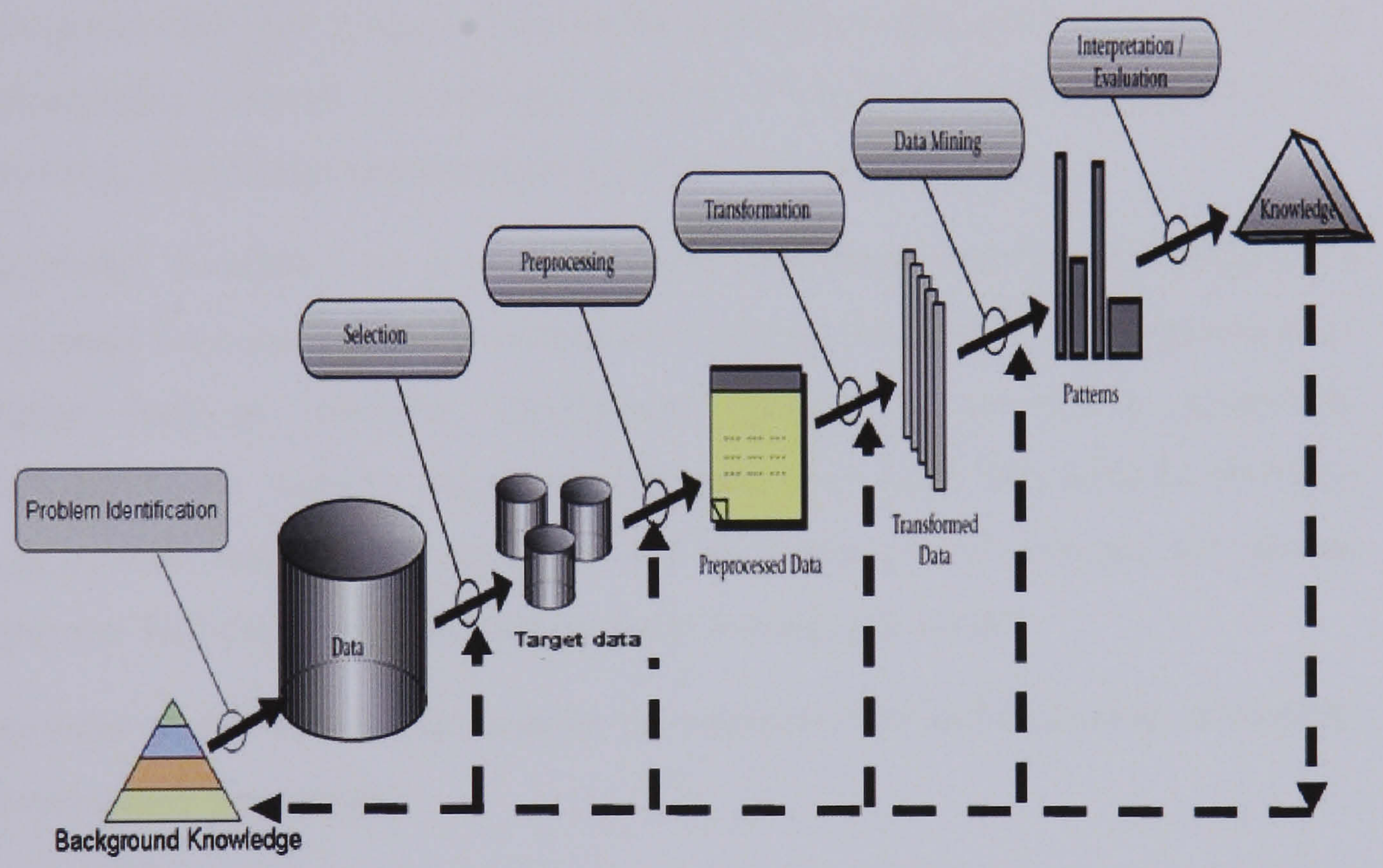


Figure 5-1: The KDD process [116, 119]

5.2.2 The KDD Process

The overall KDD process is outlined in Figure 5-1, which shows that the KDD process is interactive and iterative involving numerous steps and requiring several decisions to be made by the user [116, 119].

1 *Background knowledge:* This includes understanding the application domain and relevant background knowledge related to the application area. It is also important to clearly identify the problem and determine the goal of the KDD process from the customer's view point.

2 *Data:* The collection and selection of appropriate data includes the collection of raw data, selecting subsets of the data, and focussing on the set of variables which were most relevant to the problem.

3 *Pre-processing:*: This includes the pre-processing of data such as noise removal, insertion of missing values and data cleaning. Data from the real world are often erroneous, incomplete and inconsistent, perhaps due to the operational errors or system implementation flaws. Therefore, such low quality data needs to be cleaned. Furthermore, data should be consolidated into forms appropriate for mining.

4 *Transformation*: This includes finding useful features to represent the data depending on the goal of the task. Using several techniques such as dimensionality reduction or transformation methods, the effective number of variables under consideration can be reduced, or invariant representations for the data can be found.

5 *Data Mining* : To achieve the goal as defined in step one the kind of knowledge to be mined must be considered and various data mining functions or a combination of functions such as clustering, classification, prediction, association, regression, summarization etc. need to be performed to derive a model. This includes choosing an appropriate data mining algorithm and its parameters to perform the desired function to find the patterns in data or derive a predictive model.

6 *Interpretation and evaluation*: This includes the interpretation and evaluation of patterns to derive novel knowledge.

7 *Knowledge*: The discovered knowledge may be incorporated into a performance system for further action or it may simply be documented and reported to the decision maker. This process also includes the checking and resolving of any conflicts between previously believed (or extracted) knowledge and new knowledge discovered.

The overall KDD process can involve significant iterations and can contain loops between any two steps. In the past, most KDD works have mainly focussed on step 7 i.e. data mining. However, other steps are as important or even more for the successful implementation of KDD.

5.2.3 Data Mining

Data mining uses automated tools and employs sophisticated algorithms to discover hidden patterns, associations, anomalies and/or structure from large amounts of data stored in data warehouses or other information repositories. A data mining task can be descriptive, i.e. discovering interesting patterns describing the data, or predictive, i.e., predicting the behaviour of the model based on available data. Data mining involves fitting models to, or determining patterns from, observed data. The fitted models play the role of inferred knowledge based on whether the model reflects useful or interesting knowledge and this judgement is part of the overall interactive KDD process, where subjective human judgement is typically required. Most data mining methods are based on tried and tested techniques from machine learning, pattern recognition, and statistics. In general, a data mining algorithm can be viewed as consisting of three primarily algorithmic components as follows:

- *Model representation*: Is the language used to describe discovering patterns. If the representation is too limited then no amount of training time or examples can produce an accurate model for the data. The functions of the model include classification and clustering, and its representational form includes tools like neural network and linear discriminates.

- *Model evaluation criteria*: Are quantitative and represent the goodness of the fitness function of the model to the data. For example, predictive models are often judged by the empirical prediction accuracy on some test data, where as descriptive models can be evaluated along the dimensions of predictive accuracy novelty, utility, and understandability of the fitted model.

- *Search algorithms*: Consists of two components: (a) parameter search and (b) model search. Once the model representation and evaluation criteria are fixed then data mining is purely an optimization task i.e. finding the parameters and models that optimize the evaluation criteria.

A particular data mining algorithm is usually an instantiation of the model/evaluation/search components. The more common model functions in current data mining practices are discussed in the next subsection.

5.2.3.1 Data Mining Functions

In general, the two high level primary goals of data mining are prediction and description. However, the boundaries between these two are not sharp and some of the predictive models can be descriptive and vice versa. The goal of prediction and description can be achieved using a variety of data mining functions as described follows:

- ***Classification***: is a learning function that classifies a data item into one of several predefined categorical classes. It is one of the most studied data mining functions by machine learning and statistics community [120]. The goal is to predict the value (the class) of a user specified goal attribute based on the value of other attributes called the predicting attributes. The classification rules can be considered as a kind of prediction rule where the rule antecedent (IF part) contains a combination – typically a conjunction- of conditions or predicting attribute values and the rule consequent (THEN part) contains a predicted value of the goal attribute. For example in context of credit check in finance the classification rule may be: *IF (Unpaid_Loan?= "No") AND (Current_account_balance > £3,000) THEN (Credit = "good");*

  IF (Unpaid_Loan? = "Yes") THEN (Credit = "Bad");

  In the classification task, the data being mined is divided into two mutually exclusive data sets, the training set and the test set. The data mining algorithms discover rules using only a training dataset. Once the training process is finished, and the algorithm has found a set of classification rules, the predicting performance of these rules are evaluated on the test set, which was not seen during the training. The performance is measured using predictive accuracy. A detailed discussion of a prediction function is presented in [121].

- ***Regression***: is a learning function that maps a data item to a real valued prediction variable. Regression has a wide application for example in manufacturing, estimating the surface roughness of a turning process based on the input parameters such as feed, rate and depth of cut; predicting the consumer demand based on the marketing parameters; predicting the quality of a manufactured product based on the input parameters such as variables used at the different stages of production. Several advancements have been made in regression analysis and an advanced approach called multi parametric extended regression analysis can be seen in Figure 5-2 below. Where the X Rating value determines the quality of product.

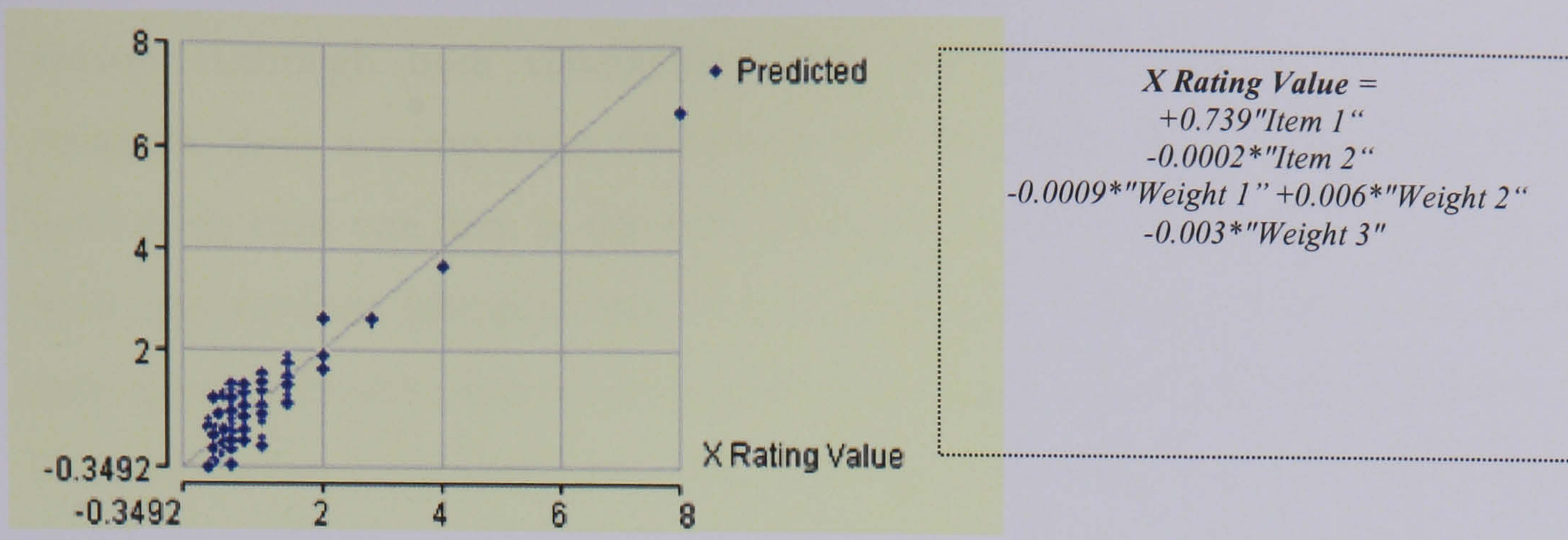


Figure 5-2: An advanced regression analysis on the manufacturing data

- ***Clustering***: Clustering also known as unsupervised learning, is a common descriptive task. Unlike classification (supervised learning), in clustering the class object of each data object is not known. Clustering maps a data item into one of the several clusters, where clusters are natural groupings of data items based on similarity metrics or probability density models. Within the same cluster data objects are similar to one other but they are dissimilar to the objects in other clusters. Liao and Wen [122] reviewed the application of artificial neural networks (ANN) for clustering and classification. A detailed review and study of clustering techniques' application areas are mentioned in [123].

- ***Summarization***: Summarization as a part of descriptive data mining describes the dataset in a concise and summarative manner and presents interesting general properties of data. More sophisticated methods of summarization include the derivation of summary rules, multi-variate visualization technique and the discovery of functional relationships between variables. Summarization techniques are often applied to interactive exploratory data analysis and automated report generation.

- ***Discovering Association rule***: Association rules mining was first introduced in 1993 and is used to identify relationships between a set of items in a database [124]. These relationships are not based on inherent properties of the data themselves (as with functional dependencies), but rather are based on co-occurrence of the data items. Generally, it is used by management for market basket type analysis to discover association rules to increase the effectiveness (and reduce the cost) associated with advertising, marketing, inventory and stock location on the floor. An association rule is a relationship of the form IF X THEN Y, where X and Y are sets of items and $X \cap Y = \varnothing$, A very simple example is; *IF fried_potatoes* THEN *soft_drink*, *ketchup*. Although both classification and association rules have an IF-THEN structure, there are important differences between them. First, association rules can have more than one item in the rule consequent, whereas classification rules always have one attribute (the goal one) in the consequent. Second, unlike the association task, the classification task is asymmetric with respect to the predicting attributes and the goal attribute. Predicting attributes can occur only in the rule antecedent, whereas the goal attribute occurs only in the rule consequent [120]. Moreover, these associations may have pre-specified strength and confidence. A detailed survey of association rule applications has been presented in [125].

- ***Dependency Modelling*:** This task can be regarded as a generalization of the classification task. It consists of finding a model that describes significant dependencies between the variables. A dependency model exists at two levels: (1) the structural level of the model specifies (often in graphical form) which variables are locally dependent on each other and (2) The quantitative level of the model specifies the strengths of the dependencies using some numerical criteria. For example, using the same example as in classification, a prediction (IF-THEN) rule is discovered, since this is a high level knowledge representation. The rule discovered may be: *IF (Salary = "High") AND (Current_account_balance > £3,000) THEN (Credit = "good"); IF (Credit = "good") AND (Age > 21) THEN (Grant_loan = "yes");* Detailed discussion of dependency modelling can be found in [116].
- ***Sequence Analysis*:** This task models sequential patterns, like time series analysis. The goal is to model the state of the process by generating the sequence or to extract and report deviation and trends over time. A more detailed discussion about sequence analysis or change and deviation detection is given in [126].

To perform these data mining functions, one or a combination of tools and techniques are applied. The following subsection discusses the basic algorithms used in data mining.

5.2.3.2 Data Mining Algorithms

Several techniques including statistical approaches, machine learning and other soft computing techniques have been applied to handle the various challenges posed by the data mining. The main constituents of these algorithms include statistical methods, regression analysis, fuzzy sets, neural network, neuro-fuzzy computing, genetic algorithm, rough set theory, association rule mining, decision tree and other hybrid techniques.

Each of them contributes a different methodology for addressing the problems in its domain. This is done in a cooperative, rather than a competitive, manner. The target result is to achieve a more intelligent and robust system providing human interpretable and low cost approximate solutions. It is important to mention that there is no universally best data mining method; choosing a particular algorithm or a combination of algorithms is entirely dependent on the particular application and requires human interaction to decide on the suitability of the approach. The following section describes some of the basic algorithms.

Statistical Approaches

Historically, statistical work has focused mainly on testing of preconceived hypotheses and on fitting models to data. Statistical approaches usually rely on an explicit underlying probability model. In addition, it is generally assumed that these methods will be used by statisticians and hence human intervention is required for the generation of candidate hypotheses and models. Statistical methods of data analysis vary from one dimensional to multivariate data analysis and provide a variety of data mining techniques including different type of regressions. There are a variety of statistical data mining techniques available in the literature. A detailed discussion and description of several statistical data mining approaches is presented in [127]. However, several methods are briefly mentioned below for the sake of completeness.

- *Summary Statistics*: In data mining analysis, it is quite informative to know about the central tendencies and data dispersions of the datasets. These describe the differences between datasets. Typical measures for central tendency includes the metrics, mean, median and mode, while measures of data dispersion include range, variance and standard deviation.
- *Predictive regression*: The prediction of continuous values can be modelled by a statistical technique called regression. The objective of regression analysis is to determine the best model that can relate the output variable to the various input variables. There are several forms of regression, such as linear, multiple, weighted, polynomial, non-parametric and robust (where robust methods are useful when errors fail to satisfy normalcy condition or when the data contain significant outliers).
- *Discriminant Analysis*: this technique is used to predict a categorical response variable. Unlike the generalized linear model, it assumes that the independent variable follows

a multivariate normal distribution. The procedure attempts to determine several discriminant functions (linear combinations of the independent variables) that discriminate among the groups defined by the response variables.

- *Time series*: there are several statistical techniques for analyzing time-series data such as auto regression methods, univariate ARIMA (Autoregressive Integrated Moving Average) modelling, and long memory time series modelling.
- *Survival Analysis*: several well established statistical techniques exist for survival analysis, which originally were designed to predict the probability that a patient undergoing a medical treatment would survive at least to time t. Now a days, it is more applicable to manufacturing where the life span of industrial equipment can be estimated. Several popular methods include Kaplan-Meier estimates of survival, Cox proportional hazards regression model and their extensions.

A detailed discussion of these and several other statistical methods are discussed in [127].

Fuzzy Sets

The modelling of imprecise and qualitative knowledge, as well as the transmission and handling of uncertainty at various stages are possible through the use of fuzzy sets. Fuzzy logic is capable of supporting, to a reasonable extent, human type reasoning, in natural form. Fuzzy set theory constitutes a powerful approach not only to deal with incomplete, noisy or imprecise data, but for developing uncertain models of the data that provide smarter and smoother performance than traditional systems [128]. A detailed discussion of the application of fuzzy sets for knowledge discovery is presented in the special issues [129, 130]. In literature, fuzzy sets have received strong attention for performing several different functions such as clustering, association rule, functional dependency, data summarization and web applications.

Inductive Learning technique

Inductive learning techniques can be divided into two main categories, namely decision tree induction and rule induction as described below:

- *Decision Tree Induction:* Several algorithms such as classification and regression tree (CART), ID3, C4.5 and C5.0 have been developed to build decision trees. These learning systems are categorized as "divide and conquer" inductive systems. The knowledge induced by these systems is represented as a decision tree, which consists of internal nodes and leaf nodes. Each internal node represents a test on an attribute and each outgoing branch corresponds to a possible result of this test. Each leaf node represents a classification to be assigned to an example. To classify a new example, a path from the root of the decision tree to a leaf node is identified based on values of the attributes of the example. The class at the leaf node represents the predicted class for that example.

CART is a binary decision tree algorithm which uses "Gini index" as an evaluation function for splitting the nodes. ID3 is a well known decision tree system which utilizes the "information gain" criteria for splitting the nodes. The information gain is computed from the entropy measures that characterize the impurity in a collection of training instances. Detailed discussion about CART and ID3 algorithms are presented in [131]. C4.5, a variant and extension of ID3 is another popular decision tree. It uses "gain-ratio" criteria because the information gain criterion has a strong bias in favour of attribute tests with many values. To relax the over fitting problem, C4.5 also uses a tree pruning method that tries to simplify the tree by eliminating sub-trees that seems too specific. The C4.5 can also transform the generated decision tree to a set of IF-THEN rules. A detailed discussion of C4.5 induction learning system is presented in [132].

- *Rule Induction:* Several rule induction algorithms such as AQ15, CN2, RIPPER, SLIPPER and RULES have been used as inductive learning technique. These learning systems are categorized as "separate and conquer " inductive system. In contrast to decision tree learning, rule induction directly generates either an unordered set of IF-THEN rules or ordered set of IF-THEN rules also called as decision lists. The general operation of rule induction algorithms is the same. They induce the rule i.e. "set one rule at a time". After a rule is generated, the instances covered by it are removed from the training data set and the same induction procedure is applied to the remaining dataset until all the instances are covered by one rule in the rule set. The detailed discussion about all the techniques of Rule Induction as mentioned above is presented in [133].

Neural Networks (NNs)

In the past, neural network were thought to be unsuitable for data mining because of their inherent black box nature. No information was available from them in the symbolic form, suitable for verification or interpretation by humans. Recently, there has been great emphasis on redressing this situation, by extracting the embedded knowledge in trained networks in the form of symbolic rules. This serves to identify the attributes that, either individually or in combination with, are the most significant determinants of the decision or classification. In general, it provides a practical method for learning real valued and discrete valued target concepts in a way that is robust to noise in the training data [134]. It has been recognized that neural network models are well suited to problems in which the training data corresponds to noisy and complex sensor data. NNs can be classified according to their mode of learning namely supervised, unsupervised and reinforcement leaning. Reinforcement learning is a special type of supervised learning where a critic is employed to evaluate the goodness of the NN output corresponding to a given input. For data mining, perhaps the back propagation network and the Kohonen neural network [46] are the most popular algorithms. As shown in Table 5-1, there are many different types of neural network models and their functions and each of them has their own advantages and disadvantages as mentioned in [134]. The selection of a model and architecture for the neural network depends on the data type and quantity of data, training and functional requirement. In the past, for data mining purposes NN was mainly used for rule extraction, rule evaluation, clustering and regression [119]. One of the main advantages of NN is their wide applicability; however they also have two drawbacks (1) complexity in understanding the produced model (2) and time consuming training process. Recently, several initiatives have been made in developing new NN methods to address these issues [122].

Table 5-1: Summary of different neural network models [135]

Model (network)	**Training Paradigm**	**Topology**	**Primary function**
Adaptive resonance theory	Unsupervised	Recurrent	Clustering
ARTMAP	Supervised	Recurrent	Classification
Back propagation NN	Supervised	Feed forward	Classification, modeling, time series
Radial-basis function NN	Supervised	Feed forward	Classification, modeling, time series
Probabilistic NN	Supervised	Feed forward	Classification
Kohonen feature map	Unsupervised	Feed forward	Clustering
Learning vector quantization	Supervised	Feed forward	Classification
Recurrent back propagation NN	Supervised	Limited recurrent	Modeling, Time series
Temporal difference learning	Reinforcement	Feed forward	Time series

Genetic Algorithms (GA)

Genetic algorithms are adaptive, robust, efficient and stochastic random search optimization techniques, which have been inspired by the process of biological evolution. They are mainly applicable to situations where the search space is large. Several GA based systems have been developed for learning classification rules and association rules. In the context of GA, rules are represented in the form of bit strings whose particular interpretation depends on the application. The search for an appropriate rule starts with a collection of randomly generated solutions called a population and each rule is called a chromosome. A collection of operations such as crossover, mutation, selection and reproduction are applied to the current population to generate an improved population. GAs use an iterative process and at each iteration the rules in the current population are evaluated relative to a given measure called a fitness function where the fittest rule is selected probabilistically as seeds for producing the next generation's population. The process performs a generate-and-test beam search of the rules, in which variants of the best current rules are most likely to be considered next. GAs have been applied to a variety of learning tasks in manufacturing systems. However, the literature in the domain of GA based data mining is not as rich as that of other algorithms. Recently, hybrids of GAs have been frequently used in combination of other algorithms such as neural networks. A detailed discussion about GA is presented in [136, 137]and [138].

Rough Set Theory(RST)

Rough Set Theory has emerged as a mathematical tool to manage vague and uncertain data for problems related to information systems, indiscernibility relations and classification, attribute dependence and approximation accuracy, reduct and core attribute sets, and decision rules. The fundamental principle of a rough set based learning system is to discover redundancies and dependencies between given features of a problem to be classified. It is an effective tool for extracting knowledge from a decision table in the form of IF-THEN rules. The RST first creates a knowledge base, classifying objects and attributes within the created decision tables. Then a knowledge discovery process is initiated to remove some undesirable attributes. Finally, the data dependency is analysed in the reduced database to find the minimal subset of attributes called a reduct. A detailed discussion of RST has been discussed in [139]. In the context of manufacturing, it has a wider applicability and Kusiak [140] used RST for prediction purposes in semi-conductor manufacturing.

Association Rule Mining

Association rule mining (ARM) is one of the most important and well researched techniques of data mining to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in transactional databases, or other data repositories. ARM is generally used to identify association rules that satisfy the predefined minimum support and confidence from a given database. It is generally a two stage process. The first stage includes finding those itemsets whose occurrence exceeds a predefined threshold in the database; those itemsets are called frequent or large itemsets. The second stage generates association rule from those frequent itemsets with the constraint of minimal confidence. For example, one of the frequent itemset is $F_K = \{I_1, I_2, I_3, \ldots, I_K\}$, the first association rule with this item set is: $\{I_1, I_2, I_3, \ldots, I_{K-1}\} \Rightarrow \{I_K\}$, by checking the confidence this rule can be determined as interesting or not. In this manner next rule is generated by deleting the last item in the antecedent and inserting it to the consequent, further the confidences of the new rules are checked to determine the interestingness of them. This process is iterated until the antecedent becomes empty. If desired, additional interestingness measures can also be applied to identify the novel rule. The apriori algorithm provides one early solution to association rule mining and recently several advances have been built upon it. A detailed discussion about association rule mining, apriori algorithms and state-of-art developments have been mentioned in [141, 142].

In this manner, it can be seen that the first generation of data mining algorithms has been demonstrated to be of significant value across a variety of real world applications. But, these seem to work best for the large sets of numerical or symbolic structured data collected in a single database with a particular decision making task in mind.

5.2.3.3 Challenges of the Data Mining Algorithms

The recent development of new generation algorithms is expected to tackle more diverse sources and types of data in order to support mix initiative data mining, where the human experts collaborate with the computers to form a hypothesis and test them. [116, 119, 143] identified the main challenges of data mining as follows:

1 *Size of Dataset:* Very large databases create combinatorial explosive search spaces and need robust and efficient algorithms. In contrast some algorithms cannot be applied to small datasets.

2 *Overfitting and Statistical Significance:* There is a possibility that due to large dataset the presence of spurious data points may lead to overfitting of the model. There is a need for regularization and resampling methodologies for designing the model.

3 *Non standard, Missing and Noisy Data*: the data can be missing and/or noisy.

4 *Understandability of Patterns*: It is necessary to make the discovered knowledge understandable by humans.

5 *User interaction and prior knowledge*: Data mining methods are iterative and interactive in nature and therefore there is a need to incorporate the relevant information including the domain knowledge into the model.

6 *Integration with other system*: Most data mining methods are standalone and need integration with the databases and other knowledge based systems to realize the true benefit of KDD process in decision making.

7 *Mixed data*: learning from data that is represented by a combination of various forms of data such as numeric, symbolic, images and text is a challenging task.

8 *Management of changing data and knowledge*: discovered knowledge and patterns need to be updated as the data in the databases are continuously modified, deleted and augmented.

5.3 Data Mining in Manufacturing: A Review based on the kind of knowledge

Data mining has emerged as an important tool for semi-automatic knowledge acquisition from manufacturing databases [143]. A very thorough review of data mining applications in manufacturing with a special emphasis on the kind of knowledge to be mined has been carried out by the author. Han [144] classified the kind of knowledge to be mined as the data mining functions to be performed on data as mentioned in section 5.2.3.1. Furthermore each knowledge type is identified in several key knowledge areas in manufacturing such as yield improvement, manufacturing processes, design, defect analysis, fault diagnosis, maintenance, supply chain, customer's relations management, planning and scheduling, manufacturing systems, decision support systems, quality control product development and process improvement. This paper was published in the Journal of Intelligent Manufacturing as Choudhary *et al.*[145] and its abstract is included as appendix 2. It shows that there is a rapid growth in the application of data mining in the context of manufacturing enterprises. From a detailed review of 150 papers, 50% of the papers were published in the last three years. This review revealed the progressive application areas and identified existing research gaps in the context of manufacturing enterprises. A novel text mining (TM) based approach was adopted to identify the research gaps and good practices using the abstracts and keywords of 150 identified papers. A detailed discussion of the approach used is given mentioned in [145]. Therefore, only a brief summary of this published paper is presented in the following sub sections.

5.3.1 Analysis of literatures and discussions

Despite the rapid growth in the application of data mining, there is still a slow adoption of this technology in some manufacturing industries primarily due to following reasons:

- In a particular knowledge area, it is very difficult to determine which data mining function needs to be performed to derive novel knowledge.
- In many cases a number of data mining techniques are possible, however, which technique should be used, or which one is most appropriate is not clear.

To address these challenges, a novel TM based approach has been applied on the reviewed literature's keywords and abstracts to identify a pattern in the application areas. A detailed discussion about text mining is presented in section 5.4. The major objectives of this application and results are as follows:

1 *Automating the process of identifying the research gaps*: This research is the first of its kind where identification of research gaps has been automated. Text mining techniques such as text analysis and link analysis have been applied on the abstract and keyword based data of 150 reviewed literatures of data mining application in manufacturing. Linkages have been established between the keywords of key knowledge areas, techniques used and functions performed. Detailed analysis as presented in [145] shows that very little or negligible work has been carried out in the areas such as data mining applications in collaborative projects, virtual enterprises and supply chain management (only 5 out of 150 papers). Data mining based intelligent decision support can be used to automate the process of decision making from data generated in areas such as RFID applications, supply chain planning and optimizations, ERP systems and virtual enterprises.

2 *Identification of overlooked and under examined areas*: it has been found that major data mining applications have been carried out in the semi-conductor industries in the

areas such as quality control, manufacturing process, fault diagnosis, maintenance, job shop type problems, and yield improvement. One of the reasons for this may be that large amounts of data are generated during these processes and this is easily available to use and does not need any dedicated equipment for data gathering. In addition, small improvements in the semiconductor industry can have a significant impact on cost. Fewer applications have been found in supply chain, virtual enterprise and collaborative projects and this may be due to the integration issues of the data mining system with the existing system. In addition, less work has been carried out in association rule mining function in comparison to other data mining functions.

3 *Identifying common practices and good practices*: The TM experiments also showed that the techniques like NN, regression analysis, rough set theory and decision tree are mostly used for prediction purposes in manufacturing. For clustering purposes, hybrid algorithms and fuzzy c means clustering have mostly been used. Most of the hybrid algorithms include neural network as a primary tool. Similarly for classification purposes hybrids of neural network or rough set theory have been used. Some of the techniques have been more frequently used in a particular type of manufacturing area. For example, NNs have been most commonly applied to analyse manufacturing process related data. Similarly, association rule mining has been applied more frequently to analyse product design related data. Some of the functions are more frequently used in chosen areas such as prediction in manufacturing processes, maintenance and job shop, association in product design, concept description for job shop type problems, classification in quality control etc. A detailed discussion is presented in appendix 2.

4 *Increased use of hybrid data mining methods*: It has been found that there is no universally best data mining method for all manufacturing contexts. Therefore, there is an increased used of combinations of traditional data mining algorithms to capture advantages from each technique used to perform the desired function of data mining at various stages of KDD process. Another noticeable fact is the use of fuzzy sets with various data mining techniques. One of the reasons for this may be, that fuzzy sets are inherently suitable for coping with linguistic domain knowledge and producing more interpretable solutions. In addition, when fuzzy sets are used in association with a different data mining technology, they are capable of handling the issues related to understandability of patterns, incomplete and noisy data, and

human interaction, and can provide approximate solutions faster in comparison to stand alone algorithms. It has been found that hybridization of NNs and fuzzy logic is increasingly used in manufacturing. In addition, researchers and practitioners have also modelled, or advanced the existing algorithms according to the problem environment. Thus, it can be said that hybridization of data mining algorithms may lead to a better solution quality when compared to existing and traditional algorithms.

In this manner, the research revealed in [145] and this thesis is primarily concentrated on the application and therefore quality of data, data preparation issues has not be taken much into consideration. However, a more generic process of data cleaning is essential to enable the growth of data mining in manufacturing. Greater focus also needs to be put on the knowledge representation and quality of rules discovered through data mining of structured data. Text mining is starting to be used by the data mining practitioners and researchers to handle the unstructured data. However, there are very limited applications of text mining in real world applications and it is yet to be adopted by manufacturing industries. Therefore, KDT and Text mining are discussed in the next section.

5.4 Knowledge Discovery in Text(KDT) and Text Mining(TM)

KDT and TM are multi-disciplinary fields of research that try to resolve the problem of information overload by identifying and retrieving useful knowledge from large stores of text-based data. KDT and TM involve techniques like information retrieval, text analysis, information extraction, clustering, categorization, visualization, machine learning, natural language processing, data mining and knowledge management [146]. KDT refers to the overall process of turning unstructured or semi structured textual data into high level information and knowledge. Following the definition of KDD by [116], [146] *defined KDT as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in unstructured data".* On the other hand, *Text Mining (TM) is a step in the KDT process consisting of particular data mining and natural language processing algorithms that under certain computational efficiency and limitations produces a particular enumeration of patterns over a set of unstructured textual data.* Text mining uses unstructured or semi structured textual data or information and examines it in an attempt to discover structure and implicit meanings "hidden" within the text.

As shown in Figure 5-3, KDT consists of three main steps.

1 *Document Collection*: The very first step in the KDT process is to identify and collect the documents which need to be retrieved for example, customer emails, technical reports, biomedical documents, patents, project reviews etc.

2 *Retrieving and pre-processing documents*: When the documents have been retrieved, they will normally need to go through a pre-processing stage to transform them into a form suitable for the TM techniques which are going to be applied. The precise nature of the pre-processing will vary depending on the characteristics of the documents and the types of TM to be used. For example, unwanted text may also be removed to reduce the size of text. Transformations may also be done to represent the documents in another form such as XML, Standard Generalized Markup Language (SGML), etc. The resulting documents may then be processed to provide basic linguistic information about the content of each document.

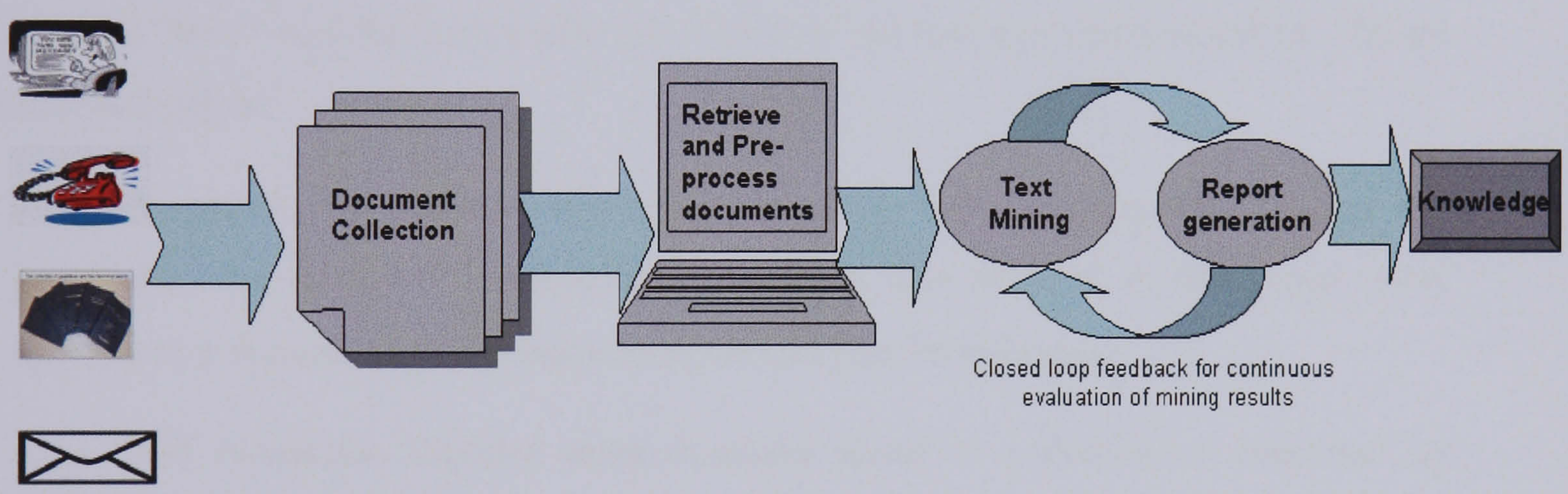


Figure 5-3: Knowledge discovery in text and text mining process

3 *Text Mining*: As a part of the KDT process, TM uses various algorithms and tools to extract metadata or high-level information and / or to discover patterns and relationship within the extracted information. Results are generated in the form of reports, which, if required can be further processed to extract the high-level knowledge. Therefore, generation of reports and text mining work as closed loop feedback for continuous improvement of mining results.

The results of TM are "knowledge" (in different forms) which can be used by decision makers to take further decisions and make improvements.

5.4.1 Text Mining Operations

TM combines techniques from several areas, including natural language processing, data mining, machine learning, information retrieval and information extraction to automatically discover patterns, extract information and generate meta data from large quantities of text. [147] described information extraction (IE) as a technology based on analyzing natural language in order to extract snippets of information. This process takes texts as input and produces fixed format, unambiguous information as output. Information extraction is different from Information Retrieval (IR) in the following sense:

- An IR system finds relevant texts and presents them to the user.
- An IE application analyzes text and presents only the specific information that the user is interested in.

Text Mining covers many different types of techniques and it is therefore difficult to identify the "best" tool for a particular job. Some of the functions performed by TM are summarised below:

- *Feature extraction*: Identification and extraction of key features from the text that can be used as the data and dimensions for analysis. The number of times each term appears in a document (word frequency) should also be indicated.
- *Text Based Navigation*: Enables users to move about in a document collection by relating topics and significant terms. It helps to identify key concepts and some of their relationships.
- *Search and retrieval*: Searching and retrieving for particular text.
- *Text Categorization*: Identification of the main themes of a document. Categorization counts the appearances of words and from the counts, identifies the main topics covered in the document. Categorization often relies on a thesaurus in which topics are predefined and relationships are identified by looking for broad terms, narrower terms, synonyms, and related terms.
- *Clustering*: Automatically groups documents on the basis of some similarity measure, without predefining the category. Documents can also appear in subtopics, ensuring that useful documents are not omitted from the search results.

- *Summarization*: Reduces the length and detail of a document while retaining its main points and overall meaning. Sentence extraction is a widely used strategy to extract important sentences from a given text by statistically weighting all the sentences in the text. Microsoft Word's Auto-Summarize function is a simple example of text summarization.

- *Trends Analysis*: Identifies trends in documents collected over a period of time.

- *Association*: Identifies relationships between various attributes (features that have been extracted from the documents) such as whether the presence of one pattern implies the presence of another pattern in a given set of documents.

- *Information Visualization*: Visual text mining or information visualization puts large textual sources in a visual hierarchy or map and provides browsing capabilities, in addition to simple searching.

- *Text OLAP (Dimensional Matrices)*: Text OLAP (online analytical process) provides an opportunity to perform quick navigation through the textual data. In order to perform that initially a dimensional matrix is created. The dimensional matrix is a collection of user defined dimensions. Each dimension contains either values from a categorical, numerical, or consists of user words or phrases from a text column. While working with the OLAP report, the user defines the value of one of more dimensions and then browses the subset of reports belonging to that slice.

A detailed discussion of the abovementioned techniques and KDT process is presented in the Handbook of Text mining by [146].

Table 5-2: Text mining products, vendors and their websites

Vendor	*Website*	*Product name*
Inxight	www.inxight.com	Smart discovery, Vizserver
SPSS	www.spss.com	Text smart1.0, Clementine
Autonomy	www.autonomy.com	Autonomy incorporated
SAS	www.sas.com	SAS Text miner
Clear forest	www.clearforest.com	ClearForest text-analysis suite
Megaputer	www.megaputer.com	Polyanalyst5.0, Text Analyst2.3
IBM	www.ibm.com	Intelligent miner for text, TAKMI
Convera	www.comvera.com	Retrieval ware

5.4.2 Text Mining Software, Features and Analysis

There are several commercial products available for text mining, as shown in Table 5-2 which lists various vendors and their commercial products used for the text mining purposes. The functionality of the Mining products marketed by the abovementioned vendors varies substantially, particularly as some of the software systems listed focus exclusively on text mining tools, whilst for other, larger vendors (e.g. SPSS, SAS), the TM tools represent only a small portion of the software developed and marketed.

General Architecture for text Engineering (GATE) [147] is free and open software for both commercial and research purposes. It has been successfully used for Information Extraction purposes. IE differs from traditional techniques in that it does not recover from a collection a subset of documents which are hopefully relevant to a query, based on key-word searching (perhaps augmented by a thesaurus). Instead, the goal is to extract from the documents (which may be in a variety of languages) salient facts about prespecified types of events, entities or relationships. These facts are then usually entered automatically into a database, which may then be used to analyse the data for trends, to give a natural language summary, or simply to serve for on-line access. But, its application is limited to information extraction only.

In Table 5-3, a comparative study has been made of all the tools available for text mining, based on the text mining operations mentioned in Section 5.4.1. It can be seen from Table 5-3 that none of the software examined provided the full range of necessary functionalities to achieve all the text mining operations.

Table 5-3: Comparative study of text mining software in terms of performing the following text mining operations

Product		*FE*	*TBN*	*SR*	*TC*	*Clus.*	*Sum.*	*TrA*	*Asc*	*In.Vs*
1	Smart discovery	X			X		X			X
2	Text miner 1.0	X			X	X				X
3	Autonomy	X	X		X	X	X			
4	SAS Text miner	X			X	X			X	
5	Clear forest	X	X	X	X				X	X
6	Polyanalyst 5.0/ Text Analytics2.3	X	X	X	X	X	X			
7	Intelligent Miner	X			X	X	X			
8	Retrieval ware	X			X	X	X			

FE: Feature extraction, TBN: Text Based Navigation, SR: Search and Retrieval, TC: Text Categorization, Clus.: Clustering, Sum: Summarization, TrA: Trend Analysis, Asc: Association, InV; Information visualization.

For experimentation purpose PolyAnalyst 5.0 [148] has been used in this research as it provided a good range of functionalities. The application of text mining in various domains is discussed below to identify the potential research scope.

5.5 Review of Text Mining Application areas

To date, very little research has been published in the area of text mining applications but the main application areas identified include analyzing the following types of databases:

1 *Biomedical documents*: [149, 150] carried out a detailed survey of text mining applications in biomedical research community. The goal of biomedical text mining is to allow researchers to more efficiently identify needed information, uncover relationships obscured by the sheer volume of available information. In this manner, the burden of information overload from the researcher is shifted to the computer by applying text mining to the vast amount of biomedical knowledge that exists in the literature as well as the free text fields of biomedical databases.

2 *Patents*: Patent documents contain important research results and consist of complex technical terminology, which makes it difficult for manual analysis. Tseng *et al.* [151] described a series of text mining techniques used by patent analysts to automate the process of decision making. These techniques include text segmentation, summary extraction, feature selection, term association, cluster generation, topic identification and information mapping.

3 *Financial reports*: Kloptchenko *et al.* [152] combined data and text mining methods for analysing quantitative and qualitative data from quarterly financial reports from three industries in telecommunication sector to see if the textual part of the report contains some indications about future financial performance. Gerdes [153] described a EDGAR-Analyzer, a text mining tool to analyze the textual portion of EDGAR database by the Securities and Exchange Commission (SEC). EDGAR consists of financial and operational disclosures of thousands of public companies. This tool allows the users to specify subject areas of interest and retrieve the relevant information to help making investment decisions.

4 *Customer relations management*: Dorre [154] demonstrated the application of IBM's Intelligent Miner for text mining purposes to provide the necessary tools to unlock the business information that is "trapped" in emails, insurance claims, news feeds, or

other document repositories. It has been successfully applied in analyzing patent portfolios, customer complaint letters and even competitors' Web pages.

5 *Product development processes*: Menon *et al.* [155] showed that text mining can be used in analyzing the service centre database, call centre databases, customer survey database. They also investigated the needs and benefits of text mining for product development processes.

6 *Medical records*: Loh *et al.* [156] used text mining to discover the useful knowledge from textual medical records representing patients' symptoms, signals and social/behaviour characteristics. An automatic system was constructed with this whose goal is to help physicians in disease diagnoses.

The vast majority of this research has been done within the last 5 years and there is negligible work carried out in the manufacturing or construction industries. Some of the difficulties facing the application areas as follows:

- *Information description*: Users in different circumstances or with different needs, knowledge or linguistic habits will describe the same information using different terms. The degree of variability in descriptive term usage is generally very high. In addition, some times one term can have more than one distinct meaning. These add to the difficulties of understanding the texts let alone analyzing them.

- *Noisy database*: Textual databases may contain spelling and grammatical errors, which makes the analysis more difficult.

- *Presence of uninformative words*: In the textual databases there are several words which do not add any value to the text mining process. These uninformative words and redundant words generally tend to adversely affect the text processing results and hence it would be important to identify and remove them during pre-processing.

- Another major difficulty comes in directly extracting the knowledge in the form of IF-THEN rules.

The discussion in this chapter shows that data and text mining has played a vital role in extracting useful and novel knowledge from a variety of industries. The last three chapters have reviewed the state-of-art in methodologies adopted for the proposed interdisciplinary research from areas including moderator technology, collaborative working teams, knowledge types and its management and finally methodologies for automated knowledge acquisition particularly KDD and KDT. The next section will identify the research gap and focus of the proposed research based on the last three chapters.

6

Identification of Research Gap

This chapter identifies and consolidates the research gaps that exist in the literature survey conducted in the last three chapters. It summarizes the focus and objectives of current research to address the identified research gaps.

Chapter Outline

6.1 Research Gaps Identified in Literature

The literature review carried out in several domains as mentioned in the last three chapters identifies the following research gaps and requirements from the state-of-the-art literature.

1. The extensive review carried out by the author into the context of knowledge management systems for collaborative projects in chapter 3 and the literature review of 130 papers in the domain of collaborative and distributed product development reveals the need for efficient learning and sharing of knowledge from multiple domain collaborative projects. Very little work has been done in the area of knowledge discovery in e- collaborative projects. None of the literature reviewed by the author and Li and Qiu [35] have used knowledge discovery and data mining as tools for ongoing learning support of the collaboration by discovering novel knowledge in collaborative projects.

2. The literature survey for all the three versions of Moderators showed that their functionalities are dependent on knowledge based approach for the extraction of useful information by interviewing experts and best practices to acquire knowledge. [36] and a paper published by the author [62] indicate the potential for Moderators to learn and update the knowledge in it's expert modules by using knowledge discovered in the operational databases of project teams.

3. The literature survey and methodologies for knowledge management and acquisition as discussed in chapter 4 indicate three major issues as follow:

    - Although capturing tacit knowledge is very difficult, it can partially be captured in the form of implicit knowledge for example post project reports. However, abundance of implicit knowledge transforms itself into data. Therefore, there is a need to further discover useful knowledge from this so called "abundance of implicit knowledge". For example, post project reports are kind of implicit knowledge as a result of brainstorming, discussions and feedbacks from experts. However if there are many reports, they transform themselves into data.

    - Another issue identified is the continuously changing data, information and knowledge with time.

- [63, 77] identified one of the major challenges for knowledge based systems as: How to produce knowledge within an organization?
- Several modelling techniques have been used for developing knowledge based systems. Also a limited amount of work has been carried out for modelling the knowledge acquisition process using Unified Modelling Language (UML) in various domains. However, according to the best knowledge of the author, none of the work in the identified literature discusses the design of an automated knowledge acquisition system using UML.

4. Several authors including [116, 119, 143, 145] have indicated the need for the integration of a data mining system with a knowledge based system to realize the true benefits of KDD. In addition, the literature review as published by the author in [145] revealed the scope for the application of data mining in collaborative projects, supply chain and virtual enterprises.
5. The review of Text Mining application areas discussed in chapter 5 and the paper published by the author show the scope, needs and benefits of text mining in the PPRs of construction project. Appendix 3, 4, and 5 presents the abstracts of these published papers. Negligible work has been carried out in this area whereas in the context of manufacturing very limited work has been done in areas such as customer's relations management and product development processes.

6.2 Focus of Research Work

Based on the identified research gaps discussed in section 6.1 and to meet the aims and objectives listed in chapter 1, the main focus of work presented in the remaining chapters of this thesis is as follows:-

1. The proposal of the KOATING (knowledge discovery and data mining integrated framework) which enables semi-automatic knowledge acquisition and update of knowledge and learning by discovering knowledge from operational databases of industries. (Chapter 7, sections 7.1 to 7.5)
2. The integration of the KOATING framework with a state of the art Moderator, called the Universal Knowledge Moderator (UKM) to enhance collaboration in e-manufacturing supply chains and illustration of this through an example (Chapter 7, section 7.6)

3. Modelling the design of the KOATING framework using the Unified Modelling Language (UML). (Chapter 8)

4. A demonstration of the application of the KOATING framework showing its use in extracting knowledge from Post Project Review (PPR) reports from construction project supply chains and showing how this can be used to update the expert modules of a Construction Project Moderator (CPM). (Chapter 9).

5. A demonstration of the application of the KOATING framework showing its use in raising awareness of business opportunities and identifying collaborative SME partners for a virtual organisation (VO). (Chapter 10)

The next chapter discusses objectives 1 and 2 as mentioned in section 1.2.

7

Proposed Knowledge Discovery and Data Mining Integrated (KOATING) Framework for Moderators

This chapter proposes a KOATING framework for Moderators. It describes the knowledge discovery module and discusses knowledge miners, knowledge managers and repositories. The architecture for the Universal Knowledge Moderator (UKM) is presented. An illustrative example has been presented to show the functioning of the proposed system.

Chapter Outline

7.1 Knowledge Discovery and Data mining Integrated (KOATING) Moderators

All previous Moderators have used a knowledge based approach to capture the relevant knowledge and information by interviewing experts and storing the knowledge in an object oriented database (OODB) based on a Knowledge Representation Model (KRM) [43]. This is a very time consuming process therefore there is a need to semi-automate the process of capturing the knowledge by embedding a learning capability that can be derived from the operational databases of participating teams. Semi-automating the process of knowledge acquisition would increase the speed and hence reduce the cost of development by decreasing the amount of time needed to acquire knowledge from experts and knowledge engineers. In addition, every piece of knowledge has a lifespan for its validity and therefore it is necessary to continuously review and update the knowledge contained in any knowledge based system. However, integrating the semi-automated knowledge acquisition system with an existing knowledge based system is not simple and is yet another challenge.

These challenges and requirements provide the context of the present research as they indicate that better knowledge management and knowledge discovery integrated systems are required to support Moderators. Knowledge discovery and data mining tools and techniques have been explored in this research to provide a semi-automated methodology to analyze the real time and historical operational data and to determine if useful information and knowledge can be discovered at various steps of collaborative projects and can then be used to update the expert modules. Data mining also has the advantage that it is "exploratory" in nature and can be carried out on existing rather than especially collected data.

To contribute towards the satisfaction of the stated objectives of this research in section 1.2, this chapter proposes a KOATING framework, which incorporates tools and techniques of data mining to provide useful information and knowledge to the Moderators. Figure 7-1 shows the existing structure of a typical Moderator, where the knowledge in any expert module can be manually updated by the knowledge acquisition module. Therefore, there is a need for the KOATING framework to enable the KAM to semi-automatically update the knowledge content of expert modules.

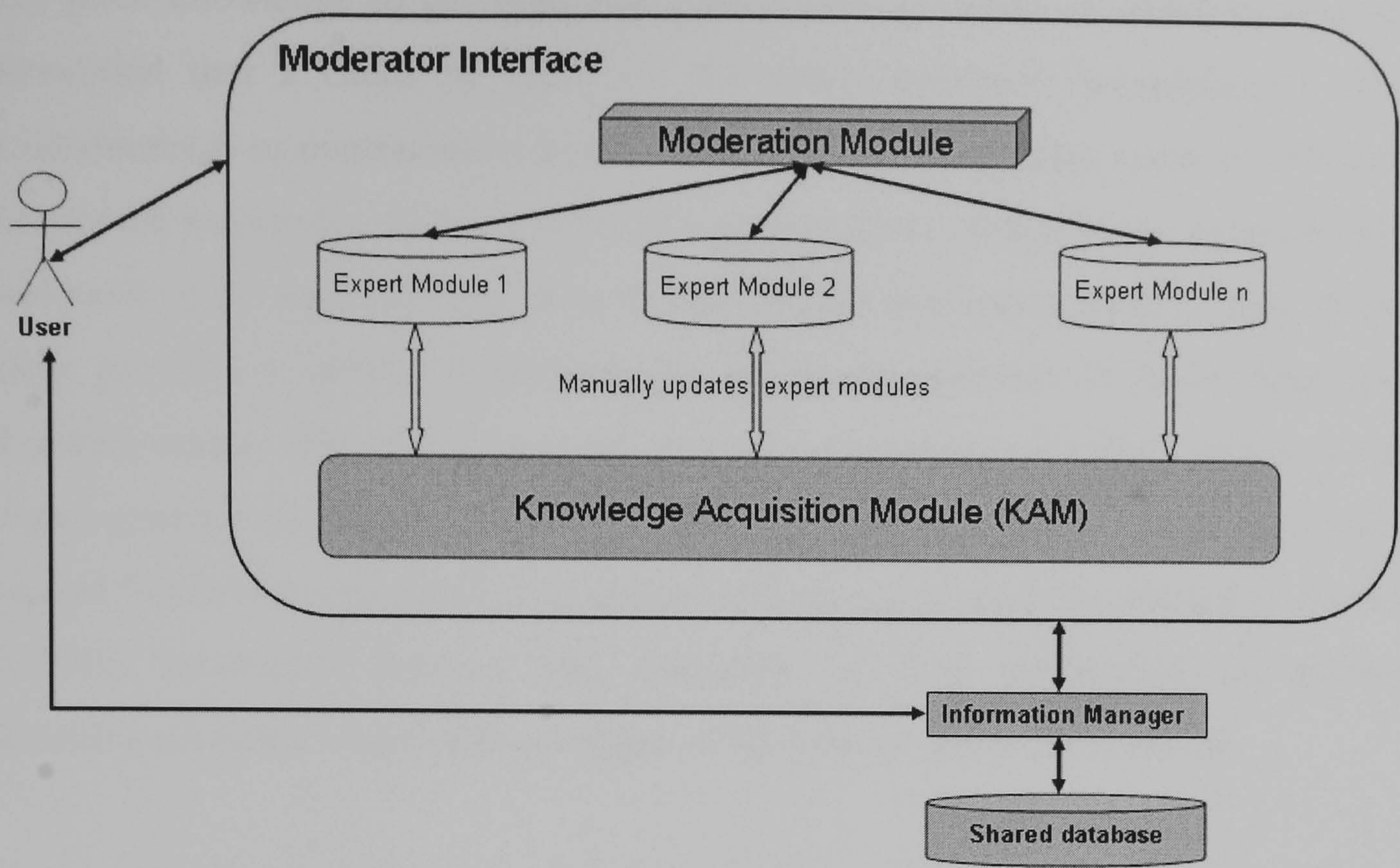


Figure 7-1: Existing structure of Moderator

Enterprises continuously generate large amounts of data during their normal operation and this data can be a valuable asset and potentially important source of knowledge. Identification and retrieval of these knowledge assets may be achieved by applying intensive and intelligent data analysis approaches to these databases with the objectives of identifying patterns, discovering rules and predicting results. Hence, the proposed KOATING Moderator services should be able to apply data mining technology to extract and exploit valuable knowledge from existing and historical operational databases.

Therefore, this research attempts to fill the research gaps identified in chapter 6 by proposing the KOATING framework for Moderators. In this framework, a knowledge discovery module (KDM) has been integrated with the knowledge acquisition module (KAM) from previous moderator systems to extract useful knowledge from large datasets and store them in the expert module for further reuse. The combined functionality of knowledge miners, knowledge manager and repository provides continuous learning, thus enabling semi-automatic update of its expert modules from time to time. This learning capability therefore overcomes some of the imperfections in previously reported knowledge based research as mentioned in the earlier chapters.

Here, it is important to remember that the KDD Guru Fayyad [116] has mentioned that the blind application of data mining to generate knowledge can be dangerous, leading to the use of meaningless patterns. Therefore, it is desirable to incorporate expertise and existing prior knowledge to get valid and interpretable patterns. In addition, it is also recommended that a check be made on the newly generated knowledge to avoid misunderstandings or contradictory knowledge within the knowledge assets by validating the discovered knowledge. Hence a form of semi-automatic update of the expert module is considered to be more realistic than a fully fledged automatic system. This chapter therefore provides a generic framework for use in the context of multi disciplinary collaborative teams. Due to the time and resource constraints of a PhD project a fully functional system and specification for hardware and software has not been provided. A conceptual framework and detailed description of the set of modules needed to develop KOATING moderator services and examples of their application in different collaborative scenarios where different types of knowledge assets are available.

7.2 Proposed KOATING Framework

The basic knowledge acquisition module (KAM) of previous Moderators must be further refined to provide the Moderators with all the knowledge that they require to create, update, learn and remove knowledge during the course of their operations. The KOATING framework does not replace the existing KAM but supports it and enables automatic/semi-automatic knowledge update. Therefore, the supporting module to perform this function has been termed as the knowledge discovery module (KDM) which is proposed as an integral part of KAM. The main contribution of this framework is the elements within the knowledge discovery module which can be seen by comparing Figure 7-1 and Figure 7-2. In the proposed framework, the project life cycle and operational data can be used to generate structured knowledge by the KDM with the primary aim that the generated knowledge should be at least as good as the knowledge provided by domain experts.

Figure 7-2 shows the proposed KOATING framework for the KAM of a Moderator system. This framework incorporates the features of a knowledge based system designed for individual as well as cooperative learning, knowledge reuse, and corresponding update of expert module's knowledge within the Moderator system. The knowledge miners can use many different knowledge discovery and data mining tools to resolve the challenges of identifying, verifying and incorporating new knowledge within the existing expert modules. For example, as shown in chapter 9, the knowledge miners within the KDM of the KOATING framework use techniques of text mining such as text analysis, link analysis, semantic search and dimensional matrix to extract novel and useful knowledge from post project report based textual data of the construction industry. As shown in Figure 7-2, the proposed framework consists of several modules, communication mechanisms and data sources.

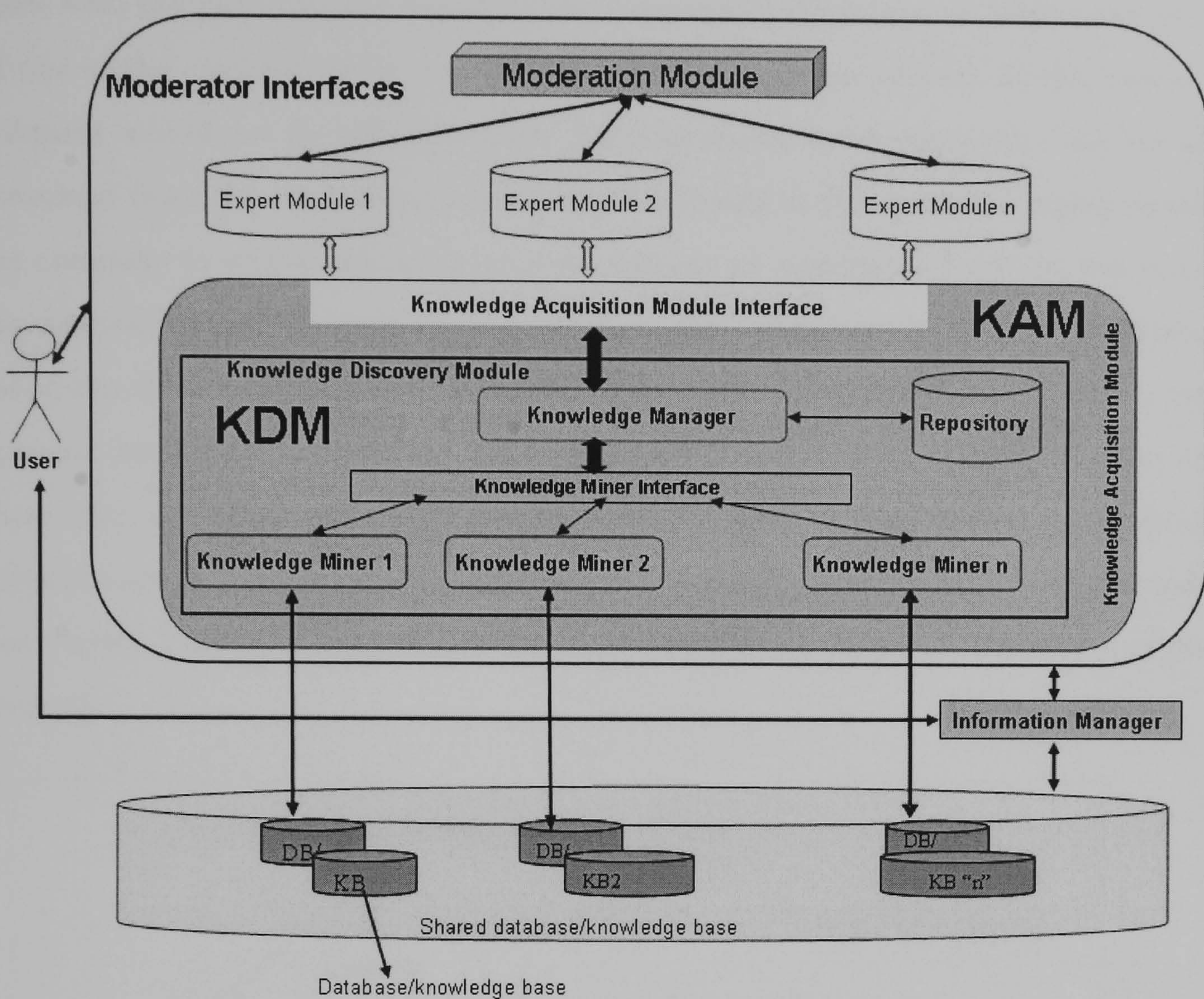


Figure 7-2: Proposed KOATING framework for Moderator [62, 157, 158]

The KAM, expert module and moderation modules are the three major modules of the system. As in the original Moderators, each expert module stores the relevant information and knowledge about a participating project team member. In the KOATING framework, it also stores the discovered knowledge delivered by the KDM. As in the original Moderators, the major function of the Moderation module is to identify the potential project problems and perform the moderation activities. These activities have not been changed in any way by the KOATING framework and therefore are not discussed in detail here. Detailed discussions of KDM as an integral part of the KAM, expert module, moderation module and information managers are given in the subsequent sections. The bottom part of the framework in Figure 7-2 shows the databases of the individual project team members. In any collaborative scenario there are issues of confidentiality and risk which need to be considered as these will potentially constrain what information should or should not be shared. Particularly in business contexts issues of intellectual property and competitive advantage are important. This is indicated in Figure 7-2 as these databases are of two types. The first database contains the data produced during a variety of operations of the company that they do not want to share with the partners. For example, the competitive advantage or intellectual property of one of the company is the processes it produce a unique product. In this context, the company would not be willing to share the data related to its processes. The knowledge generated from this data source would only be stored in the dedicated expert module of the company to enable the moderator to perform its activities effectively, but not share this knowledge directly with other partners in the collaboration. The other database is called the shared database which combines the data sources of partners with a view to generate knowledge to improve the collaboration. The double or single headed arrows show the communication between different modules. The detailed description of communication between various modules is presented in the chapter 8 while modelling the system. Further issues of risk and confidentiality are beyond the scope of present research.

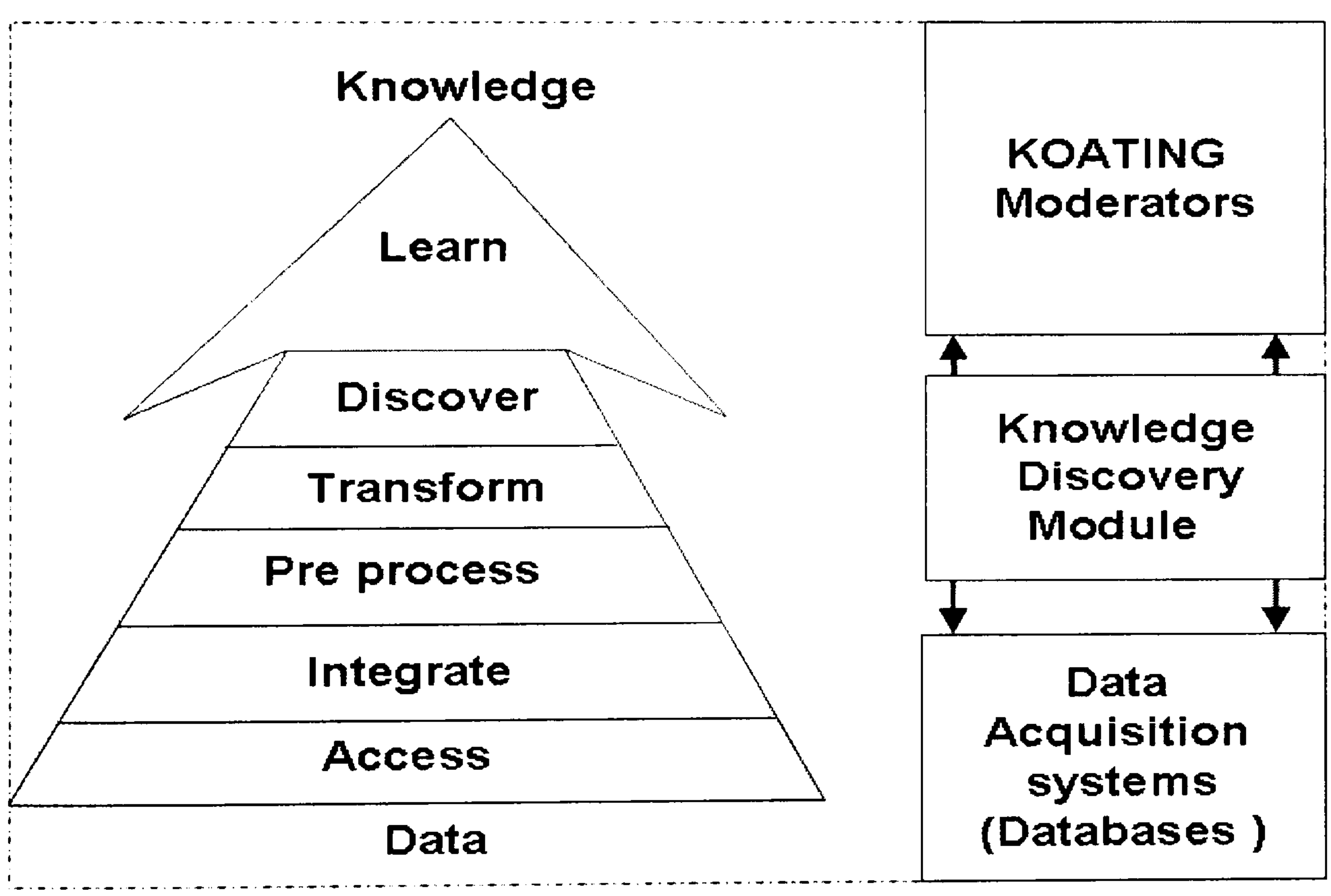


Figure 7-3: Integration of Data Acquisition System, KDM and Moderator.

7.2.1 Knowledge Discovery Module of KOATING framework

The KDM supports the KAM by providing a semi-automated knowledge acquisition mechanism to identify and retrieve appropriate knowledge from available data sources and store them in a format appropriate for further use by the Moderator. The proposed KDM requires its integration with the Moderator as well as the data source. This research assumes that the data source is accessible for analysis purposes by the KDM. Similarly, the KDM is integrated with the Moderator to provide KOATING Moderator service as shown in Figure 7-3.

A data acquisition system could take many forms and it varies from industry to industry. ERP is one of the prominent data acquisition systems. Generally, it provides firms with the ability to gain access to their data through a common graphical user interface (GUI), made possible through client/server technology. In addition, the data acquisition system must be capable of achieving integration by bringing together data from different sources within the firm. This may include disparate databases that exists across different functional units within a supply chain, customers relations management, business to business, procurements and operations etc., helping the firm to gain a more complete and realistic picture of all the data they hold. However, in this research the author is less concerned about the various approaches adopted for data collection and has focussed more on the functionality of the KDM. In this research, the assumption is made that suitable databases already exist within the collaborating companies and the KDM has the ability to access data stored in theses database and transform them through various data analysis techniques. The transformation of data provides the user with a high level view of data thus providing the Moderator with an ability to discover trends, patterns, relationships etc in the data. Once discoveries are made, the KDM supports them through all the period of learning and facilitates the update of the expert module(s) by the KAM thus enabling more accurate, timely and knowledgeable decisions to be made in the future.

Another benefit of integration of the Moderator and KDM can be extended to include the collaboration of multiple enterprises. Firstly, the functionality of the KDM can provide the Moderator with an ability to identify possible business opportunities and create awareness of any opportunity among the potential collaborating partners. This will be demonstrated through an example in chapter 10. Secondly data, information and knowledge need to be shared beyond the confines of the individual's internal environment between partners to meet the competitive demands. By doing this, the Moderator facilitates the decision making by enabling the discovered knowledge to be viewed across the entire extended value chain. This is demonstrated using an illustrative example of virtual e- supply chain in section 7.6 of this chapter.

7.2.1.1 Knowledge Discovery Approach Embedded in KDM

Knowledge discovery process can be achieved through two different approaches namely Data Mining Software Tool Approach (DMSTA) and Data Mining Application System Approach (DMASA) [159]. DMSTA, which is also known as first generation data mining approach involves the application of data mining software tools on ad hoc data mining projects and requires a significant expertise in data mining methods, databases and/or statistics. Data mining tools usually operate separately from the data source, which requires a significant amount of additional time spent with data experts, data imports, pre-processing, post processing and data transformation from various sources. The disadvantages associated with this model include the need for several experts to collaborate in a project and transferability of results and model. This implies that the results and models derived can be used for reporting, but cannot be directly utilized to integrate with other systems [159]. These limitations restrict the application of this approach in the context of the KDM.

In contrast, the DMASA approach primarily focuses on the requirement of knowledge users and decision makers enabling them to view and exploit data mining models. Models can be presented in a user understandable manner through a user friendly and intuitive GUI using standard and graphical presentation techniques. Knowledge can be discovered by focussing on a specific problem domain covered by areas of analysis with the possibility of repeated analysis in periodic time intervals, or when required by the user, or at a particular milestone such as at the end of projects. Several authors and practitioners have recommended this approach for better integration in the business environment and in decision processes [159]. Therefore, this research adopts DMASA approach. Data mining standards and platforms are important issues for any data mining based system, however, data mining standards and platforms are not the current focus of this research. A detailed study of data mining standards and platforms are presented in [160].

7.2.1.2 Process Model of KDM

The process model represents how the knowledge generation and decision making process is supported by the knowledge based system. Determining the process model is one of the key issues for the design of the KDM for Moderators. CRISP-DM (Cross-Industry Standard Process for Data Mining) process model is a data mining process model developed by the industry leaders in collaboration with data mining experts, users and data mining tool providers [161]. The analysis of various other data mining models equivalent to CRISP-DM, identifies CRISP-DM as the most appropriate process model for knowledge based system implementation [161]. In the present context, the CRISP-DM process model has been modified in order to make it applicable to the KDM. Unlike the CRISP-DM model, the process model for the KDM has been divided into seven phases and three stages. Each of these phases and stages includes a variety of tasks. The phases include: domain understanding, data understanding, data preparation, modelling, evaluation, deployment and conflict resolution. The sequence of the phases is not strict and moving back and forth between various phases is always required. Based on the outcome of each phase the next phase or a particular task of a phase that needs to be performed can be decided. The arrows indicate the most important and frequent dependencies between phases. The inner circle shows the cyclical nature of knowledge discovery process itself. It means that a knowledge discovery process continues even after knowledge is discovered during the deployment phase. The lessons learned and experiences gained during this whole process can benefit the subsequent data mining processes. It is important to mention that the elements of KOATING framework have been shown as an actor in Figure 7-4. As shown in Figure 7-4, the modified structure of CRISP-DM has been adapted to the needs of the KDM as three stages:

- the preparation stage
- the knowledge production stage and
- the implementation stage

As shown in Figure 7-4, the preparation stage of the process model prepares the area of analysis for production and implementation uses. This stage focuses on performing the first five phases, i.e. from domain understanding to evaluation in an iterative manner. The major reason the multiple iterations may be carried out is to achieve step by step improvement in all the phases. For example, a slight redefinition of an objective might be required in the domain understanding phase depending on the results of the evaluation phase. Similarly in the data preparation phase, the procedures required for the recreation of datasets might be improved. Datasets must be created automatically on a periodic basis, say every night based on the current state of the existing databases, data warehouse and the transactional data. The problems identified in the data preparation phase may demand changes in the data understanding phase.

In the modelling and evaluation phase the models are created and evaluated for multiple times with an aim to fine tune the data mining algorithms through finding the adequate values of the tuning parameters for the algorithm. Therefore, it is necessary to perform enough iteration in order to monitor the level of changes in datasets and models acquired. With this approach, the stability of the data preparation phase is reached and optimal or near optimal values of parameters of the data mining algorithms are identified. In this manner, according to results gained in the evaluation phase through multiple iterations, the preparation stage can either reject the area of analysis due to insufficient quality of model or approve it with or without a slight modification in the objectives in the domain understanding phase and consequently changes in the other phases.

The second stage is called the knowledge production stage, which mainly focuses on modelling, evaluation and deployment; it does not mean that other phases are not encompassed in the production stage. At this stage the models are created and evaluated multiple times with an aim to fine-tune the algorithms and parameters used. These functions are performed by knowledge miners with support from knowledge managers and repositories. These are discussed in further detail in the next sections. The third stage, called the implementation stage, involves the updation of the fine-tuned knowledge into the expert module and consists of the modelling step through to the conflict resolution step.

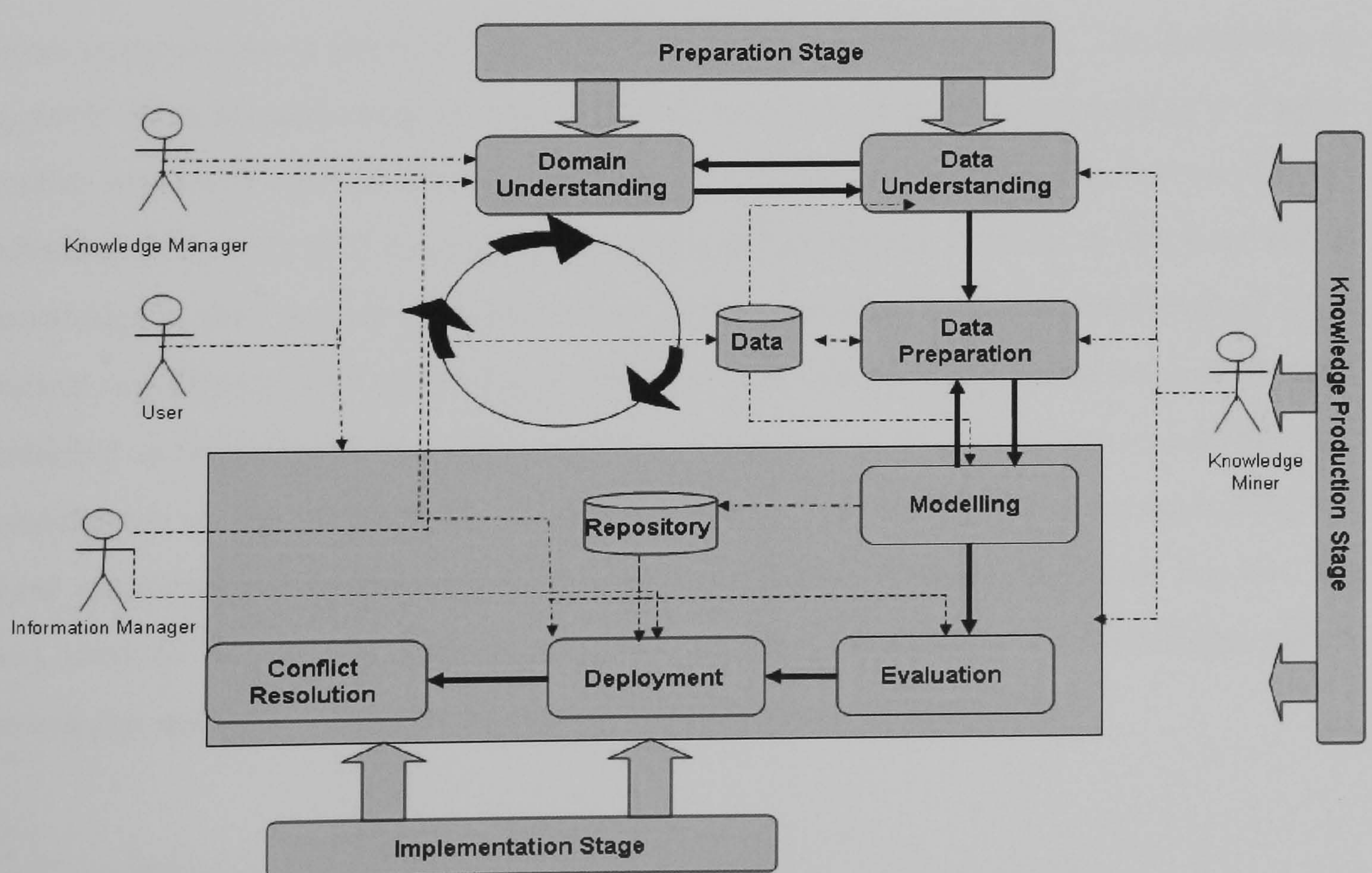


Figure 7-4: The Process Model of Knowledge Discovery Module

This stage provides inputs to all the prior stages based on the extracted knowledge. This stage requires interaction from users, knowledge miners, knowledge managers and repositories as shown in Figure 7-4. Conflict resolution between the existing knowledge and the discovered knowledge requires input from users based on their knowledge of the domain.

The development of a process model for the KDM provides a basis for the development of knowledge integrated moderator services. As shown in the KOATING framework in Figure 7-2 and the process model Figure 7-4, KDM mainly consists of four components:

(1) Knowledge Miners: to actively discover hidden data relations, patterns or models within data associated with the team member.

(2) Knowledge Manager: to provide coordination and communication of knowledge, and management support between various components of the KDM;

(3) Information Manager: to implement the changes being made to each expert module during the project process as this relates to changes in the information and knowledge stored about individual team members. A detailed discussion and its relation to KAM is discussed in section 7.3.

(4) Repository: to provide a mechanism for current intelligence, temporary storage of rules and tuning parameters specific to an algorithm.

These components of the KDM have been treated as intelligent agents. The definition of agents is still a research issue and the boundary between an intelligent program and agent overlap with each other as can be seen from several literatures [162-166]. Several authors including [135, 167, 168] have used agent based data mining systems to discover useful knowledge in the contexts of multi criteria decision making, e-commerce and shop floor control respectively. One of the major reasons to choose the agent based modules is their flexibility to interact with the other modules of the system. However, this is not the main contribution of this research and therefore less emphasis has been put on describing the agent oriented architectures and their integration issues. Considering these aspects, the next sections discuss the internal structure and the functionality of knowledge miner, knowledge manager, information manager and repository respectively.

7.2.2 Knowledge Miners

Knowledge miners are an integral part of the KDM and their major functionality is to extract patterns, relationships and useful knowledge from the operational databases to equip, build, populate and update the expert modules associated with the team members. In this manner, each expert module consists of knowledge about an expert, their objects of interest, priorities etc., which can be reused to support the moderation process. In the present context, the performance target of a knowledge miner is to generate knowledge which is as good as or even better than the knowledge associated with human experts in the same situation with the same input datasets. At present, a successful application of data mining generally relies on the experience and expertise of both the data mining expert and the domain expert, and therefore is a semi-automatic process.

Figure 7-5 schematically shows the architecture of a knowledge miner, which acts as an intelligent agent. It consists of a data interface, operational facility, agent knowledge base and knowledge interface. The knowledge interface manages the communication between the knowledge manager and the knowledge miner. The communication is based on message passing based on a shared ontology; this means that when the knowledge miner receives messages that are represented in a common ontology, the knowledge miner interface converts these messages into local format based on the common ontology. In a similar way, when the knowledge miner sends messages to the knowledge manager, the knowledge interface translates them into a common format first and then sends them to the knowledge manager. A detailed study of ontology models are presented in [169] and the complexity associated with this approach is beyond the scope of this research.

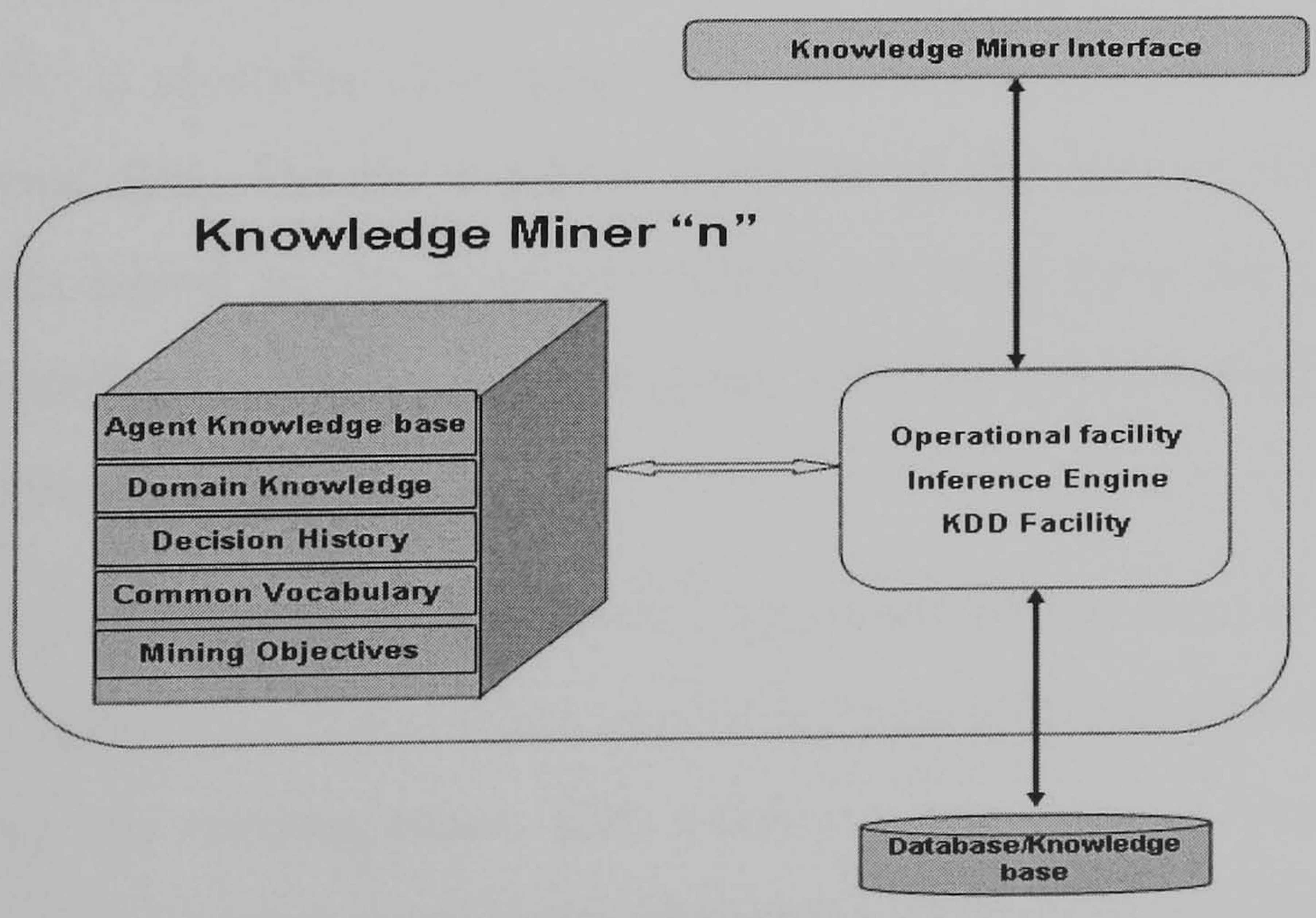


Figure 7-5: Architecture of the Knowledge Miner

The operational facility component is the central control and action part of the knowledge miner. It consists of sub components such as an inference engine, and KDD Facility. The KDD facility is one of the main components of the knowledge miner and it could be implemented in several different ways, e.g. as code, software or an expert system. It performs the mining task of discovering causal and interesting relations from the dataset and presents them in a form compatible with the knowledge manager. This component must carry out several functions that are required to carry out the data mining functions mentioned in section 5.2.3.1.

The knowledge required to perform the mining task, common vocabulary; knowledge about different users, past decisions, mining objectives and domain knowledge are stored in the agent knowledge base. The data interface component of the system provides a mechanism to extract data from the external data source such as data acquisition systems. In order to perform the mining task in a variety of application areas (as described in Appendix 2), KDD facility uses a variety of tools, techniques and functions as discussed in chapter 5. After the knowledge miners complete their tasks, they send the knowledge mining results to the knowledge manager using the knowledge interface and the knowledge miner then terminates.

The functionality of the process model (Figure 7-4) accomplishes its main objectives of knowledge discovery through the implementation of 4 modules in the KDD. These are described as follows:

1. *Data Acquisition Module*: This module performs three major tasks, firstly it acquires data from the current data acquisition systems/data warehouses of the company.. Secondly, it identifies data types such as structured numerical data or unstructured text based data. Thirdly, it selects a sub set of the data or focuses on a subset of its attributes based on the objective. Finally, it must have the capability to acquire an understanding of the application area, relevant prior knowledge and the goals of performing the KDD.
2. *Data Preparation Module*: This module performs all the functions needed to transform the raw data into a form which can be fed into different algorithms. This involves (1) replacing any missing values with a suitable one such as mean, median or one with maximum frequency or removes incomplete records if no suitable substitution can be used for missing values. (2) Eliminating noisy data to improve the efficiency of algorithms and improve the accuracy of the results, (3) If required, it normalizes the

database in order to avoid duplication, and possibly eliminate various kinds of logical inconsistencies that could lead to loss of integrity of database (4) Transform the data to different values depending on the problem domain and, (5) create derived attributes in order to reduce the computational burden. The identification of critical process attributes may make it possible to achieve a desired level of quality knowledge by optimizing only a small number of variables.

3. *Modeling Module*: based on the goal of data mining, this module performs one or more combinations of the data mining functions such as classification or prediction, etc,. To perform these functions, this module consists of a set of data mining tools and techniques. A detailed study of such techniques and functions are presented in chapter 5. These algorithms are applied to obtain and extract the patterns from the databases.

4. *Knowledge Evaluation Module*: this module evaluates the generated knowledge based on certain criteria to make sure that it is novel and useful. It applies statistical techniques to determine the validity of the identified knowledge over the universal set.

In this manner, one can see that the provision of these facilities validates the process model developed in the earlier section as they provide all the functionality to achieve all the stages mentioned in the process model.

Moderators require two types of update of the expert module (1) regular or periodic update and (2) User requested update. Periodic update is required as there is a change in the database and correspondingly new knowledge can be generated from the updated data. User requested update is required when the user finds some unusual knowledge or wishes to identify a specific type of knowledge. Considering these requirements of the Moderator system, there are two types of knowledge miners, which work in two different manners. The first type is termed as Periodic Knowledge Miner and the other type is called Task Oriented Knowledge Miner. A periodic knowledge miner starts at the beginning of the life cycle of project. Usually a periodic knowledge miner works periodically and generates knowledge based on changes in the database of the corresponding team member. For example, say every day at 3 AM in the morning data acquisition system updates data warehouse. The addition of data triggers could enable a periodic knowledge miner to also perform the mining task every day. Another type of periodic knowledge mining may occur whenever a project finishes and new data such as a Post Project Report is added to the database. Addition of the report will trigger the knowledge mining facility to perform the knowledge discovery process. Periodic knowledge mining will follow a sleep-work-sleep-work cycle and will be destroyed when the entire moderator system terminates. A similar approach has been used by [167] for group decision making purposes. In contrast, task oriented knowledge miners are activated for a particular data mining task. This may be based on a request from the knowledge manager when an expert's role or interest changes or whenever something unusual happens, or the knowledge about an expert is identified to be incorrect. Task oriented knowledge mining also provides the user with an opportunity to capture some particular type of organizational knowledge from defined data. The user passes this message to the knowledge manager, and the request is then relayed to the knowledge miner. After the knowledge miner has completed the task, the results are sent to the knowledge manager for further processing.

7.2.3 Knowledge Managers

The knowledge manager acts as the heart of the proposed KOATING framework and plays a vital role of manager, mediator and communicator between the different knowledge miners, the information manager, the EMs and the repository for knowledge sharing. The knowledge manager is an integral part of the KAM and makes the decisions to create or delete EMs based on the extracted information, recommendations from the various constituent elements and information about the project team. The knowledge manager mainly consists of four components: miner interface, knowledge acquisition interface, functional facility and the knowledge manager agent knowledgebase that provides support for localized reasoning. The basic structure is represented in Figure 7-6.

The knowledge manager checks what special types of knowledge are relevant to specific EMs and consequently which types of databases and files are appropriate to mine in order to update the knowledge content of any particular EM. It activates the responsible knowledge miner to perform the mining task. When mining and update tasks are completed, it stores the knowledge about these activities into the repository for possible future use and transfers the relevant new or updated knowledge into the EMs through the knowledge acquisition interface. The updated EMs can then be enabled, so that they can be used in the ongoing moderation processes. The operational facility provides the mechanisms for knowledge transactions.

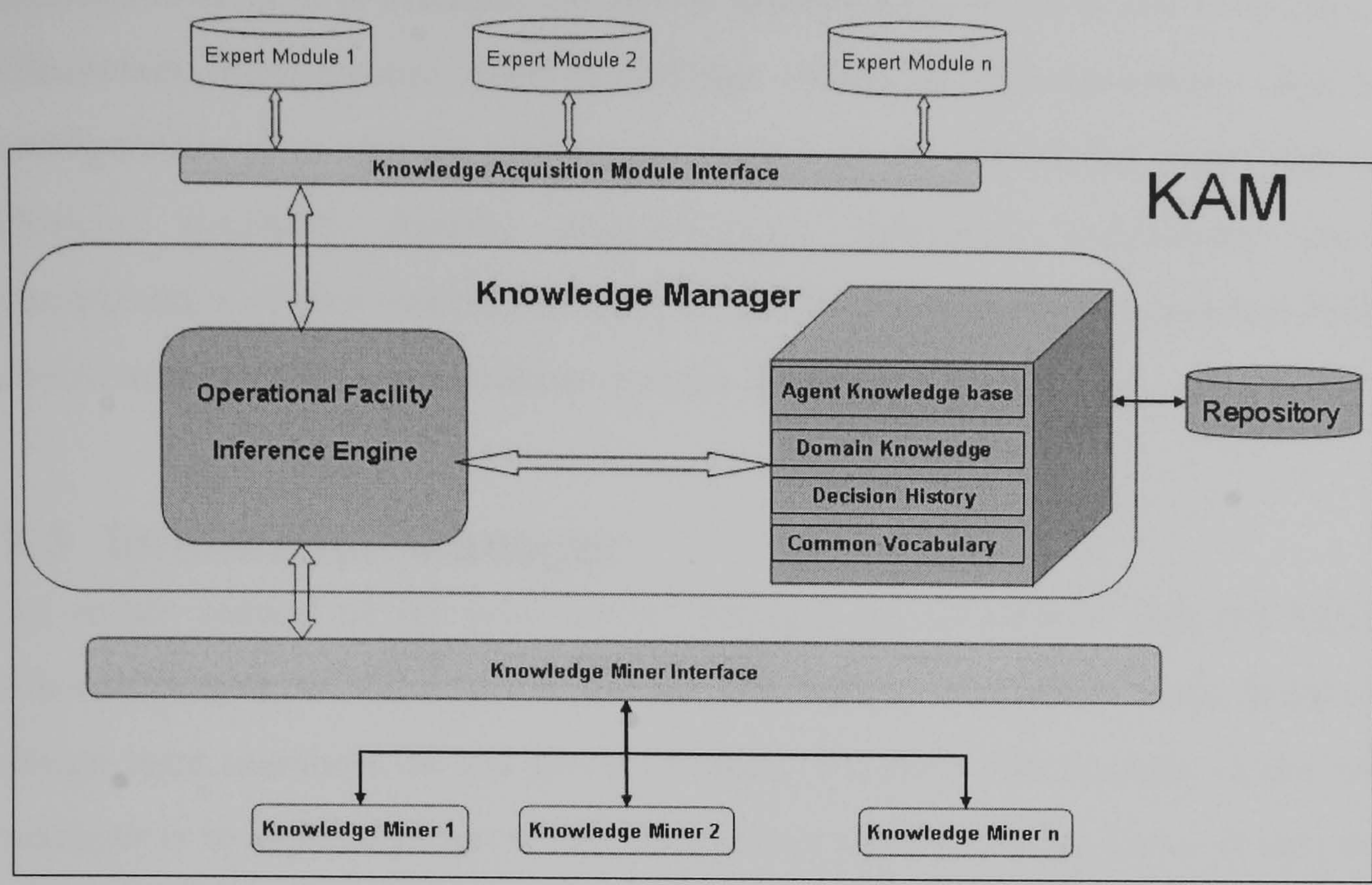


Figure 7-6: Structure of Knowledge Manager

The knowledge manager mediates requests from the user and analyzes these requests through its localized knowledge and inference engine and then initiates a knowledge miner to perform the desired task through the miner interface. It communicates the type of knowledge to be mined i.e. the function that needs to be performed on the data. The information manager, the repository and the knowledge miner are interfaced with the knowledge manager. It also communicates with the knowledge miner to start the task oriented mining process whenever the knowledge about a particular expert is not correct or appropriate. It also communicates the task driven knowledge miner to identify particular knowledge whenever something unusual happens or when information about a particular situation is needed.

7.2.4 Knowledge Repository

The repository temporarily stores the mining results and tuning parameters and helps the knowledge manager by providing a set of knowledge required by knowledge miners. When the knowledge manager receives the mining request for information, it first queries the repository to see if relevant knowledge pertaining to the request has already been discovered. If it has not been found, then the knowledge manager initiates the knowledge miner(s) to mine the appropriate knowledge/data bases. In addition, the repository provides the mechanisms for using a common vocabulary. As all the components within the KDM work on the same problem domain and communicate with a set of valid message objects, it is essential for all the components of the KAM to share a common vocabulary. Furthermore, meta-knowledge stored in the repository, such as system configuration, (e.g. various mining parameters of the knowledge miner like number of clusters, similarity criteria, interestingness measures, confidence etc.), design consideration of data mining framework such as generic software and knowledge can be shared and reused by other similar systems in the future.

7.3 Information Manager

An earlier version of the information manager was interfaced with the MSEM in the Mission project, to share and access the information stored in Oracle databases for the design team members. In the present context, the main functionality of the information manager is to share and access the shared data and knowledge bases. It also notifies the knowledge manager about particular contributions or changes to the shared data desired by particular project team members. The information manager also maintains its original functions during the Moderation process as it signals the Moderator whenever a change in the project data is recorded in the project database. However this does not affect the knowledge acquisition process, so will not be considered further here. In this research, the information manager also assists the knowledge manager in deciding when to create, delete or update a particular expert module. This will happen whenever a new project team member joins a project or when existing contributors are changed in any way, or when new knowledge or experience is identified within the enterprise which may influence how a contributor to the project will make his or her project decisions.

7.4 Expert Module

Moderators use expert modules (EMs) to represent each team member and therefore EMs are populated with all the knowledge about the team members in the collaborative projects. Hence, the collection of EMs provides the moderator with the background knowledge that it requires to support the multidisciplinary team. Harding [43] mentioned that the information and knowledge for Moderators should be captured in a flexible manner to ensure its reusability in a variety of situations by a variety of users or applications over a period of time. They have used the KRM [43] concept to represent the knowledge as a groups of associated interactive objects called modules. The objects can exhibit different types of behavior including collecting information, processing information, checking the truth of specific conditions, and communicating with users etc.

The KRM can be used to produce a knowledge base by storing the objects in a object oriented database. Based on earlier Moderator concepts, in the present context, knowledge about individual team members, knowledge of their area of interests, their competencies, and the knowledge about changes that are important to them and actions that need to be taken when such changes occur are stored in objects called Expert Module. The interactive objects associated with Expert Modules come from various classes, including ruleset, rule, condition and action objects. The processing of knowledge is achieved by message passing between instances of these classes.

In the present context, particular types of knowledge can be mapped in the form of rulesets and particular knowledge can be mapped as rule. The KDM can produce a variety of rules including, classification rules, prediction rules, association rules, which will be in the form of IF (Condition) THEN (Action). By producing the IF-THEN rules, KDM follows the KRM concept and updates the knowledge which can be applicable to the Moderator. When the Moderator updates an expert module, a notification is sent to the user indicating of what knowledge have been updated and also of any conflicting rules found in the EM. In this manner, generated knowledge needs to be verified by the user before it is used in the moderation process. In addition, it is also feasible to update the area of interests and competencies of individual team members based on their past work or performances. As discussed in chapter 9, this feature will be achieved thorough the application of text mining. Again, the verification of this discovered knowledge is required from the user. Section 7.6 of this chapter provides an illustrative example of how the knowledge can be generated in the form of IF-THEN rules, which can be further updated based on changes in the dataset.

It is important to integrate the proposed framework with the current state-of-art Moderator and to achieve this, the next section discusses the Universal Knowledge Moderator (UKM). Here, the major emphasis is to show how the KOATING framework can be integrated in the context of heterogeneous data sources.

7.5 Universal Knowledge Moderator for E-Supply Chain

The complexity of moderator technology increases when manufacturing projects are large and members are globally distributed in the context of an extended enterprise (EE) or virtual enterprise (VE) [9]. Manufacturing projects operating within EE and VE environments face additional problems that different information models are likely to be used by different parts of the manufacturing project teams. Supply chain partners inevitably use different vocabularies and terminologies in their work resulting in misunderstandings and confusions. Moreover, the escalating use of web technologies has also accelerated the growth and complexity of manufacturing digital information. The consequent enormous amounts of heterogeneous data (e.g. structural heterogeneity or semantic heterogeneity) make it increasingly difficult to communicate between different project teams and organizations. In response to this problem and to achieve true information interoperability, Lin *et al.* [55] adopted the ontology and semantic web technologies within a Manufacturing System Engineering Moderator (MSEM) to enable semantic interoperability across extended project teams [57].

This research builds on previous work on the MSEM and therefore, this section of the research proposes how the MSEM may be extended to provide knowledge discovery for globally distributed and collaborative e-supply chains on the semantic web. The aim of this research is to develop and establish a flexible method for knowledge discovery from semantically heterogeneous data for the moderation of project teams in globally cooperative e-manufacturing chains by integrating KOATING framework with state-of-art Moderator. The proposed universal knowledge moderator should therefore be able to:

- Analyze and define the specification of a common manufacturing ontology for the manufacturing industry in an ontology server.
- Enable WWW information exchange between partners in cooperative manufacturing chains via common mediated meta-models across different disciplines within engineering project teams through semantic mapping.
- Enable the moderator's KAM to incorporate "learning", updating and reuse elements which exploit knowledge discovery techniques.

Previous research by Lin [55, 57] has addressed the challenges of different partners within an e-supply chain using different vocabularies and terminologies and therefore the first two of the above objectives are beyond the scope of this thesis. This section now sets the KOATING framework into the context of a Moderator to support collaborating working in extended supply chains or similar organizations.

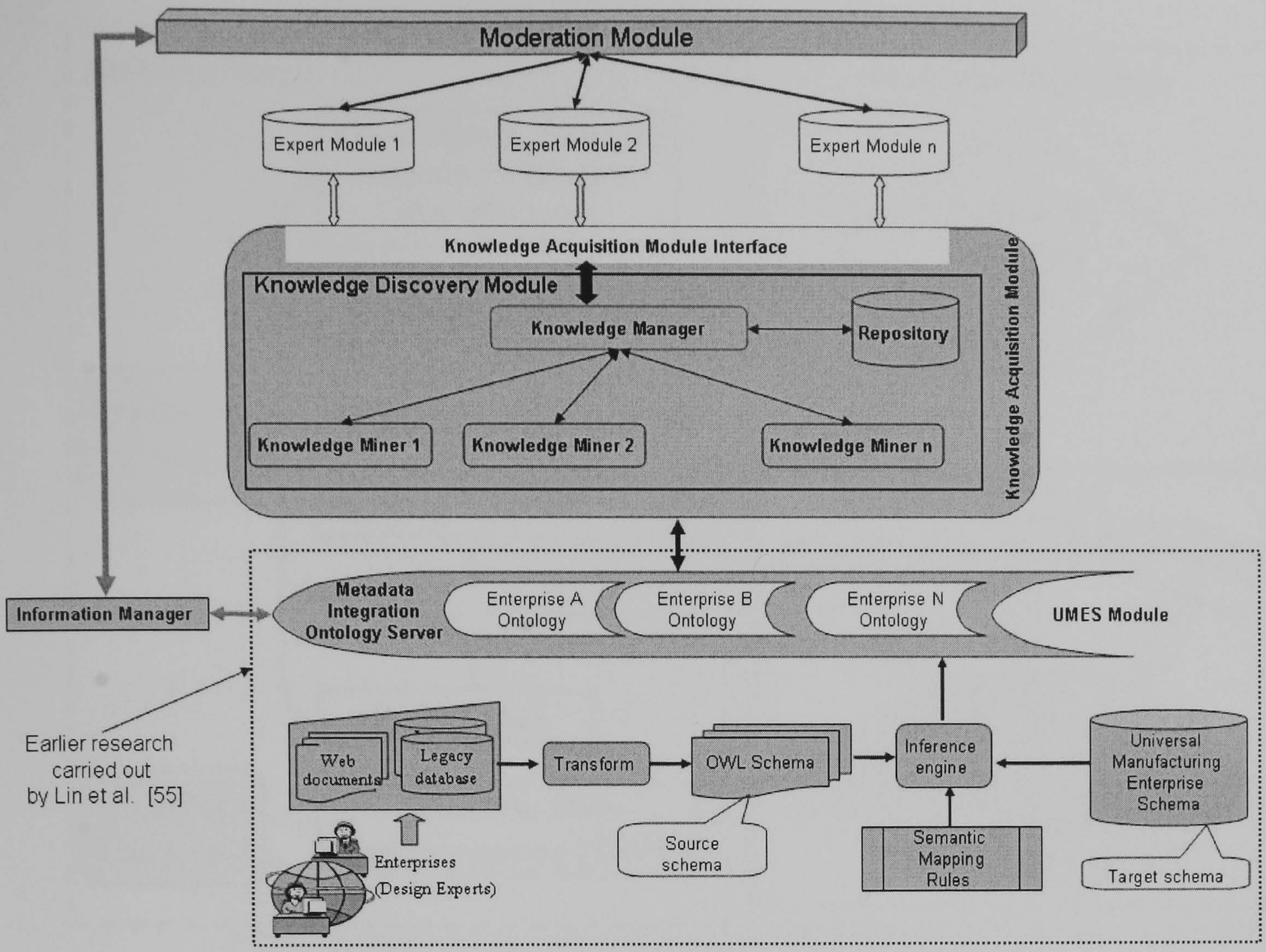


Figure 7-7: The Architecture of Universal Knowledge Moderator (UKM)

The main contribution of this thesis is therefore how to integrate KOATING framework with the other different element of the existing EEMSEM services to provide the above-mentioned functionalities. The previously listed research objectives are discussed in the context of an architecture model for UKM to enable semantic integration of geographically distributed knowledge discovery services. Three main modules have been identified as shown in Figure 7-7:

- Universal Manufacturing Enterprise Schema (UMES) Module [55,57]
- Knowledge Discovery Module
- Moderation Module

This is an extension of a previous architecture for MSEM that was implemented during the MISSION project, and provides a means of moderating semantically heterogeneous databases distributed over the Internet

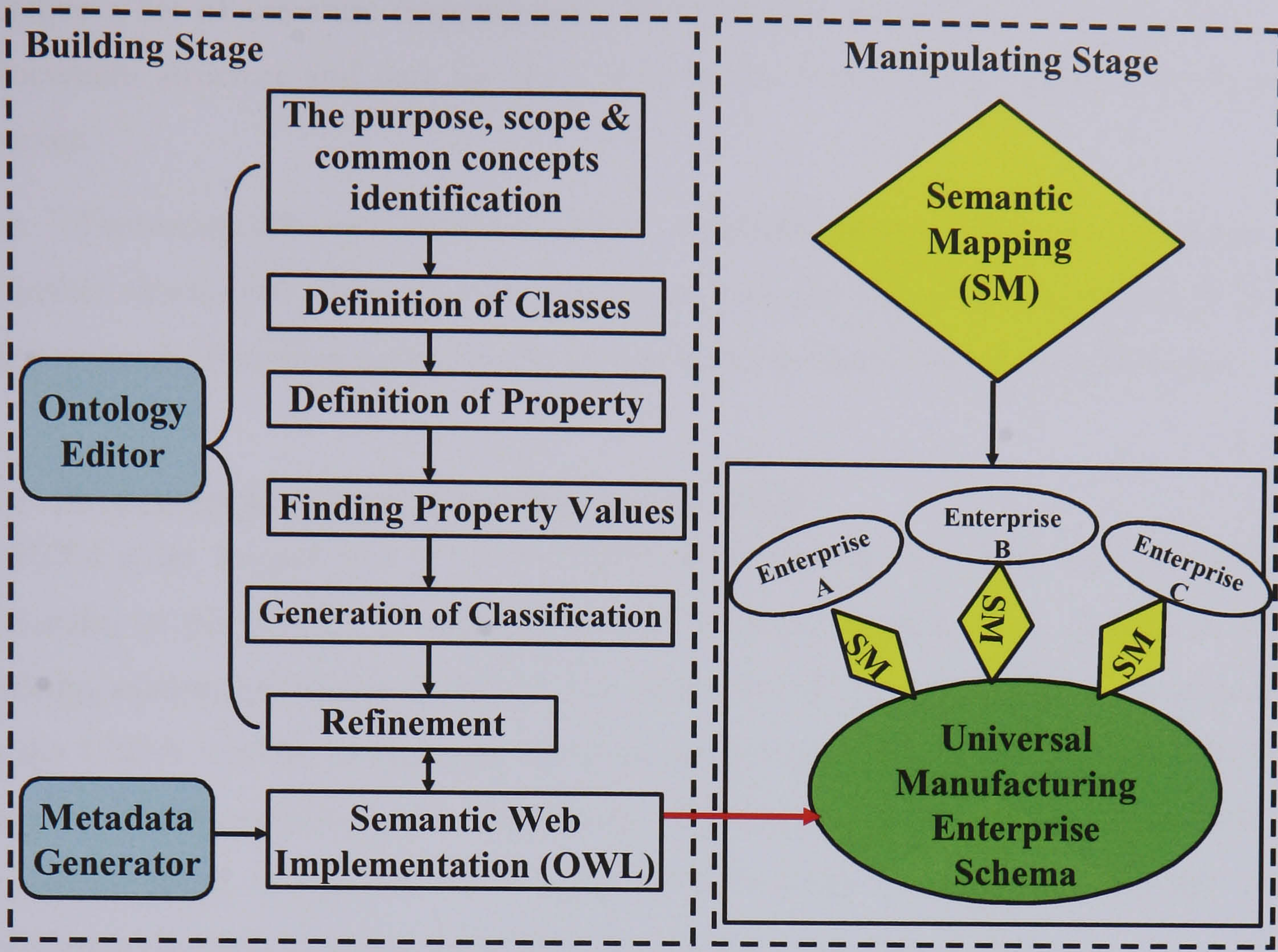


Figure 7-8: The Universal Manufacturing Enterprise Schema - Mediated Ontology ([57]).

7.5.1 Universal Manufacturing Enterprise Schema Module

This architecture of UKM uses a dynamic mediated and shared ontology model for manufacturing enterprises, in order to achieve information interoperation for a UKM within an internationally collaborative environment. The mediated ontology model may involve simple logical reasoning for semantic and syntax mapping. The methods of the UMES are listed and briefly described as follows:

- Analyze and identify the terminology, representation and classification of the manufacturing system for UKM activities in the context of a global high tech industry e.g. PC, IC manufacturing, operating in a globally co-operative e-manufacturing chain.
- Convert the UMES into a web-based ontology language, e.g. Resource Description Framework (RDF) and Web Ontology language (OWL).

- Define a set of semantic mapping rules for automatic reasoning of heterogeneous document structure and data for the UMES in the Metadata Integration Ontology server.

Figure 7-8 schematically represents the UMES. A detailed study of these functionalities and further discussions are presented in Lin *et al.* [55] and therefore not the focus of present research. Therefore only brief details have been included here for completeness.

7.5.2 Knowledge Discovery Module of UKM

The KDM is an integral part of the KOATING framework. The basic structure and functionality of this module remains the same as discussed in section 7.2. It consists of knowledge miners, knowledge managers and miners interface. This module is interfaced with the UMES module to deal with the semantic heterogeneity of heterogeneous data sources and interoperability issues. Knowledge discovery is based upon the defined and common ontology, so that the KDM generates knowledge in appropriate language or vocabulary to be used to update and enrich the expert module knowledge.

7.5.3 Moderation Module

The major functions of the Moderation Module are to identify the potential project conflicts and to perform moderation activities. It does so by continuously reviewing the current state of the activities associated with the project and information about recent project decisions, and comparing this with the knowledge it has about team members' interests and requirements as stored within the EMs. The moderation process is activated whenever a project decision is made and this is identified by a change being made to the project information within the shared database. The information manager can notify the UKM of each change. The UKM checks the interests of the team member in its knowledge base by examining the EMs. If the Moderator finds that one or more team members have an interest in the current type of change, the interested team members are contacted by the UKM and it remains in dialogue with these team members until conflicts are resolved.

Clearly when team members came from different cooperating companies, it is very important that information is understood by different parties and the use of ontologies has previously been covered in detail in [55, 57, 157].

7.6 An Illustrative Example of KOATING Framework on E-Supply Chain

The e-supply chain is the communication and operations backbone of a virtual network that links suppliers, business partners and customers together as one cohesive collaborative entity. A virtually enabled supply chain network is a series of value added processes or stages owned by one or more enterprises, starting with a material or information supplier and ending with consumers. In this structure, each intermediate stage is a supplier to its adjacent downstream stage and a customer to its upstream stage. Although the participants in the chain can play various roles, all their relationships are limited to supplier and customer roles. An open fast communication mechanism is essential for the companies entering into supply chain network activities, allowing its members to jointly forecast, develop, produce, synchronize and deliver their product or services, and anticipate dynamic customer requirements. A typical example of a e-supply chain is schematically shown in Figure 7-9.

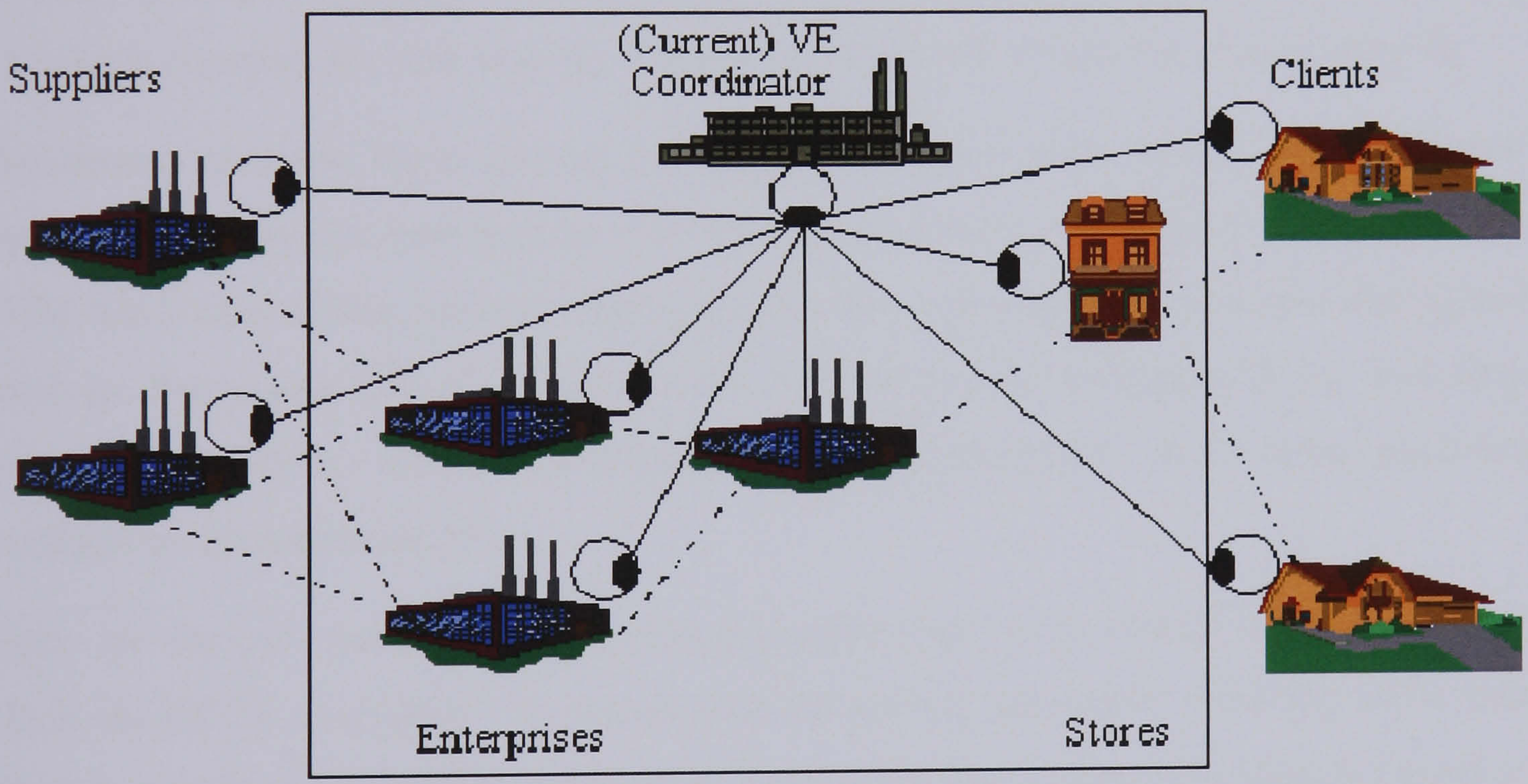


Figure 7-9: A generic view of Virtual e- Supply Chain

This example assumes that a VE has been created by a leading UK automobile manufacturer (X) which operates in UK and Europe. The automobile enterprise "X" wishes to contract a supplier to produce 2 parts (P_a and P_b) in order to build a new prototype in the UK. However, no local manufacturer can be found with enough resource to meet the demand. The solution adopted was to form a collaborative virtual-e-chain with 4 enterprises including suppliers and manufacturers. 4 EMs are created, each containing knowledge about one of the enterprises in the e-supply-chain. These will be referred as supply chain agents in this example. The UKM has access to all of these EMs and therefore has knowledge about the each Supply chain agent.

Supply chain operations start when an order for the two products (P_a and P_b) has been sent to the 3 suppliers (S_1, S_2, and S_3) through a web interface with requirements such as lead-time, quantity, and product type. Two different parts P_{a1} and P_{a2}, and P_{b1} and P_{b2} are assembled to produce the products P_a and P_b respectively. For product P_a supplier S_1 can only provide $P_{a1,}$ P_{a2} has to be ordered from S_3. Similarly, for product P_b, S_2 can only provide P_{b1}, P_{b2} has to be ordered from S_3. On the other hand, S_3 can supply all the parts P_{a1}, $P_{a2,}$ P_{b1} and P_{b2} to produce P_a and P_b. All the suppliers are able to produce different products with different lead times, quantities and prices. All the suppliers use different terminologies for the same context. For example S_1 uses lead time, S_2 uses due date, and S_3 uses delivery time for the same context of delivering the product to the customer. Similarly, S_1 uses quantity, S_2 uses number of products, and S_3 uses number of pieces for the same context. The product information for each supplier is given below in Table 7-1, which contains the data that the supplier wishes to share with the Coordinator X.

Therefore, a semantic heterogeneity for the product information exists, where different suppliers use different terminologies for the same context. The UMES module of the UKM takes care of this semantic heterogeneity by a developing a common and agreed ontology. For example in the present context ontology can be developed for lead time, due date and delivery time. More details about the development of common/mediated ontology are discussed in [57].

There are several combinations of price/quantity/time over several ranges of values which should be considered to ensure that the orders are placed with the most cost effective suppliers. The UKM can help to ensure that this is done providing it has up to date information about current price/quantity/delivery combinations that are best for each supplier. All the EMs must contain the knowledge in the form of rules stating that which combinations of supplier are cheapest. This expert module also contains the items of interest for each supplier such as orders detail, quantity, lead time, due date and price etc. The focus of this research is limited to the knowledge acquisition aspect therefore this example only shows that how knowledge can be generated for the EM associated with company X from shared database of suppliers. In present context, the data presented in Table 7-1 is used as shared data. There is a need to transform these data into knowledge in order to obtain the effective combinations of quantity, price and due date. This knowledge can then be stored in the expert module for further Moderator activity.

Table 7-1: Supplier details and their capabilities for producing products

Product information for supplier S_1		
Lead Time	Quantity	Unit price (£)
1-4	1-30	48.5
5-8	1-100	47.5
9-12	1-150	45
Product information for supplier S_2		
Due date	Number of products	Cost per product
1-3	1-10	49
4-7	1-70	48
8-12	1-150	44.5
Product information for supplier S_3		
Delivery Time	Number of pieces	Selling price per piece
1-3	1-50	50
4 -8	1-100	46.5
9 – 12	1-150	46

In order to discover the knowledge, knowledge manager instructs the knowledge miner to find patterns, relationships and rules within the shared data associated with supply chain agents considering minimal price/product. The mining engine component of knowledge miner finds rules for minimum price. In the present context, IF-THEN rules were discovered. For example for the given data set the discovered rule may be as follows:

1 *IF (LeadTime>8) THEN Select Supplier* S_1 & S_3.

2 *IF (Quantity ≤ 30) AND (LeadTime≤3) THEN Select Supplier* S_1 & S_3.

3 *IF (Quantity ≤ 100) AND (4≤LeadTime≤8) THEN Select Supplier* S_3.

4 *IF (30<Quantity ≤ 50) AND (LeadTime≤8) THEN Select Supplier* S_3.

5 *IF (LeadTime>7) THEN Select Supplier* S_2 & S_3.

6 *IF (Quantity ≤ 10) AND (LeadTime≤3) THEN Select Supplier* S_2 & S_3.

7 *IF (Quantity ≤ 100) AND (4≤LeadTime≤8) THEN Select Supplier* S_3.

8 *IF (10<Quantity ≤ 50) AND (LeadTime≤8) THEN Select Supplier* S_3

These rules are stored in the expert module of Company X and need to be checked whenever a order is raised. Similarly, expert module of suppliers S_1, S_2 and S_3 must be populated with knowledge about its area of interests. For example, for $S_1$the knowledge in the EM would be:

IF (Lead time ≤ 3) AND (Quantity ≤ 30), THEN notify of order.

IF ($9 \leq$ Lead time ≤ 12) AND (Quantity ≤ 150) THEN notify of order in competition with S_3

Similarly for S_2, The knowledge in the corresponding EM would be.

IF (Lead time ≤ 3) AND (Quantity ≤ 10), THEN notify of order.

IF ($8 \leq$ Lead time) AND (Quantity ≤ 150) THEN notify of order in competition with S_3

For S_3, Knowledge in EM would be

IF (Lead time ≥ 4) AND (Quantity ≤ 100), THEN notify of order.

IF ($8 \leq$ Lead time ≤ 12) AND (Quantity ≤ 150) THEN notify of order in competition with S_1 & S_3.

These are very simple rules and could be generated manually, but, consider a situation where several other qualitative and quantitative attributes such as supplier reputation, quality of product, physical location etc, are considered. In the present context scenario, a decision tree algorithm can be used by the data mining engine of a knowledge miner to generate these rules. These rules need to be updated whenever the data changes. It means that whenever the supplier changes their capability, the dataset will change. For example, after a few orders supplier S_1 may have improved their way of production and be capable of producing more products at reduced cost. Based on this fact, they have changed the information related to the product and the combination of lead-time, quantity and price. Changes in the dataset will trigger the knowledge manager to prompt a message to the knowledge miner to initiate the mining task. The knowledge miner will therefore apply its data mining algorithm to extract new rules, patterns and relationships, and thereby generate and update new knowledge within the expert modules. At this stage, if it finds that there is a conflicting rule, it will trigger a message to the user to resolve this conflict based on its domain knowledge.

Now when the e-supply chain is operating and production is going on, the UKM discovers a delay in the delivery of part P_{a1} by supplier S_1. Details of each delivery and order are stored in the company's databases so regular updates of UKM knowledge could identify that there is an error in the usual rules as one of the supplier has been delivering "later" than the quoted date. This delay is critical to the lead time requested by the customer. This delay can be a hindrance to the successful completion of the order by the suppliers involved. Therefore, UKM must notify this delay to relevant supply chain agent responsible for Supplier S_3 and Company X. This alert can be sent in the form of an e-mail about the problem occurrence.

In this case, to overcome this problem UKM may recommend company X to:

- Send the order to an alternative supplier or;
- discuss the quality of deliveries with the existing supplier and negotiate new terms or;
- Increase the working hours of corresponding supplier.

There might be situation where the manufacturing enterprise "X" needs to collaborate with another supplier 4. In this case, chain will be increased based on an agreement of the entire supply chain. In this case, knowledge acquisition module will generate a new expert module corresponding to supplier 4. Above mentioned example is just an instance of several activities involved in the operation of virtual enterprise supply chain.

7.7 Summary

Competitive advantage can be gained through targeted exploitation of proprietary knowledge of all types, but especially of expertise and experience. In earlier research projects Moderator technology, in the form of knowledge based software support systems, has been successfully demonstrated in both the product and manufacturing system design domains. However, knowledge acquisition, learning and updating of knowledge has not previously been studied fully. Therefore this chapter presents a KOATING framework to provide semi- automated knowledge acquisition for moderator technology in collaborative projects to update the expert modules. This enables the reuse of discovered knowledge from operational databases within collaborative projects and facilitates the exploitation of the right knowledge at the right time in the right context.

In addition, a Universal Knowledge Moderator (UKM) system, consisting of a Universal Manufacturing Enterprise Schema Module, Knowledge Discovery Module and Moderation Module, has been proposed to improve the moderation activities. The Universal Manufacturing Enterprise Schema Module enhances the interoperability of the chain on the semantic web. This shows how the KOATING framework can be integrated with the state-of-art Moderator.

The proposed KOATING framework will facilitate the decision making and moderation process through the following multiple capabilities.

- Access to past experiences, designs, projects and data to analyze the changes as they occur in collaborative projects and where necessary provide the corresponding response to increase the awareness within the teams.
- Incorporating the "learning" and "knowledge reuse" element within moderators.
- Semi-automated knowledge creation, updation and retrieving capability.
- Capturing implicit knowledge and transforming this into explicit knowledge.
- Text mining for enhanced awareness of business opportunities between the partners in a collaborative pool.
- Consolidating the knowledge, if necessary transforming the identified patterns and /or models to alternative representations and resolving conflict or contradictions with previously extracted knowledge.

However, the application of Semantic Web technologies and tools require considerable technical expertise, and are thus not well suited for users outside the field of computer science. This makes it hard for domain experts and ontology engineers to work together on e-manufacturing tasks. One of the major challenges for the UKM research is to face the ease of interaction and operation for mass collaboration and knowledge sharing.

8

UML Modelling of Knowledge Acquisition Module

This chapter presents a method for developing and documenting the KAM of the Moderator for manual and semi-automatic update of knowledge, using the Unified Modelling Language (UML) in order to meet objective 3 of section 1.2. The KAM has been modelled to clarify and better understand the requirements and operation of this module. The use of UML to model this module has also enabled greater detail to be added to its design. UML has been employed to explore the static structure and dynamic behaviour, and describe the system analysis, system design and system development aspects of the proposed KOATING framework as discussed in the previous chapter.

Chapter Outline

8.1 Introduction

In order to develop a knowledge based system, there is a need to model the system considering its various knowledge requirements. This will provide the user, knowledge engineer and developer with guidelines to analyse, design and implement the proposed system. Therefore, the next section discusses and identifies the way to model the system and further document it. Structured knowledge based system methodologies are relatively immature in comparison with other traditional branches of knowledge engineering. These include GEMINI, POLITE and KADS [112]. Until now, the knowledge acquisition process remains one of the major challenges in knowledge engineering. A key issue is how to solicit the expert to clearly articulate their knowledge, which may be tacit or intricate. The knowledge acquisition task should focus on a representation of expertise that is natural to domain experts, so that they can think about their expertise rather than how to represent it in a computer [112]. Several authors including [112, 170-172] have suggested and recommended the use of Unified Modelling Language (UML) to model the knowledge based system for the knowledge acquisition process. In addition, UML has been recognized as a *de facto* standard for modelling software systems. Therefore, the author has adopted a UML based approach to demonstrate the design and development of the proposed KOATING framework for Moderators discussed in chapter 7. In this manner, this chapter meets objective 3 of section 1.2.

Rhem [112] has proposed the development of a knowledge acquisition workflow for a framework to assist the knowledge engineer. This workflow will guide the knowledge engineer in modelling specific domain knowledge. In the present context, the approach adopted is similar to the approach adopted by [112] to develop a knowledge management system. Once the type of knowledge, e.g., tacit, explicit, declarative, or procedural etc., that needs to be captured for the expert module has been determined, the knowledge acquisition workflow helps the user to understand the general nature of knowledge oriented tasks and guides the knowledge modelling effort using UML.

UML is a notation that can be extrapolated to include the development of knowledge models. Models are used to capture the essential features of real systems by breaking them down into more manageable parts that are easy to understand and manipulate. According to Booch [173], "A Model is a simplification of reality". Models are used in systems development activities to draw the blueprint of the system and facilitate the communication between different elements of the system and the users at different levels of abstraction. People may have different views of the system and models can help them understand these views in a unified manner. In addition, this is a widely accepted and used approach. The knowledge engineer can easily use it and successfully capture, apply, validate and retrieve the knowledge. Bearing all these factors in mind, UML has been used to construct the conceptual model of knowledge intensive activities in the proposed system. Simple visual UML diagrams have been used to understand the various elements of the system, sources of knowledge, input and output, flow of knowledge, interaction with knowledge bases, knowledge experts and users.

8.2 Unified Modelling Language

The major objectives of knowledge engineering are to support the development process of knowledge based systems by providing appropriate tools, techniques, concepts and languages to the knowledge engineers. An important approach within knowledge engineering is the use of conceptual models to represent the real world application domain and model the problem solving skills of the experts and the system. One of the major problems associated with conceptual modelling, also known as knowledge modelling is that there is no standard language available to model the knowledge for developing knowledge intensive systems. Most of the possible languages have been adopted from software engineering. Some of the widely used and famous languages in knowledge modelling are project based using a mix of notations such as Unified Modelling Language(UML), Integrated Definition Method(IDEF), Structured Analysis and Design Technique (SADT) etc. The software engineering community has adopted UML as the *de facto* standard for modelling object oriented system, and it has been advised and recommended that the knowledge engineering community should do the same [112]. This would be beneficial in the long term as KBS can be easily integrated with other enterprises systems, particularly if their design is based on a standard language as this would help facilitate communication and sharing of blue prints among developers [174]. As discussed in chapter 4, UML is rarely used for modelling automated knowledge acquisition systems; therefore the study of UML in the context of Moderators is an added contribution of this research.

8.2.1 Object Oriented Development and UML

UML is a general purpose visual language that can be used as a blueprint to develop and implement the software system. It is a standard language to specify, visualize, construct, model and document the artefacts of a software system. It provides the ability to capture the characteristics of a system using notations. The history of UML is archaeological and comprehensive. UML was pioneered by three software engineering gurus: Jim Rumbaugh, Ivar Jacobson and Grady Booch, who contributed significantly to the development of the notion of UML. They each originally had their own competing methods (OMT-Object Modelling Technique, OOSE-Object Oriented Software Engineering, and Booch respectively). Rumbaugh's OMT was strong in system analysis, but he put little emphasis on the design stage. The main emphasis of Jacobson's OOSE was on user's experience and it was lacking in terms of analysis and design. Booch's approach was strong in system design; however, he neglected the analysis aspects. Eventually, in the 1990s they joined forces and introduced an open standard to establish the basis of UML. Since then, UML has become a standard modelling language for analysts, designers and architects because it is programming language independent. Also, the UML notation set is a language and not a methodology. As such, this language can easily fit into any company's way of conducting business without requiring change [112].

UML provides (but is not limited to) several types of diagrams that when used within a given methodology, increase understanding of an application under development. The underlying premise of UML is that no one diagram can capture all the different elements of the system in their entirety. The most useful and standard UML diagrams that can be used to model a system at different points of time in the software's development lifecycle include:

1 *Use case diagram*: This illustrates a collection of functionalities performed by the system, showing which actors and processes interact with each use case. It either represents a whole use case for the complete system or the breakout of a particular group of use cases with related functionalities.

2 *Activity flow diagram*: The activity diagrams are similar to a flow chart and emphasize the procedural flow of control between two or more class objects while processing an activity. Activity diagrams are best used to model higher-level process; however, when used in a lower level system, they are associated with a use case.

3 *Sequence diagram*: A sequence diagram is also known as an interaction diagram, and shows the interactions between a set of objects and their relationships. This also includes the messages that can be passed among them in a time sequence order.

4 *Class diagram*: A class diagram is a diagram that shows the classes in the system, interfaces, collaborations, and varieties of relationships such as dependency, generalization, association, aggregation and inheritance etc.

5 *State chart diagram*: the state chart diagram models the different states that a class can be in and how that class transfers from state to state. It is used to model the event ordered behaviour of any kind of object including classes, objects, interfaces, components and nodes.

6 *Collaboration diagram*: the collaboration diagram displays the object interactions in event sequence order under relatively complicated circumstances. They convey the same information as a sequence diagram, but they focus on an object's role rather than on the times that messages are sent.

7 *Component diagram*: the component diagram shows physical components, interfaces and usage dependency between components of the system.

8 *Deployment diagram*: the deployment diagram shows how physical components of the system under development are distributed across the physical environment in which it runs.

In this manner, using all or a combination of these diagrams, UML can present the static structure and dynamic behaviour of the system. The static structure defines the kinds of objects that are important to a system, its implementation and the relationships among the objects. The static behaviour of the system can be mapped by using use case diagrams, class diagrams, component diagrams and deployment diagrams. The dynamic behaviour defines the history of the objects over time and the communication among objects to accomplish the goal. The dynamic view consists of state chart diagrams, activity diagrams, sequence diagrams and collaboration diagrams.

8.3 UML Modelling of KAM and Analysis

This section details the modelling of the KAM using UML. Here the concept of UML grows from analysis to implementation mainly focussing on the knowledge acquisition aspect of the Moderator. In the following section, the use of "system" primarily refers to the KAM. It mainly consists of three stages: system requirement and analysis, system design and system implementation.

8.3.1 System Requirement and Analysis

This phase of modelling focuses on the set of system's requirements, the available resources and the user's desire along with the idea of the system. The behaviour of the KAM (that is what functionality must be provided by the KAM) is documented in this phase using a use case model and activity diagram model. These are discussed as follows:

8.3.1.1 Use Case Analysis

A use case diagram illustrates the interaction between the users and system's functions. Here the user is referred to as an actor, which are not the part of the system, but represents anything or anyone that interacts or inputs information to or receives information from or both receives and inputs information to and from the system. Use cases represent the functionality provided by the system that is what capabilities will be provided to an actor by the system. There is no need for one use case diagram to capture everything about a system. A well structured use case diagram focuses on communicating only one aspect of the system. In the context of the Moderator, only the knowledge acquisition aspect has been considered and therefore in the proposed use case model only those use cases and actors are considered which are essential for understanding the knowledge acquisition aspects. As shown in Figure 8-1, there are three actors which interact with the system:

- User: represents the person who uses the Moderator system and possesses the domain expertise and has a basic understanding of knowledge discovery. The user verifies the knowledge and needs to be aware of the actions that might be required for a particular condition in a rule(i.e. if a particular problem occurs)

- Database: represents the past project databases, project summary data and operational data of the enterprises etc., (changes to project information may necessitate changes to the expert module.

- Expert module: is a kind of knowledge base. More details have been presented in section 7.4.

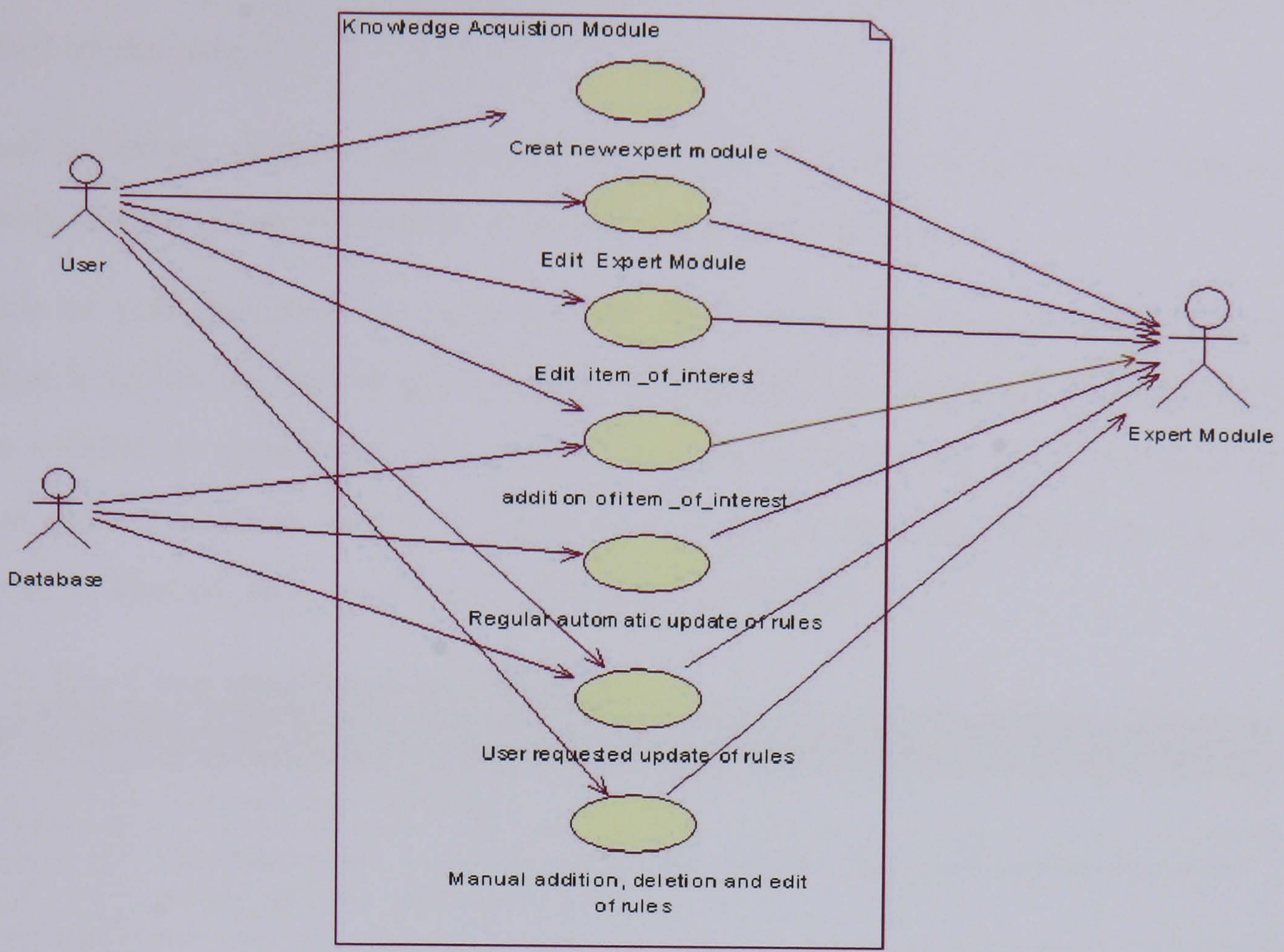


Figure 8-1: Use case diagram of Knowledge Acquisition Module

Generally, indicating the interactions between actors and the system can help to identify the use cases. Use cases mainly depend on the level of abstraction. The seven use cases that have been identified for the KAM are:

- Create expert module: This use case represents the creation of a new expert module when a new partner joins the collaborative project and agrees with the rules and regulation of the collaboration.
- Edit expert module: this includes the modification or deletion of an expert module.
- Addition of item of interest: Automatic and manual update of an item of interest. Here, the item of interest has been categorized into competencies and objects of interest related to changes.
- Edit item of interest: this can be either the modification or deletion of an item of interest.
- Regular automatic update of rules: this include the regular update of knowledge in the form of rules in the expert module corresponding to the addition of data to the databases and verification of knowledge from the user.

- User requested update of rules: this identifies a special type of knowledge which is required by the user.
- Manual addition, deletion and modification of rules: this includes the manual addition, deletion or modification of rules in the expert module.

The details of each use case are documented with a flow of events to extract more information from the use case diagram. The flow of events for a use case is a description of events needed to accomplish the required behaviour of the use case. The flow of events for each use case is written in terms of *what* the system should do, not *how* the system does it. The use case specification is detailed in Table 8-1.

Table 8-1: Use Case specifications of KAM

1. Flow of Event for the Use case *Create Expert Module.*	
Actors	User and Expert module
Pre-conditions	The collaborative team must validate the creation of an expert module after a new partner joins the collaborative project.
Post-conditions	A partially populated expert module exists. This will require addition of item of interest and regular automatic updates of rules use cases.
Basic Flow	This use case executes when a new team member join the collaboration. User enters his/her password. The system verifies that the password is valid.(E-1.1) and prompts the user to create an expert module's profile(E-1.2). The system prompts the user to enter the information required in its profile such as name, contact details and communication method etc., or quit the system.
Alternative flows	E-1.1: An invalid member and wrong password is entered. The user can re-enter its id and password or terminate the use case. E-1.2: An invalid name of expert module is entered or an expert module with that name already exists. The user can re-enter the name or retrieve existing expert module or terminate the use case.
2. Flow of event for the use case *Edit Expert Module*	
Actors	*User and Expert Module*
Pre-conditions	The expert module must exist before this use case executes.
Post-conditions	Other partners need to be aware of changes after this use case executes. This is required to make other partners aware of the current role and new area of interest.
Basic Flow	For this use case to execute, the user must be logged into the system. This use case can execute either of two main activities: delete expert module or modify expert module. User chooses the corresponding expert module (E-2.1). The system prompts the user to select the desired activity, DELETE, MODIFY or QUIT. If the activity selected is DELETE, the system will ask for confirmation to delete the expert module and if the user confirms, the system will delete the expert module (E-2.2). This will happen whenever the project is finished, or the assigned role of a partner is completed and that partner is no longer required for collaboration. If the activity selected is MODIFY, the system will help the user to modify its name or change its content in some way (E2.3). This will happen when the role of the partner has changed and it is playing a new role in collaboration. It might also happen that it has completely changed its interest and moved to a new business.

	If the activity selected is QUIT, the use case ends.
Alternative flows	E-2.1: User chooses a wrong expert module which doesn't correspond to user. The system prompts the user with invalid message. The user chooses a valid expert module or terminates. E-2.2: the system prompts the user to either permanently delete the expert module or temporarily inactivate the profile. E-2.3 An invalid name is entered, the user re-enter the name of expert module or terminate.
3. Flow of event for the use case addition of item of interest	
Actors	User, expert module and past project summary database.
Pre-conditions	Create expert module use case must execute before this use case and expert module must exist.
Post-conditions	User verifies the item of interest if it has been extracted using automated methods.
Basic Flow	For this use case to execute, the user must be logged into the system and choose the relevant expert module (E-3.1). The item of interest refers to a list of objects or attributes that are important to the user or the user's competencies. Moderator will raise the awareness based on these items of interest and competencies. When user chooses a particular expert module, system will prompt the user to either AUTOMATICALLY update competency based on summary of past project or MANUALLY add the item of interest. If the activity selected is manual addition of item of interest, system prompt the user to type the desired item/s of interest (E-3.2) and update the items of interest. Similarly, if user selects to add competency, system prompts to type competency and update the competency list (E-3.3). If the activity selected is automated update of competency, the system will prompt the user to choose the relevant database of project summaries. The competencies in the form of keywords are extracted. User verifies the extracted keywords and correspondingly prompts the system to update the list.
Alternative flows	E-3.1: User chooses wrong expert module. The system prompts the user to choose right expert module or terminate. E3.2: An invalid item of interest is entered. System flags invalid message. E3.3: Competency already exists, it prompts the user of repetition and terminates.
4. Flow of event for the use case edit item of interest	
Actors	User and expert module.
Pre-conditions	Corresponding expert module exists and addition of item of interest use case must execute before edit item of interest use case.
Post-conditions	Expert module has updated list of items of interest.
Basic Flow	For this use case to execute, the user must first choose the corresponding expert module (E4.1). This use case executes two main activities: delete item of interest and modify item of interest. User chooses the corresponding item of interest (E-4.2). The system prompts the user to select the desired activity, DELETE, MODIFY or QUIT. If the activity selected is DELETE, the system will ask for confirmation to delete the selected item of interest and if user confirms, system will delete the item of interest (E-4.3). This will happen whenever the partner feels that the selected item of interest is no longer in its operational domain. If the activity selected is MODIFY, the system will help the user to modify its item of interest or change it(E4.4). If the activity selected is QUIT, the use case ends.
Alternative flows	E-4.1: User chooses a wrong expert module which doesn't correspond to the user. The system prompts the user with an invalid message. The user chooses a valid

	expert module or terminates. E-4.2: User chooses a wrong item of interest, which s/he doesn't want to delete or modify. The user chooses a valid item of interest or quits. E-4.3: the system prompts the user to confirm the deletion of the item of interest. E-4.4 An invalid name is entered, the user re-enters the item of interest or terminates.
5. Flow of event for the use case automatic regular update of rules	
Actors	Database, user and expert module
Pre-conditions	System has access to relevant database and corresponding expert module exists
Post-conditions	User must be notified to edit, verify and resolve any conflict of rules. The edit of rules use case must execute after the execution of this use case
Basic Flow	This use case begins when new data is added to the data base. The data acquisition system passes this message to KAM. KAM has access to the database (E-5.1). It verifies the type of data and correspondingly activates the knowledge miner within the system to perform the knowledge discovery process (E-5.2). A set of rules are generated as a result of the knowledge discovery process (E-5.3). The rules are compared with the existing rules to check for a conflicting rule (E-5.4). If there is no conflicting rule update the expert module. This use case terminates when the rules have been updated. .
Alternative flows	E-5.1: KAM is unable to access the database; user must be notified of this problem. E-5.2: Unknown type of data identified, user needs to be informed. E-5.3: KAM is unable to capture knowledge in the form of rules; in which case derived knowledge should be notified to user for manual entry of knowledge in the form of rules. E-5.4: There exists a conflicting rule; user should be notified of this conflict and initiate the edit of rules use case.
6. Flow of event for the use case user requested update of rules	
Actors	User, database and expert module
Pre-conditions	System has access to relevant database and corresponding expert module exists
Post-conditions	User must be notified of the discovered rules and edit rule use case executes after this use case.
Basic Flow	This use case is initiated by the user, when s/he needs any special type of knowledge. In this case, it prompts the system to the database and specifies the kind of knowledge to be mined (E-6.1). The system identifies the data type and applies the knowledge discovery process to discover the knowledge in the form of rules (E-6.2). A comparison is made with the existing rules for any conflict (E-6.3). Rules are added to the expert module (E-6.4) This use case terminates after notifying the user of discovered knowledge.
Alternative flows	E-6.1: User selected a wrong database, it can re-select the database or quit. E-6.2: Unknown data type identified, user need to be informed. E-6.3: There exists a conflicting rule; user should be notified of this conflict and initiate the edit of rules use case. E-6.4: System is unable to capture knowledge in the form of rules; in which case derived knowledge should be notified to the user for manual entry of knowledge in the form of rules.
7. Flow of event for the use case manual addition, deletion and modification of rules	
Actors	User and expert module
Pre-conditions	Create expert module and automatic regular update of rules must have executed before this use case executes.

Post-conditions	Expert module has updated list of rules
Basic Flow	This use case begins when the new rules are automatically discovered and need verification from the user for its authenticity. Identified conflict or rules are prompted to users to resolve them (E-7.1). The user resolves the conflict by performing activity such as ADD, DELETE, MODIFY and UPDATE of rules or QUIT the system. If the activity selected is ADD RULE, system adds a new rule to expert module (E-7.2). If the activity selected is DELETE RULE, system deletes the chosen rule from the expert module (E-7.3). If the activity selected is MODIFY RULE, system facilitates the user to modify the rule based on domain knowledge (E-7.4). If the activity selected is UPDATE RULE, system updates the expert module with updated rules. If the activity selected is QUIT, the use case ends.
Alternative flows	E-7.1: No conflicting rule found, system updates the expert module E-7.2: An invalid rule is added, the user can delete or modify that rule and re-enter the desired rule or terminate the use case. E-7.3: An invalid rule is selected, the user can re-select the desired rule to delete or terminate the use case. E-.7.4: The rule chosen for modification is invalid, the user can re-select the desired rule and modify.

8.3.1.2 Activity Diagram

An activity diagram is also created to demonstrate the requirements of the system at the system analysis stage. Activity diagrams are used to model the dynamics of the system. These diagrams are essentially a flow chart that is used to show the workflow of the system; that is, they show the flow of control from activity to activity in the system. Activities can be done in parallel or along any alternative paths through the flow. It also shows how the system will respond to the user's request. In the present context, activity diagrams are created to represent a flow within a use case; this depicts the start state, activities that the system performs and decisions that affect the workflow of the proposed system. Activity diagrams contain activities, transitions between the activities, decision points, and synchronization bars. Figure 8-2 to Figure 8-8 shows the activity diagram corresponding to each use case used for the proposed system.

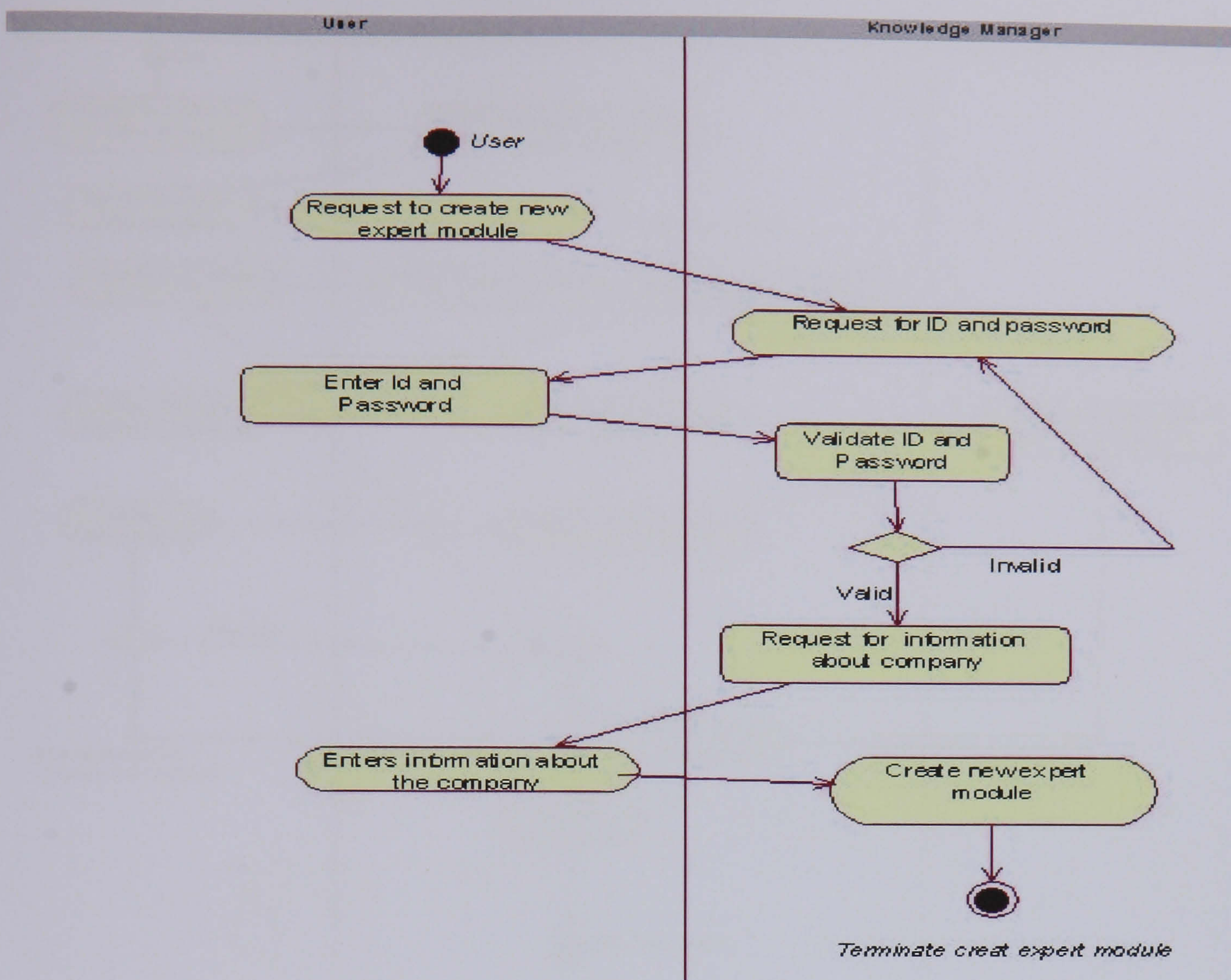


Figure 8-2: Activity diagram for use case create expert module.

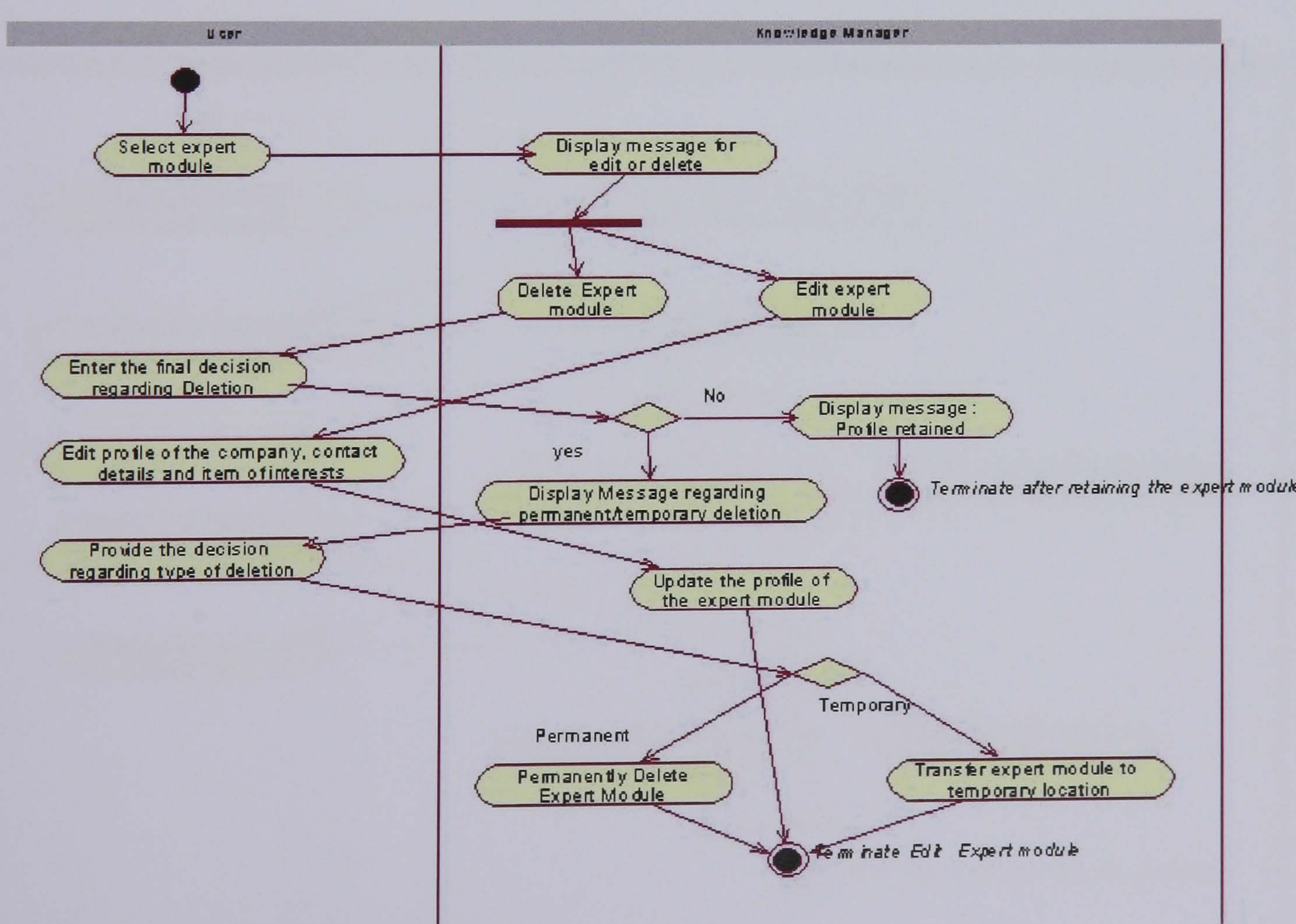


Figure 8-3: Activity diagram for use case edit expert module

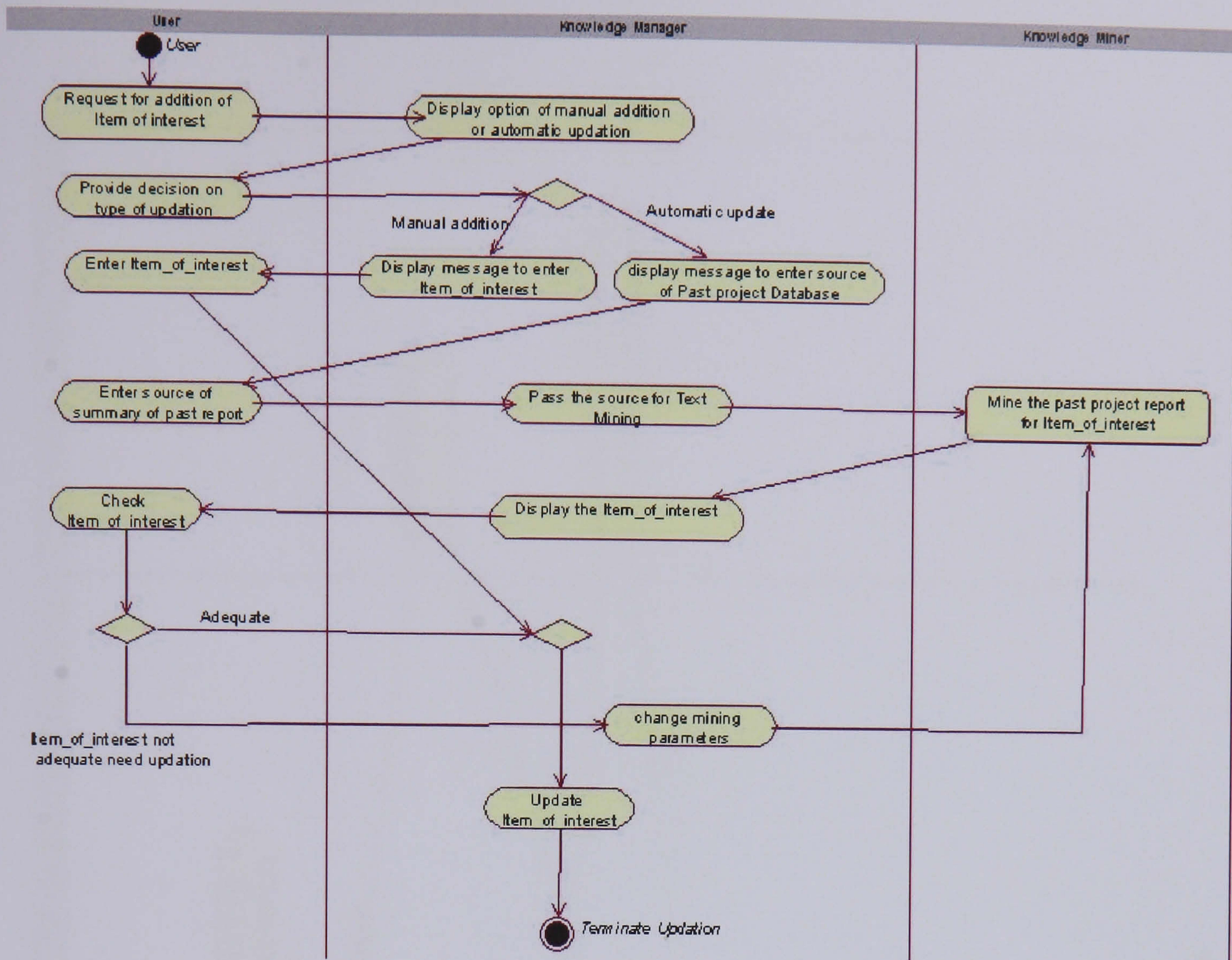


Figure 8-4: Activity diagram for use case addition of item of interest

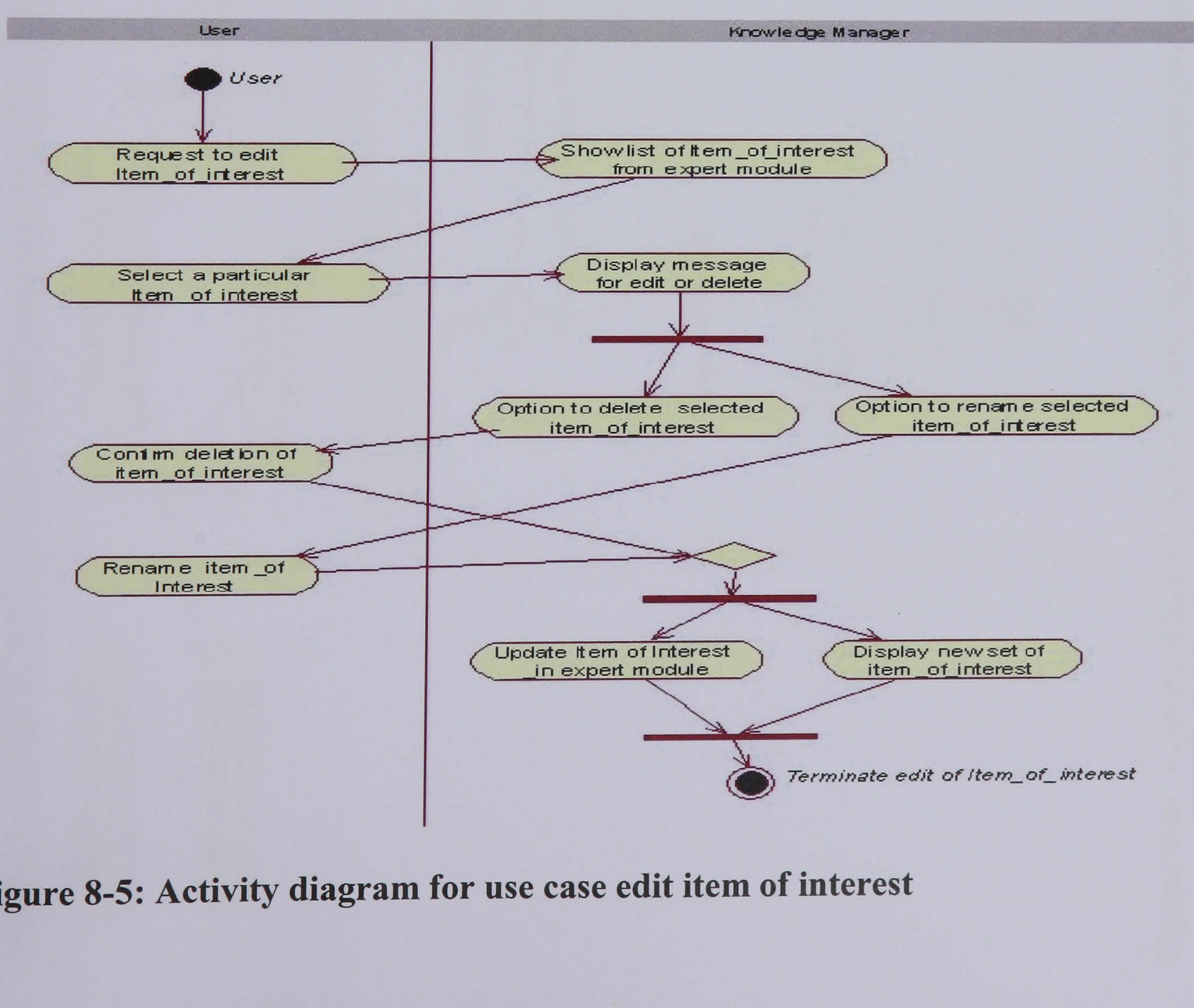


Figure 8-5: Activity diagram for use case edit item of interest

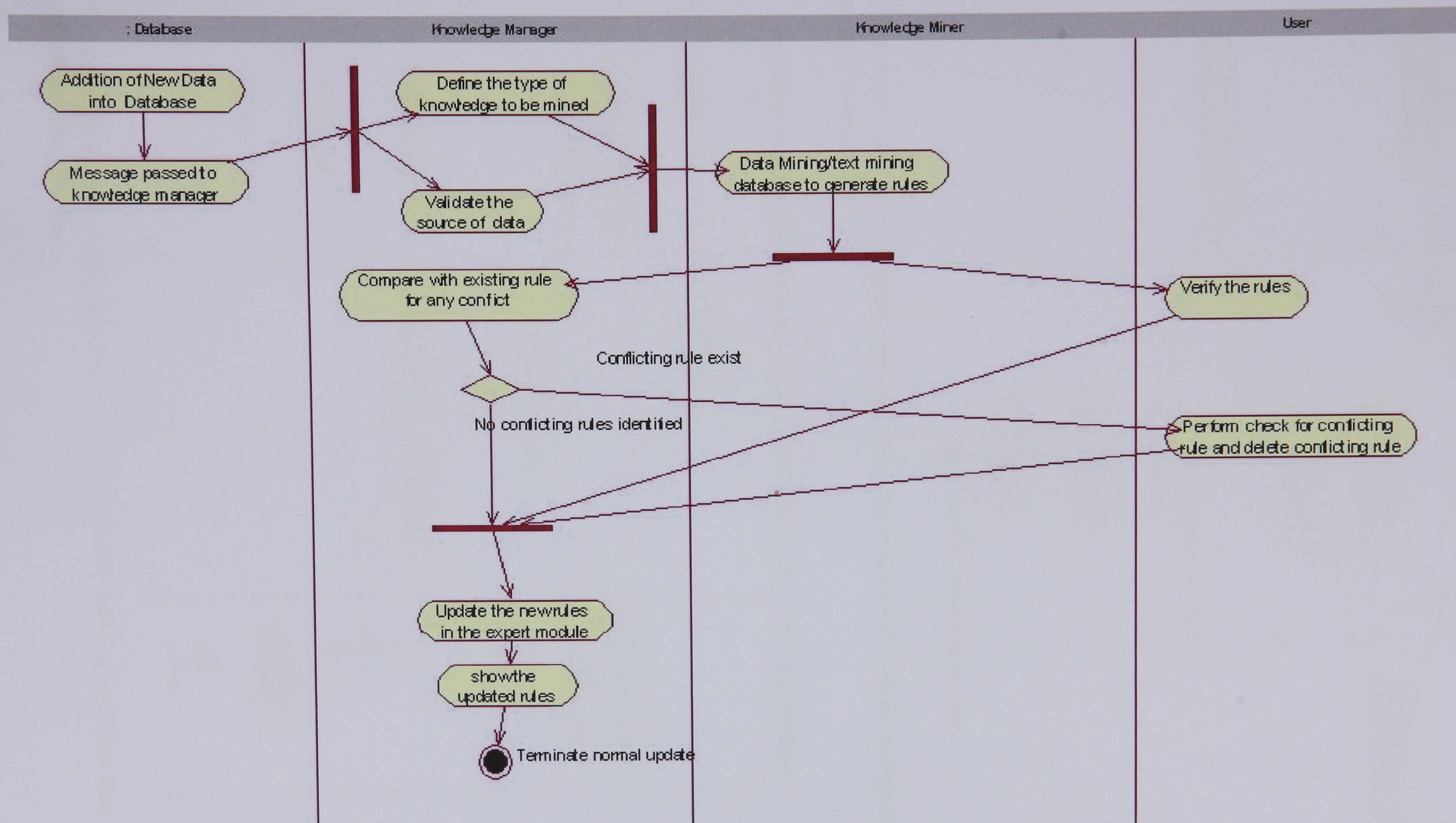


Figure 8-6: Activity diagram for use case regular automatic update of rules.

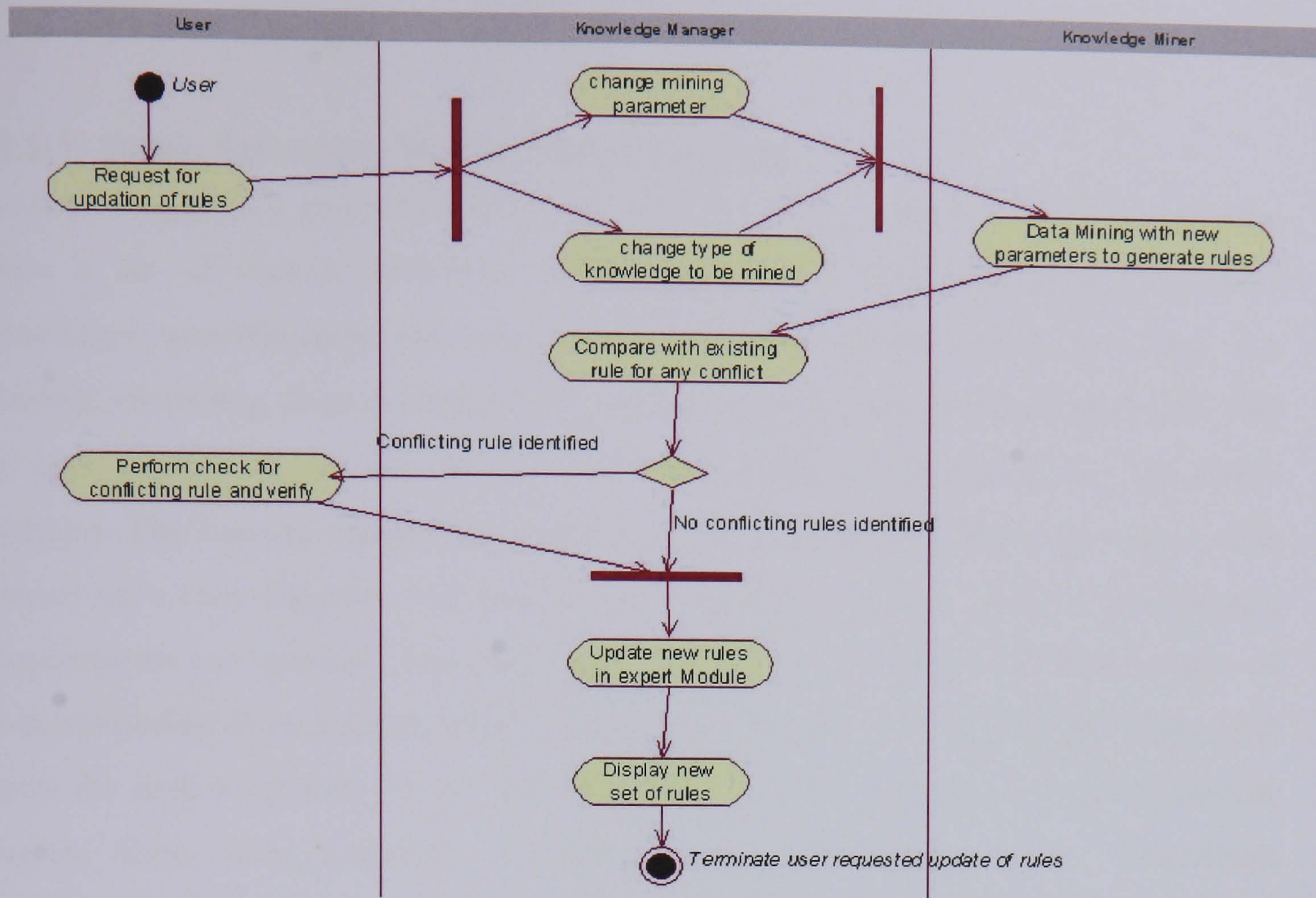


Figure 8-7: Activity diagram for use case user requested update of rules in expert module

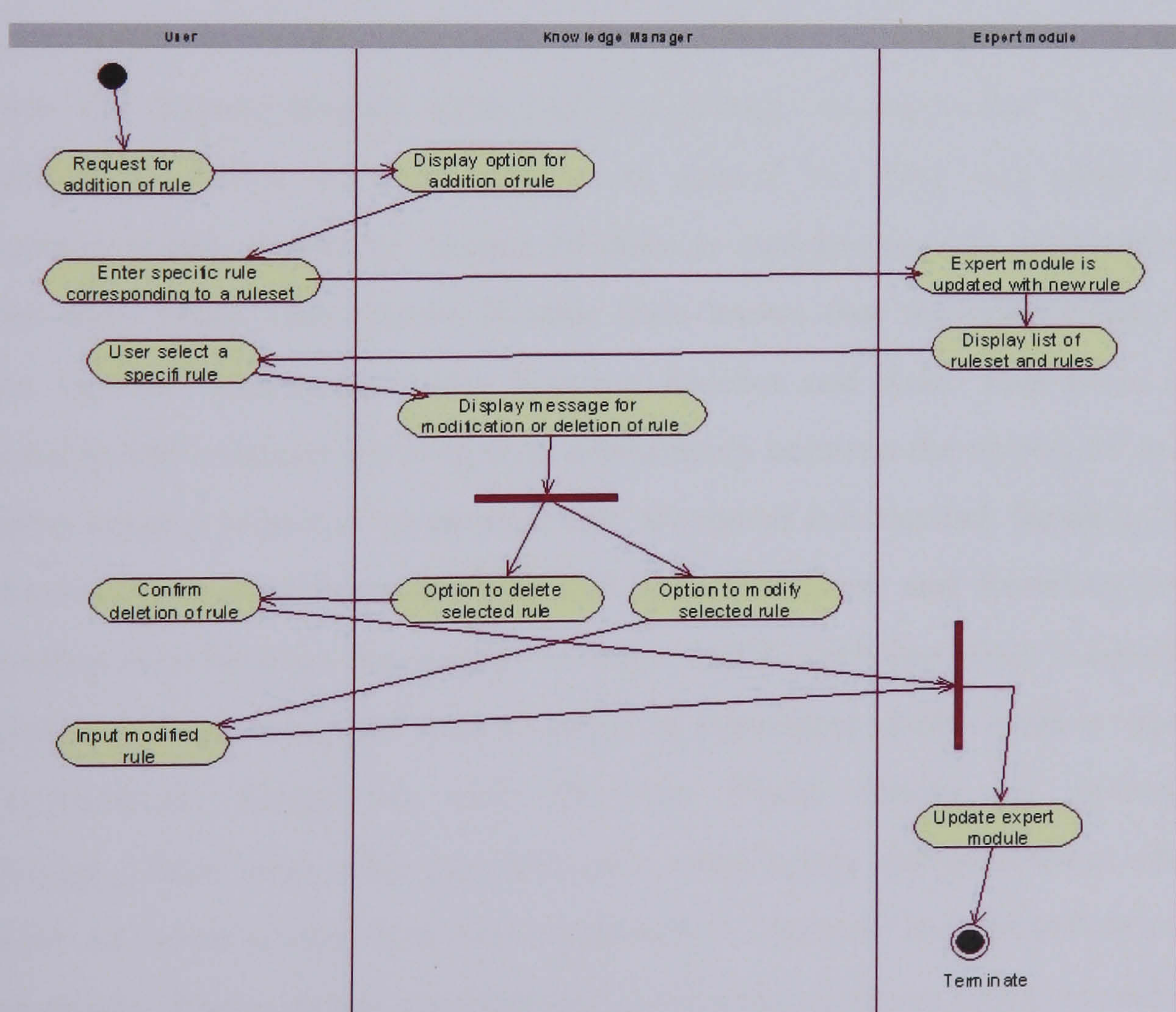


Figure 8-8: Activity diagram for use case manual addition, deletion and modification of rule.

8.3.2 System Design

8.3.2.1 Static Structure Model: Class Diagram

The class diagram is a graphical view of the static structure of the model. A class diagram shows a set of classes, interfaces, collaborations and their relationships (such as dependency, generalization, and association). The UML representation of a class is a rectangle containing three compartments stacked vertically, as shown in Figure 8-9. The top compartment shows the class's name. The middle compartment lists the class's attributes. The bottom compartment lists the class's operations. When drawing a class element on a class diagram, one must use the top compartment, and the bottom two compartments are optional. The class diagram shows that how the different entities of the system relate to each other, their internal structure and their relationships. Figure 8-9 shows the high level view of the class diagram. It mainly consists of Expert_Module, RuleSet, Rule, Rule_Condition, Resulting_Action, Knowledge Miner, Knowledge Manager and Repository. Relationships between classes are represented by lines and labels, arrowheads and notation. The upper structure in the boundarly line shows the KRM structure used in [43],. therefore, the details are beyond the scope of this thesis. The class name attributes and operations are represented in the diagram. The relationship between the Expert_Module class and Knowledge Manager class is uni-directional association. It shows that two classes are related but only one class knows that relationship exists. Similarly, Expert_Module is uni-directionally associated with the RuleSet class. Here, only Expert_Module class knows that the relationship exists. The second type of relationship exists between RuleSet and Rule. This is an association relationship and indicates the long term relationship between the classes. In other words, a RuleSet keeps a reference to another rule, whenever it is needed. Similarly, association relationships also exist between Rule and Rule_Condition and Resulting_Action. The relationship class between Knowledge Manager and Knowledge Miner is association. The knowledge miner is equipped with a variety of algorithms classes such as Apriori, C4.5, Neural_Network, Rough_Set and Stat_App. These classes are connected with Knolwedge_Miner class with generalization relationship. Generalization provides the capability to create a superclass that encapsulates structure and behaviour common to several classes. These classes are examined for commonality of structure and behaviour. Repository class is also associated with knowledge miner and Knowledge Manager with association relationship and is represented by a dotted line. In the next section, author discusses the dynamic action model for system design that is sequence diagram.

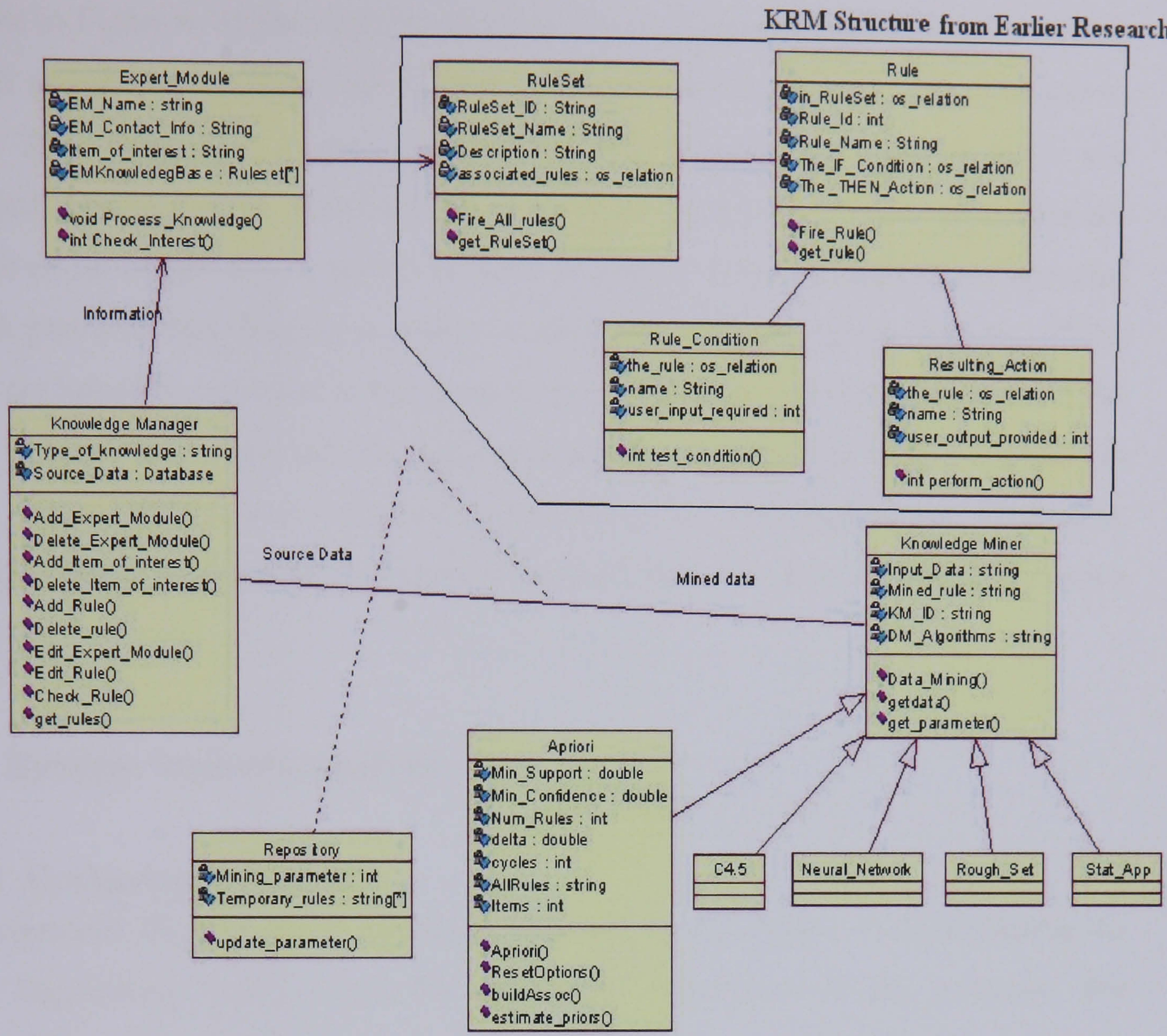


Figure 8-9: Class diagram of the proposed system.

8.3.2.2 Dynamic Action Model: Sequence Diagram

A use case diagram presents an outside view of the system. Scenarios are used to describe how use cases are realized as interactions among societies of objects. A scenario is an instance of a use case - it is one path through the flow of events for the use case. Scenarios are developed to help identify the objects, the classes and the objects interactions needed to carryout a piece of functionality specified by the use case. A sequence diagram depicts object interactions arranged in time sequence. It shows the objects and classes involved in the scenarios and the sequence of message exchanged between the objects. Sequence diagram are typically associated with use case realizations in the logical view of the system under development. In the sequence diagram, an object can be named in three ways: the object name, the object name and the class, or just the class name. In present context, only one diagram capturing the features for design of KAM has been considered.

As shown in Figure 8-10, the sequence diagram shows objects arranged along the X-axis from left to right and message ordered in a timely manner along the Y axis from top to bottom. They commonly contain objects, links and messages, along with notes and constraints. It also has an object life line, a vertical dashed line, which represents the existence of an object over a period of time. In Figure 8-10, the objects are extracted from the previous class diagram as shown horizontally at the top of the diagram. All the objects are linked by association instances as per their sequence of participation in the process. The objects send call messages and invoke a method to start an action. In present context, the sequence of process started with the user logging into the system, creating its expert module and ending either deleting expert module or editing the expert module.

8.3.3 System Implementation

8.3.3.1 Component Diagram

The component diagram shows the static structure of the system. It demonstrates the various dependency types among the executable component of the program. The component to which arrow is leaves is said to be dependent on the component to which it goes. The dependency is shown by a clashed arrow between components. Stereotypes may be used to define the type of dependency. In present context, as shown in Figure 8-11, the system is divided into 5 major components. They are Interface, AppsPackage, DataMiningPackage, repository.db and KBSSupp.db.

- Interface: this component of the system provides the user a mean to communicate with the system. It facilitates the user with many capabilities by performing various tasks, visually showing the rules, item of interests and profiles etc.
- AppsPackage: is a package that offers many built in functions for common usage of the proposed system based on the user's request or message received from other components of the system. It obtains results from the DataMiningPackage and verifies them by the user using interface and deciding if they should be stored in the KBSSupp.db.

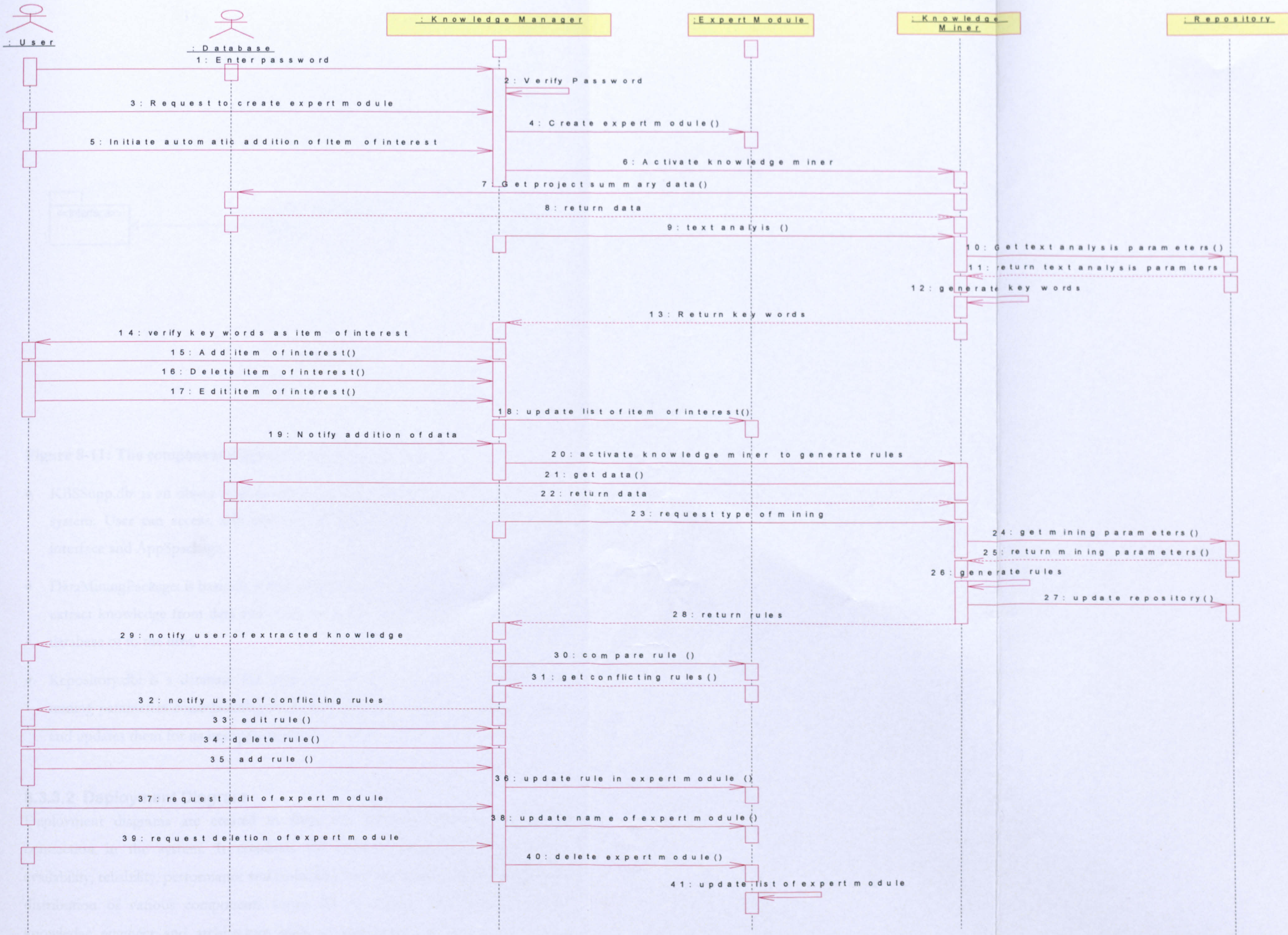


Figure 8-10: The Sequence Diagram of KAM

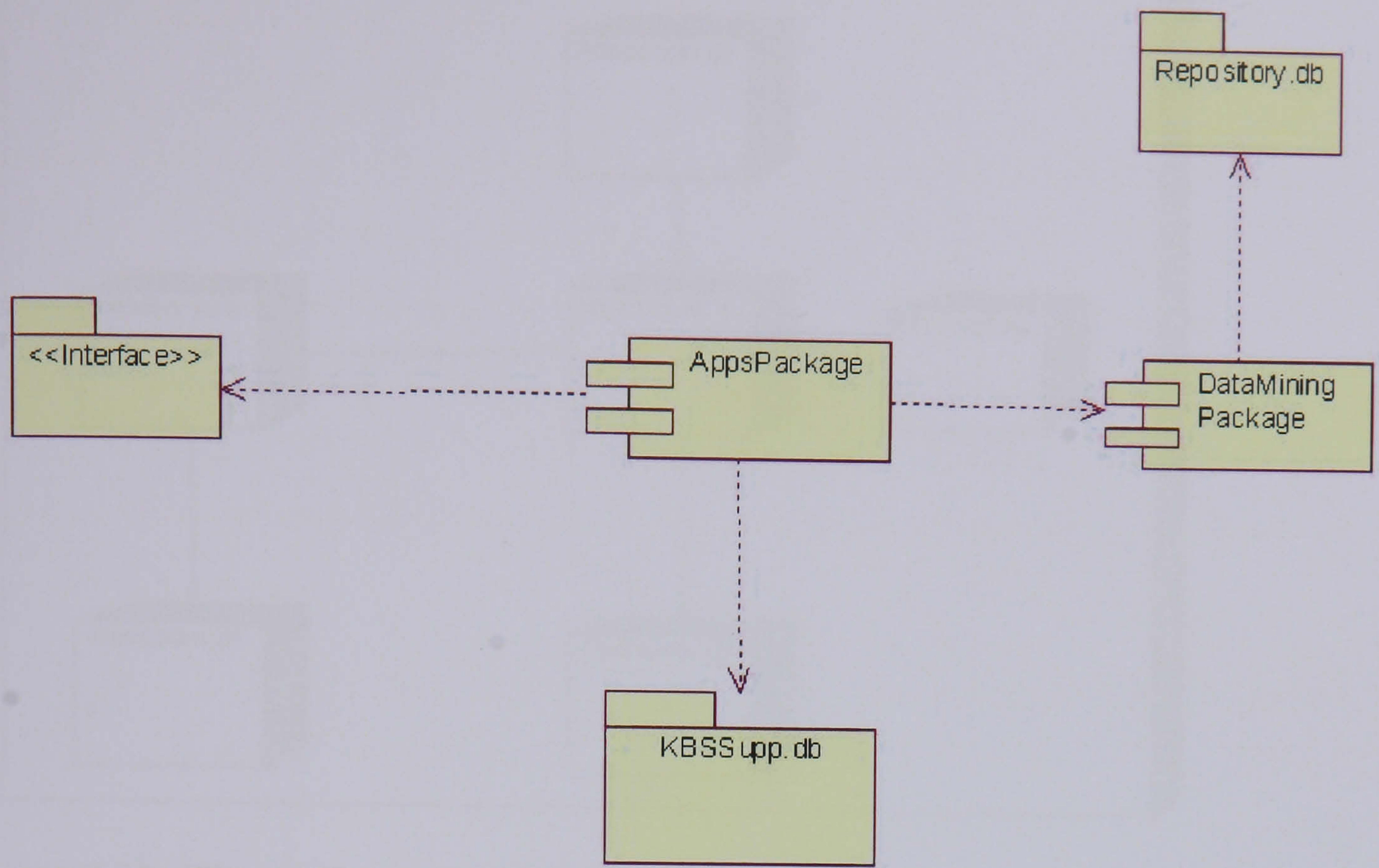


Figure 8-11: The component diagram of the proposed system.

- KBSSupp.db: is an object oriented database storing rules and other attributes of the system. User can access, edit and add information in to this database using the interface and AppSpackage.
- DataMiningPackage: is basically a data mining engine containing several algorithms to extract knowledge from data and deliver it to the AppSpackage for update into the database or to the user.
- Repository.db: is a database for temporary storage of rules. It stores the various mining parameters corresponding to several algorithms used by DataMiningPackage and updates them for its best performance.

8.3.3.2 Deployment Diagram

Deployment diagrams are created to show the different nodes along with their connection in the system. It considers the system requirements such as system availability, reliability, performance and scalability. The deployment diagram visualizes the distribution of various components across the enterprise. This diagram allows the knowledge engineer and architecture team to understand the system topology and provides aid in mapping components to executable processes.

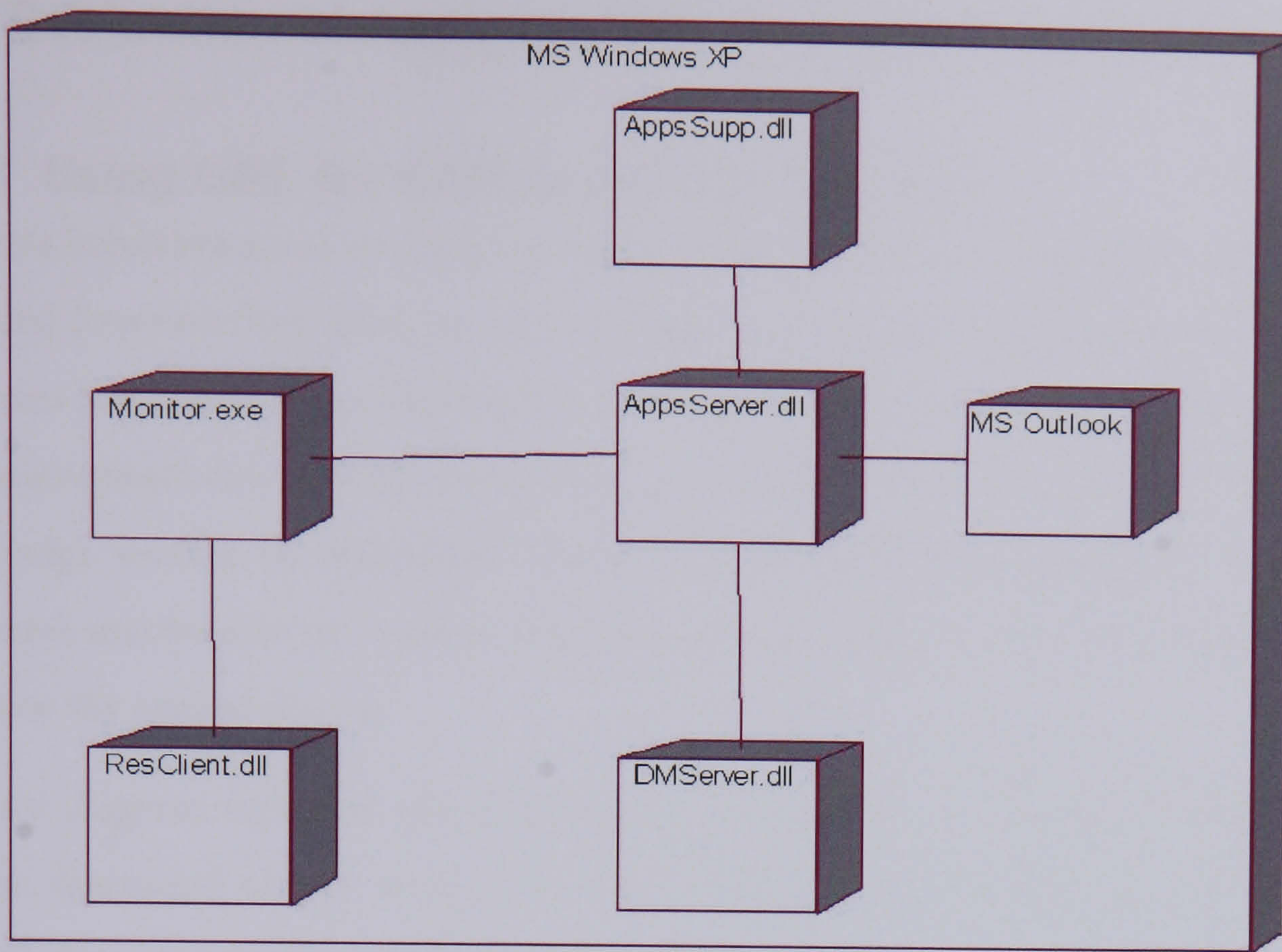


Figure 8-12: The deployment diagram of proposed system.

As shown in Figure 8-12, the deployment diagram shows what runs where. Only Monitor.exe, ResClient.dll, AppsSupp.dll, AppsServer.dll, DMserver.dll and MSOutlook are needed in the user PC. The design time component, such as KbsSupp, AppsPackage, DataMiningPackage and KbsSup.db are not necessary. In other word, the user does not need to install the expert system tools packages and copy the related program source/database to the target PC during deployment. Monitor.exe is the main executable program to execute the program. During execution, it obtains the business rule logic and interface information from the exported package called AppsServer.dll and AppsSupp.dll. The user only needs to copy the above files into a dedicated directory and create a shortcut for use. Moreover, email is needed to communicate between partners. Thus, MS Outlook should be available on the users PC to enable the aforementioned service. This program will automatically call MS Outlook. Although the email is represented in the diagram by MS outlook but any preferred mail service could be used. In this manner, one can see that the main difference between the component diagram and the deployment diagram. The former do not show instances of components, merely the dependencies which apply between all components of one type and all components of another. Instances are shown in the later. The following subsection evaluates the use of UML for the proposed system development.

8.4 Evaluation of Applying UML for KAM development

8.4.1 Using UML for KAM knowledge Analysis

System's behaviours can be captured using a use case model which illustrates the system's intended function (use case), its surroundings (Actors) and the relationships between the use cases and actors (use case diagrams). The most important role of a use case model is to communicate the system's functionality and behaviour to the end user. This helps the knowledge worker to understand the role of different actors and their corresponding processes involved in the system. Here the main emphasis is what the system should do, not how the system does it.

Activity diagram represent the dynamics of the system. Technically, workflow analysis can be illustrated clearly using the activity diagram. At this point in the life cycle, an activity diagram may be created to represent the flow across the use cases or to represent a flow within a use case. Later in the life cycle, activity diagrams may be created to show the workflow for an operation.

Without these diagrams, the knowledge engineer cannot understand the scope of development of the system. It helps the knowledge engineer to correct any misunderstanding at the beginning of the system development.

8.4.2 Using UML for KAM Knowledge design

A class diagram can show the relationships between various classes accurately and thus accelerate the development process at a later stage. However, it is very tedious task to present all the possible choices in a class diagram. A sequence diagram shows object interactions arranged in a time sequence. This diagram represents the graphical view of the scenario. In conclusion, without these two diagrams, the knowledge engineer may not be able to realize the knowledge structure, representation, and interactions before programming and implementation.

8.4.3 Using UML for KAM knowledge implementation

The component diagram and the deployment diagram are used in this research for knowledge implementation purposes. These diagrams can be very useful to transfer the necessary knowledge to the developers and IT implementers. The component diagram can illustrate the static structure of various components used for editing the source code. It is beneficial for a knowledge engineer to maintain and update the inference engine and the knowledge base. A deployment diagram is created to show the hardware configuration that is used for the system under development. It provides instructions on how to install the system onto a users' PC. Since the Moderator system will be distributed over several partner's desktop computers, the run time configuration should be illustrated clearly to reduce the installation and version upgrading time frame.

8.5 Conclusion and learning outcome.

In conclusion, this chapter has articulated a way of modelling the KOATING framework for Moderators using UML to meet objective 3 in section 1.2. UML has been recommended by several authors as mentioned in section 8.1 and also recognized as a standard for modelling knowledge based systems and software applications. UML has been used to analyze, design and develop the proposed KOATING framework using a set of diagrams. It can also capture the static and dynamic models of the system. Use case diagram and Activity diagrams were used to analyze the knowledge requirements of the proposed system. A class diagram and sequence diagram were used to design the system, and component and deployment diagrams were used to provide guidelines for the implementation of the system. In this manner, the author has provided guidelines on how the proposed system can be designed, developed and implemented.

The application of UML to develop the proposed KOATING framework provided the author with a better understanding of the system under development. However, this experience can be extended in a number of ways such as designing software, communicating software or business processes, capturing details about a system for requirement or analysis and documenting an existing system, process, or organization. In the present context, it has been found that the UML attempts to bridge the gap between the original idea for a system under design and its development. Furthermore, a set of codes can be generated from the developed diagrams that will ease the process of software development. Being proficient in UML modelling means that one will have an ability to capture ideas, relationships, decisions and requirements in a well defined notation that can be applied to many different domains.

9

Extracting Knowledge from Post Project Reports

This chapter discusses the application of text mining on PPRs with a view to extracting useful knowledge that can be used to update the expert modules. It first presents an overview of the problem context by briefly describing construction project supply chains, PPRs, issues involved with PPRs and their limitations. It further discusses the construction project moderator and the type of knowledge needed by the expert modules. A variety of text mining techniques have been applied to discover useful knowledge that can be presented to the user for addition and updating of expert modules. Finally, it discusses the limitations of the proposed approach.

Chapter Outline

9.1 Introduction

This chapter provides a case study related to a construction project supply chain in order to prove the concept of the proposed KOATING framework for Moderators using a proof by demonstration methodology. It illustrates how different kinds of knowledge can be captured by knowledge miners from unstructured text based data in the form of post project review (PPR) reports, using a variety of TM techniques. The discovered knowledge can be used to semi-automatically update the knowledge content of the Moderator's EM.

This case study shows how different TM algorithms can be used to extract knowledge as patterns, associations and trends from the PPRs. The generated knowledge in the form of rules relates to improving the processes, identifying recurring problems or good practices, addressing problem areas, improving customer relationships and enhancing the coordination between collaborators of the construction project. The details of various data mining and text mining approaches have already been discussed in chapter 5, and this chapter demonstrates their application. It also provides guidelines for the application of knowledge miners on the post project reports data related to CPSCs. Knowledge miners are used to identify hidden knowledge and then embed this learning in the Moderator's EMs. Availability of this knowledge in the EMs will enable the Moderator to identify potential problems or item of interest due to operational changes and make the team members aware of these and recommend appropriate action.

This chapter presents an approach used by the knowledge miners to extract knowledge from unstructured PPR based textual data. The extracted knowledge is not available in the form of IF-THEN rules and therefore requires interpretation so that the identified knowledge can be manually inserted into the expert modules. The next section presents a brief overview of CPSCs, PPRs and the current limitations and issues involved with PPRs to provide a context for the case study.

9.2 Overview of PPRs in Construction Projects.

The construction industry is one of the most diverse and unstable sectors within the UK economy. It faces widely fluctuating demand cycles, project specific product demands and uncertain production conditions, and has to combine a diverse range of specialist skills within geographically dispersed short term project environments. A CPSC may contain hundreds of firms, contractors, subcontractors, material and equipment suppliers, engineering and design teams and consulting firms [175]. Collaboration between the various entities of the CPSC is temporary and may vary from project to project. The lifecycle of a CPSC is limited to a particular project. PPRs of construction projects are one of the most important and common approaches for the capture of knowledge and lessons learned from the operation of a CPSC. They provide opportunities for the project team members to share, discuss and explain their experiences through face-to-face, facilitated interactions before a project is closed and the team is dissolved. PPRs therefore allow multi-disciplinary teams to critique a project to determine both positive and negative aspects, potentially capturing tacit knowledge as learning points to improve the planning, execution and design of new construction projects.

9.2.1 Construction project supply chain (CPSC)

A CPSC involves all the processes, activities, tasks and information flows (both upstream and downstream) within various networks of organizations involved in the delivery of quality construction projects or services. A general CPSC may contain several firms, contractors, sub contractors, material and equipment suppliers, engineering and design firms or teams, consulting firms or teams, etc. It remains highly fragmented and involves many small and medium size suppliers and sub contractors. The upstream responsibilities of the CPSC consist of the activities and tasks leading to the preparation of the production on site involving the construction client and design team. The downstream activities and tasks involve the construction suppliers, subcontractors and specialist contractors. The CPSC therefore needs a high level of coordination among various stakeholders, who may have conflicting interests. This thesis only considers the physical view and therefore does not take into account the management philosophy of supply chains. CPSC is not the focus of this thesis but a more detailed study of CPSC is presented in [175]. The linkage between a CPSC and PPRs is shown in Figure 9-1.

9.2.2 PPRs of Construction Project

PPRs are a rich source of data and information for organisations - if organisations have the time and resources to analyse them. Too often these reports are stored, unread by many who could benefit from them. PPRs attempt to document the project experience – both good and bad. If these reports were analysed collectively, they may expose important detail and experiences which have perhaps been repeated across a number of projects. However, because most companies do not have the resources to thoroughly examine these PPRs, either individually or collectively, important insights are missed thereby leading to missed opportunities to learn from previous projects. This research attempts to determine whether hidden knowledge and experiences could be captured using KDT and TM approaches to uncover patterns, associations, and trends in unstructured data. The results might then be used to identify specific problem areas and enhance processes, and improve customer relationships.

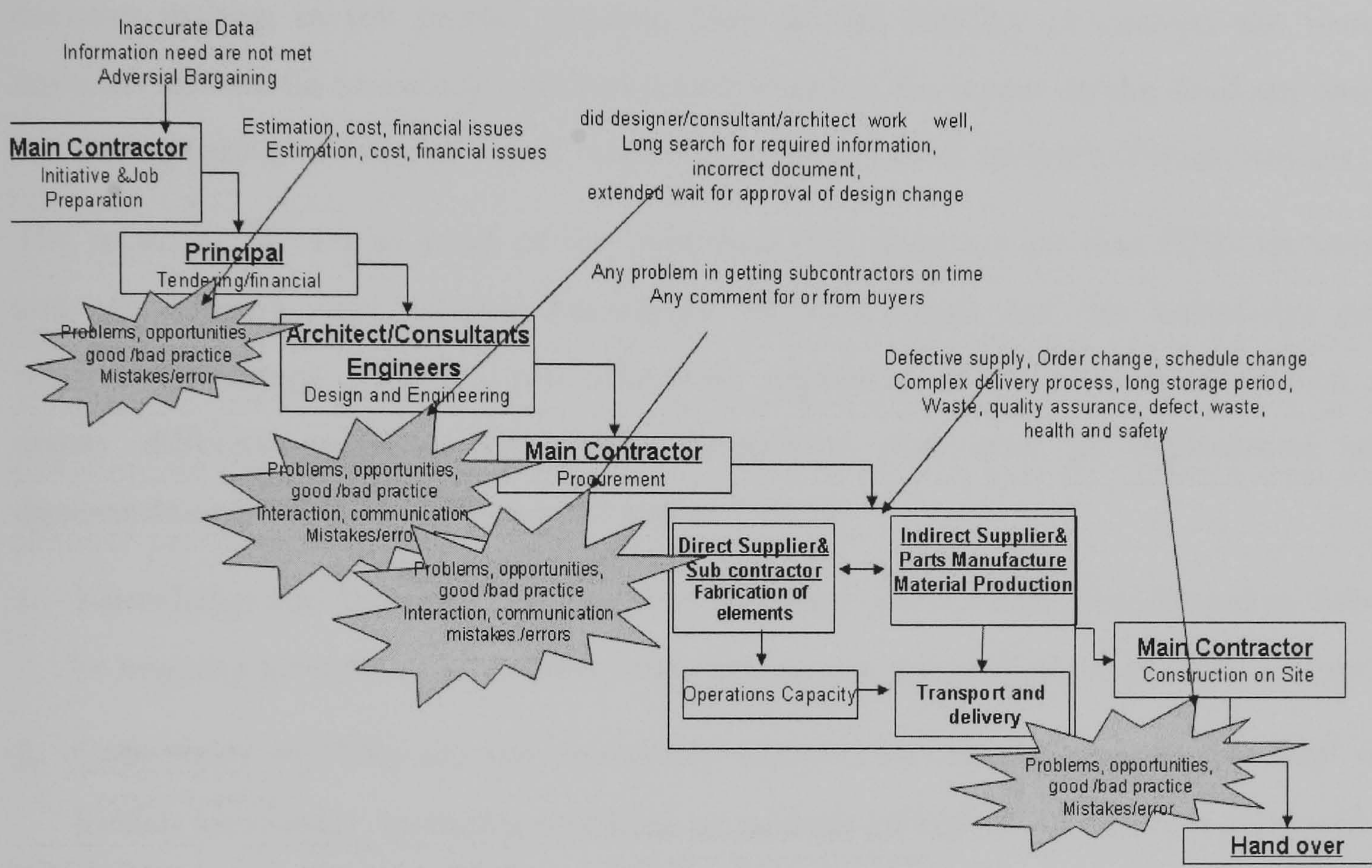


Figure 9-1: Generic view of construction project supply chain and linkage with PPR.

A PPR is a process through which an organisation looks at the project process retrospectively with a view to learning from activities carried out, to avoid mistakes in the future and also to learn from successes and failures. A PPR is also defined as a formal review of the project which examines the lessons which may be learnt and used to the benefit of future projects. A detailed review of PPRs approaches, benefits of conducting PPRs and problems associated with PPRs is presented in [176]. It is clear from this review that many authors acknowledge the benefits of conducting PPRs. These include facilitating collective learning, providing utilizable knowledge, benefitting client organization, better project phase management and preventing knowledge loss. However, several recurring problem areas have also be identified as summarized in the next sub-section.

9.2.3 Issues involved with PPRs

As reviewed in [176], several authors have indicated that organisations are aware of the benefits of PPR but do not utilise the full opportunities to learn from these. Newell [177] explored reasons why organisations capture project knowledge but do not seem to subsequently utilise that knowledge in future projects. One of the conclusions of Newell's research is that project teams lack awareness that such critical knowledge exists and could improve their processes. The PPR is sometimes a huge silo of information which rarely gets analysed critically to reveal patterns of information that could help decision making in the project process. Due to this inability to convert the review contents into useful knowledge, project teams abandon the report on the shelf and move on, thus ignoring knowledge which might be useful and even critical to future projects.

The recurring themes in much of the published PPR literature are that PPRs are useful and can generate very valuable knowledge, of many types, but this knowledge (and consequent lessons learnt) is not effectively exploited throughout organisation(s) or across different projects. The major limitations that exist in exploitation and dissemination of PPRs can be summarised as follows:

1. Knowledge and learning from previous PPRs are not routinely transferred to future or ongoing projects and therefore learnt lessons are not properly exploited or reused.
2. Collections of PPRs are not commonly analyzed or explored together to find the hidden knowledge, recurring problems or patterns of behaviour that may exist within them.
3. PPRs may identify good and bad practices but existing PPR processes do not detail how these should be disseminated in order to improve project processes or performance.
4. PPRs commonly generate large quantities of documentation which may be stored on a company's information network. However research is still required to determine how to effectively disseminate analyzed PPRs throughout an organization.

Drawing from the above, a major challenge for organisations is how to extract knowledge from PPRs and disseminate this appropriately across the organisation to ensure optimum improvements in the organisation's project processes. The application of KDT and TM methods on PPRs should provide opportunities for better exploitation of the potential benefits of these PPRs by extracting useful information to identify and replace poor practices, improve process performance, avoid reinventing solutions, re-use lessons learned on previous projects etc. The use of KDT and TM techniques would also facilitate the detailed analysis of multiple PPRs simultaneously and therefore could substantially increase the potential for useful information and patterns of operation being found.

In the remainder of this chapter, a knowledge miner of a Moderator is implemented on the PPR based database as part of a knowledge acquisition process. This is done to determine how explicit knowledge can be extracted from the PPR based database, so that it can be used to (1) equip the expert module with knowledge which was previously only available to the project team members by reading the PPRs and (2) update the content of an expert module as soon as a project finishes. The ultimate aim is to show how a knowledge miner can be used to extract useful knowledge.

9.3 Construction Project Moderator

A construction project is a type of multi-disciplinary project which requires the involvement and application of different areas of expertise including a main contractor, design and engineering, consultants, suppliers, sub-contractors, quality assurance, health and safety and financing etc. In addition, the main contracting company may have different types of projects going on simultaneously at different sites involving several stake holders. Inevitably, during the design and operation of a CPSC, problems will arise due to the different objectives of the team members and this may lead to conflicting decisions, which may be beneficial to the objective of one function, but detrimental to the objective of another function. In such situations, an intelligent software tool such as construction project moderator might be used as a special manager to raise the awareness of potential problem areas or decisions that might affect the team members during the design and operation of CPSC. The purpose of the construction project moderator is to continuously monitor the decisions related to the project activities, to identify as far as possible each potential occurrence of conflict or problems, and to orchestrate a dialogue between the interested team members until the conflict is resolved.

The basic functionalities, structure and contents of a construction project moderator (CPM) are the same as the earlier moderators described in chapter 3 and chapter 7. The methodology for the design and development of the Moderator has been discussed in chapter 8. Although the focus of the research in this thesis is on the knowledge acquisition element of the Moderator, it is useful here to briefly review the types of knowledge used by a CPM as it supports a collaborative team during the lifecycle of a project.

A basic assumption of Moderator technology is that the project team members share and add to the same shared knowledge relating to the project and/or the deliverables of the project during the lifecycle of the project. This shared knowledge may be distributed or located in different places. The schematic representation of this concept is shown in Figure 9-2.

During the project, the CPM monitors the shared knowledge and each time a change is made to it (e.g. addition, deletion or edit), the CPM checks its internal knowledge to see if the "change" might be important to one or more of the team member and if it is, what action should be carried out to make that team member aware of the change (and potential problem). To carry out these functions, the CPM's internal knowledge must contain:

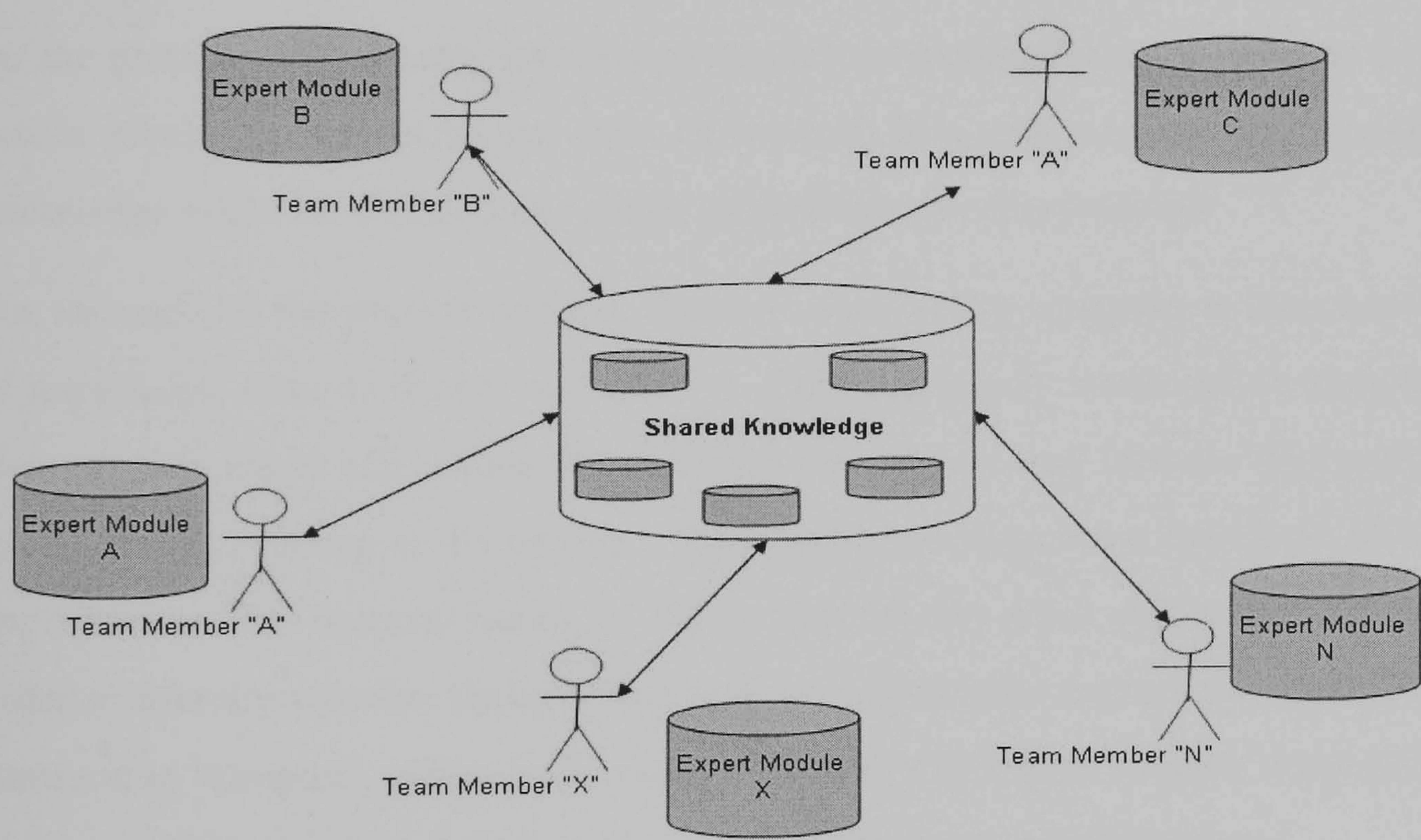


Figure 9-2: Shared knowledge of the project

- Knowledge of how to carryout the appropriate checks on the shared information to identify important changes.
- Knowledge of what is important to each team member.
- Knowledge of how to carry out the appropriate actions when an important change has been identified.

Part of the Moderator's internal knowledge is stored in expert modules, and there is one expert module for each member of the project team as shown in Figure 9-2. The knowledge in each expert module is structured into several parts e.g.

- Identification and contact info for the team member represented by the expert module.
- Items of interest.
- Knowledge of what to do when an item of interest is identified.

A detailed description of expert module is presented in chapters 3 and 7. The CPM needs to be populated with knowledge about all the team members and things that are important to each of them. The expert modules also contain information about areas of interest for each team member that is an expert module contains a list of items in which the corresponding team member is interested and also knowledge about what the CPM should do if one of these "items of interests" is changed. Therefore, in any project the Moderator's internal knowledge will be distributed in several expert modules. However for the process of discussing knowledge acquisition (and updating of items of interest, or details relating to any particular item of interest) it is only necessary to consider how knowledge might be changed in a single (or individual) expert module.

For instance, in the present context, the main contracting company is interested in a list of knowledge areas as shown in Table 9-1. They are mainly interested in identifying any changes that might affect these knowledge areas as this may indicate that problems or conflicts are occurring in the project. These key knowledge areas have been identified as important to the success/failure of the project by the main contracting company. In order to identify relevant changes and perform appropriate actions to support the main contracting company, when such changes occur, the expert module representing the main contractor must contain knowledge related to these knowledge areas.

As discussed earlier, the main aim of this chapter is to show how the knowledge miners of the knowledge acquisition module can be used to equip and semi-automatically update the content of an expert module by discovering new knowledge related to these key knowledge areas from the post project reports. The next section discusses the application of knowledge miners to extract useful knowledge.

Table 9-1: Key knowledge areas of the contracting company

1	*Financial Issues* Additional costs Budget Contract sum Financial reporting Liquidated damages Pricing Profit in excess of regional targets Tender allowance	2	*Time* Contract programme Extension of time Lead-in times Practical completion Project hand over
3	*Safety* Access Accidents Health and Safety	4	*Environmental Issues* Transporting materials Waste disposal Whole life performance
5	*Quality* Aftercare Internal audit Key Performance Indicators Process Product Quality Scope of work Service	6	*Individual Trade Packages* Drainage Drilling Electrics Excavation Plumbing Scaffolding Security Procurement
7	*How work won* Design and build Discounted tender Negotiation Tender		

9.4 Proposed KOATING framework for Construction Project Moderators

Figure 9-3 shows the proposed KOATING framework in the context CPM to extract knowledge from PPRs based textual data. It shows that PPR based database has been used as a source of knowledge. In the present context for experimental purposes, PolyAnalyst 5.0 software system has been used in this research as an instance of a knowledge miner which performs as a data mining engine. The benefits provided by this software and its comparative study with other softwares are presented in Table 5-3. PolyAnalyst has access to PPR database. As an instance of knowledge miner, PolyAnalyst applies several techniques such as text analysis, link analysis, text OLAP, semantic search, summarization and summary statistics with a view to discovery variety of useful knowledge from PPR database. A brief overview of these techniques is presented in section 5.4.1, however, a detailed study is presented in [146]. The application of these techniques is discussed in the later subsection. Once the knowledge is discovered by PolyAnalyst, it is presented to the user for verification and knowledge can be updated in the appropriate expert modules using the knowledge acquisition module interface.

As discussed in the sections 5.2.2, the whole process of knowledge discovery works in several stages including understanding the problem domain and process, identification of data source and its type, data cleaning, data transformation, data selection, data mining, pattern evaluation and knowledge representation. However, these steps can be iterative at any of these stages. As stated in chapter 7, a knowledge miner is equipped with several modules to perform these functionalities at each stage of knowledge discovery as required. These modules consist of several heuristics and algorithms to perform the knowledge discovery process and extract different kinds of knowledge. The algorithms and techniques presented in this case study are only examples to show the application of the knowledge miner, since no single technique can perform best on all types of data. A data mining module of the knowledge miner should be able to apply a variety of different data mining algorithms depending on their appropriateness for the data that is being analyzed.

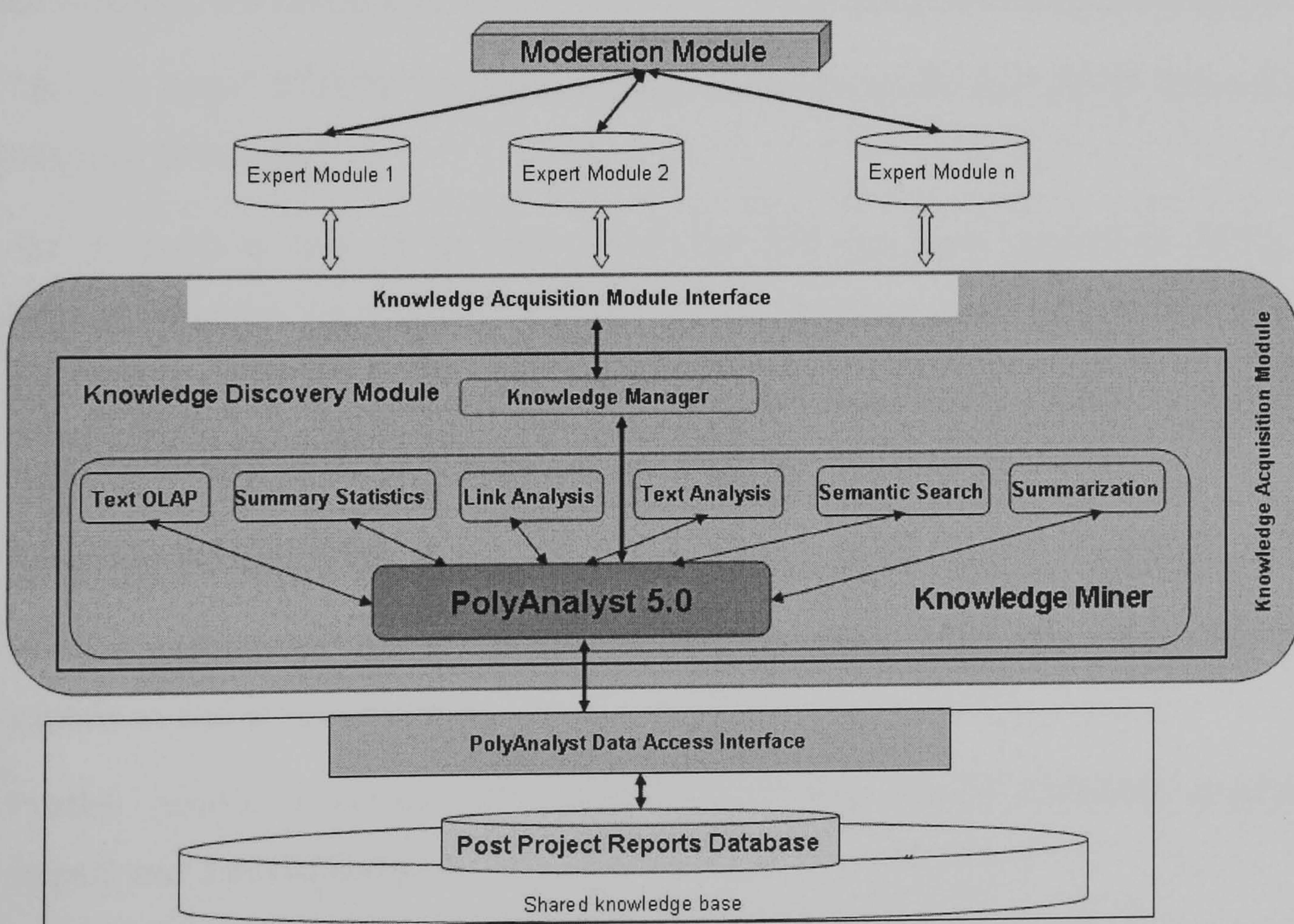


Figure 9-3: Proposed KOATING framework in context of Construction project Moderator

9.4.1 Domain understanding and collection of documents

This example is based on PPR documentation of two UK based construction companies. 40 PPR documents were collected from work projects carried out during the last three years. Additional information was also gathered from the companies, including PPR agendas, explanations of the process, structure of PPRs, and the criteria and expectations of the companies when they carry out PPR. There are two quite different types of reports. The first containing textual narration and a description of the project, with the review divided under sixteen headings plus textual information providing the lessons learned during the project. These reports were quite long and typically vary from 15- 25 pages in length. The second type of reports does not have any specific headings and used pictorial notations to express the views of its members. Every PPR includes key people who have detailed knowledge of the project and the procedures and format of review are uniform across all the projects. Analysis has been carried out separately for both the companies with an aim to discover useful knowledge for future project management, process development, helping decision makers determine appropriate decisions, avoiding mistakes, optimizing the process and identifying any patterns of good or bad practices. Once the key knowledge are discovered, they can be presented to the user for update of expert modules. An example of the key knowledge in the context of this case study is:

IF *"change in design"* ***THEN*** *"alert the contracting company for possible loss"* ***AND*** *"negotiate with designing team for redesign".*

As this research is first of its kind where the TM has been applied to PPRs, its exploratory in nature. An iterative methodology was therefore designed for this research.

1. Manual inspection of PPR reports and discussions with domain experts to determine types and examples of knowledge which should be found within the reports.
2. TM using the same PPR reports
3. Manual examination and evaluation of the TM results (with reference to domain experts as and when required).
4. Further (usually directed) TM experimentation and manual evaluation (repeating steps 2 and 3 as required).
5. Representation of knowledge and manual update of the expert module.

It should be emphasised here that all the PPR documentation used in this case study were existing documents. i.e. reports had not been prepared for TM purposes. Before any TM experiments were carried out on these PPRs, the reports were studied manually and the format, terminology used, agenda and key issues were considered and discussed. Utilising an iterative process, the key knowledge areas where KDT and TM should be concentrated were identified, discussed with the contracting company and prioritised using their domain expertise. This enabled the KDT process to concentrate on key knowledge areas and processes that were most important to the contracting company. The next section discusses an ontology development process that was used to support the knowledge discovery process.

9.4.2 Ontology development

It is necessary to determine an ontology of important and common terms within the PPR reports to facilitate the processes of knowledge search and knowledge discovery. In practical terms, to satisfy the requirements of simple exploratory TM experiments, it would be sufficient to analyse example PPR reports, find the common terms (which indicate the types of knowledge that are likely to exist in the reports) and then combine and refine these outputs with information provided by domain expert related to key areas of interest. The results from this analysis and refinement could then be used to direct the TM experiments to identify any useful knowledge that may lie hidden within the reports. However, the aim here was to develop an approach for knowledge discovery from PPR reports, so that particular types of knowledge can be targeted for knowledge discovery. With this in mind, an ontology was developed and extrapolated based on the combination of key terms identified in the example reports provided by the two companies and discussions with domain experts. The details of this approach are discussed next.

This approach involves the development of hierarchies to capture the various key issues identified within the PPRs. Each hierarchy has a single root which indicates the main topic areas shown in Figure 9-4. The Top Level hierarchy shows that the learning from a project will relate to one or more of the following topics, Finance, Quality, Communication, Building, Health and Safety, Labour, Environment, Time, Security, Materials, Plant or Project Stages of a construction project. Each of these areas is then considered in further detail in separate hierarchies, using key words from items of learning identified in the example PPR reports. For example, Figure 9-5 shows the hierarchy for finance related key words.

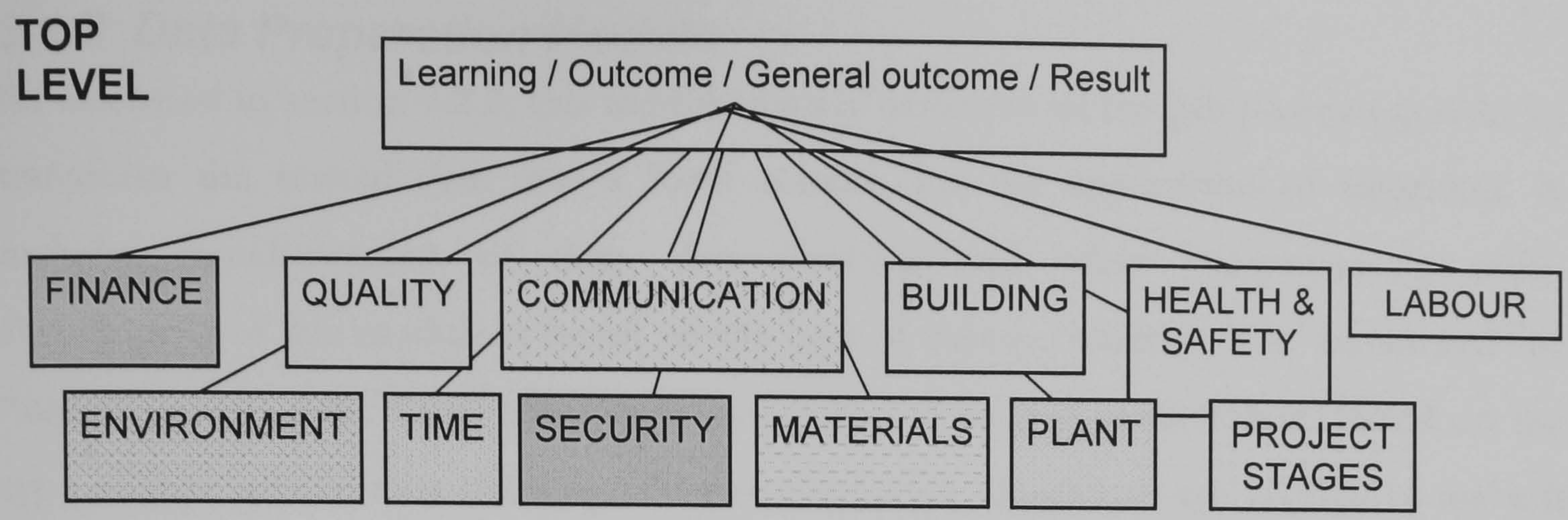


Figure 9-4: Top Level hierarchy relating to construction PPRs

Appendix 6 shows the rest of hierarchies corresponding to each category of top level hierarchy. In order to focus the text mining on particular topics of interest it is likely that two or more hierarchies will be utilised simultaneously. For example, to identify knowledge about delays on particular parts of buildings, both the "Time" hierarchy and the "Buildings" hierarchy would be used.

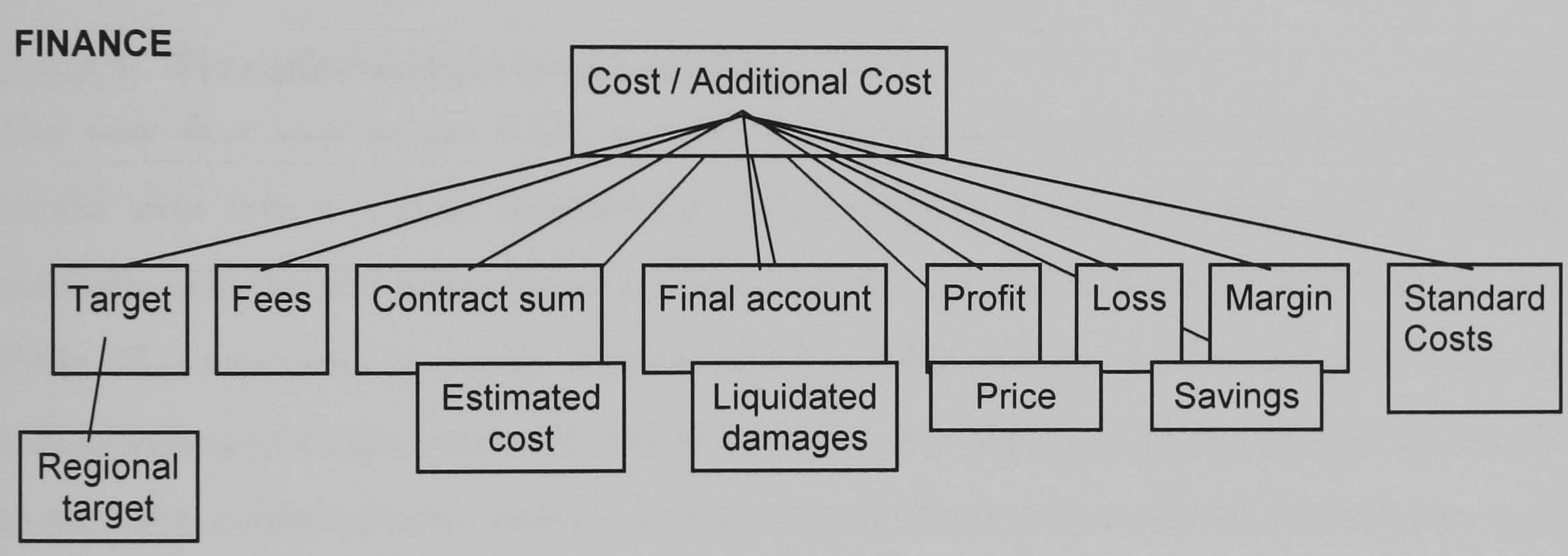


Figure 9-5: Hierarchy relating to finance from top level hierarchy

Another motive for development of an ontology based approach is to deal with the semantic issues and issues of multilingualism. The different terms may be used for the same context by different members of team or the same term can be used for different contexts. In addition, a dictionary feature has also been used to add any particular term and modify it based on the context. The dictionary feature also helps in identifying the meaning of a particular keyword/phrase which has several meanings for different contexts. Therefore, the combination of modification of the dictionary feature and an ontology based approach has been used to take care of semantic and multilingual issues. The next section discusses the functioning of the data preparation module of the knowledge miner to perform the pre-processing tasks.

9.4.3 Data Preparation Module

As discussed in section 7.2.2, this module mainly performs all the pre-processing tasks to transform the textual data into a form suitable for the application of algorithm. It includes transformation of data, data cleaning and other pre-processing tasks. Functioning of this module is based on the type of data i.e. whether it is "structured" or "unstructured" data. Firstly the data is acquired and its type is identified. Based on the type of data several data cleaning and transformation algorithms are applied to make it suitable for the mining engine of PolyAnalyst. In the context of textual data, this involves removal of unwanted and uninformative texts, stemming and shallow parsing etc. Data transformation is another step that is required to transform the data into a form suitable for input into the chosen data mining algorithm. For text based data, generally one type of text based file is converted to another type of file such as .pdf to .txt, to ease the computational burden and provide better knowledge representation.

9.4.3.1 Transformation and Loading

The very first step in the KDT process for PPRs is the transformation of gathered textual data into a format suitable for TM. Company 1 reports were semi-structured manually and divided into categories based on the original headings (issues) within the PPRs. This was done to enable exploration of possible patterns within particular issues as well as across different issues. All the textual data was imported from .doc file to .txt file to ease the computational burden. In the original document a tabular format was used containing 2 extra columns which did not provide any specific information. Therefore, these two columns were deleted and only the description column was considered as a knowledge source. In the context of company 2, the pictorial notations were replaced by textual information. Analysis for each company was done separately. In each case, the whole file was loaded as a new project into the software system.

9.4.3.2 Pre-Processing:

Unlike other data mining applications, in this research pre-processing of the textual data was carried out after loading the data into the PolyAnalyst system. Pre-processing is mainly done to reduce information overload and generate metadata. The textual data pre-processing steps are as follow:

- Removal of "Unwanted" text: In TM, punctuation delimiters such as commas, apostrophes, exclamation marks, as well as alpha numeric text and numerals are referred to as "Unwanted texts". These delimiters do not help to differentiate between the textual inputs, since they are reasonably well distributed throughout the report. Therefore, they are removed from the text.

- Removal of non informative words: A common technique to improve the accuracy of TM results and to reduce redundancy in the computation is to remove frequently occurring, common words from the text. These words are often defined by a "stop list" which typically consists of 200-300 words, including articles, prepositions, conjunctions and some high frequency words for example *are, the, from, can, may, etc.* In the present context, pre-identified words, which did not contribute to the essential information within the report, were removed to reduce the size of document and avoid information overload.

- Stemming: This is a common pre-processing step in textual databases and refers to a simplified form of morphological analysis by simply truncating a word. For example, agree, agrees, agreed, agreeing can all be stemmed to agree. The inbuilt stemmer of the PolyAnalyst software helps to reduce the document space and can provide a more concise document space representation.

- There were several words which were unlikely to contribute towards identification of useful knowledge. These key words need to be ignored while analyzing the text documents. For example in the context of finance, certain key words such as pay, wage, bonus, adjustment etc. were ignored. Key words need to be agreed with domain expert.

- Synonyms or different representations or abbreviations used were identified and marked in the TM dictionary so that they can be treated as one word. For example, "sub-contractor" is the same as "subcontractor" or "sub contractor", as are the terms "KPI" and "Key performance Indicator". Similarly, profit and financial success, happy and glad meant the same thing. Therefore, these need to be included in the dictionary to ease the text mining process.

The dictionary is a part of the software system used for the experiments and customizing the dictionary serves the purpose of increasing the likelihood of extracting useful knowledge from PPRs. Here it is important to mention that domain knowledge also known as background knowledge plays a vital role in text mining pre-processing operations to enhance concept extraction and validation activities. Although text mining pre-processing operations play a critical role in transforming unstructured content of raw documents into more tractable concept level data representation, the core functionality of TM system lies in the analysis of concept co occurrence patterns across documents in a collection. Indeed, TM components rely on algorithms and heuristics approaches to consider distribution, frequent sets, and various association concepts at an inter-documental level to enable users to discover the nature and relationships of concepts within a collection of documents as a whole.

9.4.4 *Text Mining Module*

The text mining module should apply an appropriate text mining algorithm to perform a specific function on the pre-processed data. After pre- processing, the textual data are ready for TM, which involves the application of various algorithms and functions to extract patterns, trends and discover useful knowledge. Text mining generally involves techniques from information retrieval, information extraction and corpus based computational linguistics to discover useful information and knowledge. As discussed in chapter 5. The main focus here is on application of technology therefore less information of algorithms has been presented. A detailed discussion of text mining technology and algorithms are presented in the handbook of text mining by [146]. A range of functions such as text analysis, link analysis, dimensional matrices, semantic search and summary statistics are applied in this case study to demonstrate that a variety of information and knowledge that can be presented to the user. The following subsections discuss the application of these functions in detail.

9.4.4.1 Text Analysis (TA)

The TA provides the morphological and semantic analysis of unstructured textual PPR reports in a database format. TA extracts and counts the most important words and word combinations from the textual PPR reports, and stores terms-rules for tokenizing database records with pattern of encountered terms. Here tokenization refers to the breaking of text into sentences and words. TA can be used either in unsupervised mode-to provide a better understanding of the most important terms in the investigated textual notes, or in supervised mode when a user focuses the exploration on a specific subject to accurately track issues important to the user. The terms or a combination of terms are generated as rules which record the number of times each term or combination of terms exist and where the occurrences are. These rules can be applied on the PPRs to find patterns of events or activities, issues, causes or achievements. For example, as shown in Figure 9-6, the "extension of time" rule applied to the dataset, brings a subset of reports where extension of time has been granted on the project. Similarly, For example: "Profit in excess of regional target" was identified by creating a rule using two keywords and phrases –"profit" AND "regional target" and applying them on the dataset.

Domain expertise may be needed to determine the relevancy and importance of combinations of terms identified in this manner. For example, if the word "scaffolding" occurred several times in the Health and Safety sections of the PPRs, perhaps in some combination with other words, this may well indicate that there is a recurring problem with scaffolding in a particular context or with the safety of a particular type of scaffolding. Further, these rules can be used as an input to other visualization techniques (e.g. Link Analysis) and classification tools (e.g. Decision Tree, or Decision Forest) in the software system. TA therefore can be used to highlight the commonly used words in various areas of PPRs such as planning, estimation, errors or mistakes, quality, health and safety, defects and many more. Identifying problems, issues and possibly their causes in this way may help managers to avoid them in future projects. They can then be captured as "item of interest" in the EM so that the moderator can make the partner aware of potential problems that may arise when they occur in future projects. TA can also be used to find where a particular word or its synonyms are used in various reports. Figure 9-6 shows the application of TA on PPR reports and generated rules. Left hand column shows the high-level hierarchies, middle column shows the time hierarchy and the corresponding keywords and phrases automatically extracted by text analysis in the form of rules and the right column shows the reports. This also facilitates the team member of a project to look at their areas of importance. For example, a user responsible for financial activities might be more interested in identifying knowledge related to additional cost, budget, contract sum, financial reporting, liquidated damages, pricing, loss, tender allowance and profit in excess to regional target, negotiations and discounted tenders. Furthermore, a combination of generated rules can be used to create a new rule based on the OR, AND and XOR operators. Based on the key knowledge areas, a set of rules were created using a set of keywords and phrases. These rules were further applied to the textual database to identify a subset of reports dealing with similar issues. "Extension of time" AND "additional cost" AND "loss" was used to identify a set of reports where extension of time resulted in additional cost and furthermore led to loss in the projects.

The application of this rule enabled four reports to be found where an extension of time and additional cost had resulted in a loss.

The results of the rule application experiments include an abundance of examples of rules that were successful in retrieving specific occurrences of certain keywords and phrases. The few examples above have simply been chosen to demonstrate how rules can be applied to retrieve knowledge. Rule application is an area in which domain experience is very important as an expert's input is needed to identify a set of phrases which should be used to create a useful rule and the expert is subsequently also needed to identify the relevancy of the results and issues of importance.

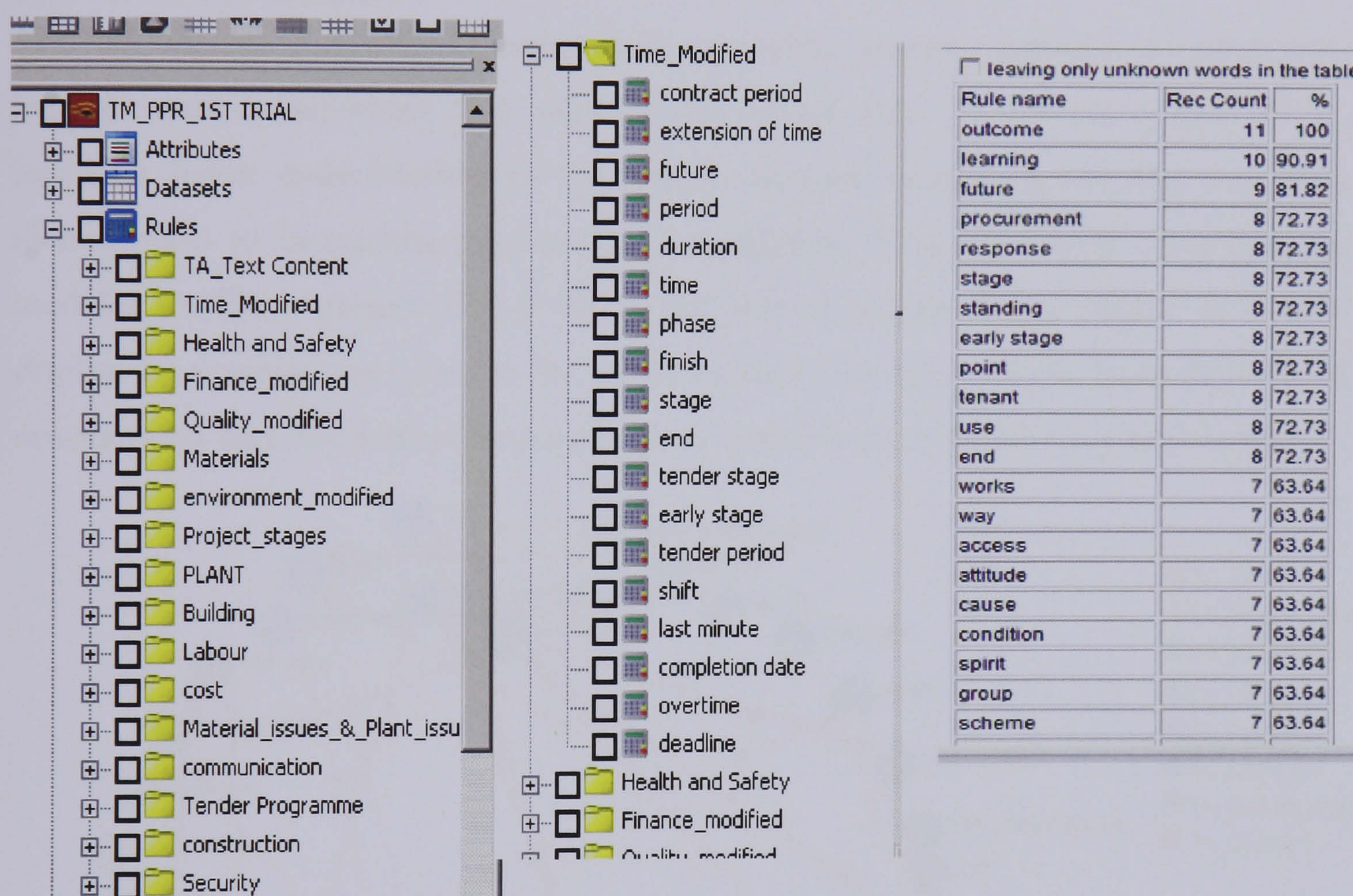


Figure 9-6: Text Analysis on textual PPRs and generated rules

In this manner, a set of specific knowledge can be generated through the application of TA and further application of rules on the data set. The User can use these knowledge to form rules in the form of IF (Condition) THEN (Action) to make it applicable to current moderator system. For example some of the rules that can be generated are as follows:

1. *IF ((project type==office building) AND (no change in the initial planning) AND (no liquidated damage) AND (No additional cost)) THEN (up to 4 Week extension may be granted if required to achieve a profit).*

2. *IF (extension of time is allotted for project) AND (additional cost is not shared between contractors and the company) THEN (predict a loss in the project)*

3. *IF (project type ==YYZ) AND (design team == XYZ) THEN (Predict the quality of work to be below expectations).*

4. *IF (there is any change in design at later stage of project) AND (redesign is required at extra cost) THEN (project might incur a loss).*

Supervised and unsupervised text analysis not only helps the user to identify key knowledge but also facilitates other text mining tools such as link analysis to identify the key knowledge.

9.4.4.2 Link Analysis

The link analysis (LA) reveals and visually represents complex patterns of correlations between various keywords that exist in the textual data. The visual output of LA facilitates better understanding of the hidden structure of investigated data and helps identification of interesting patterns for investigation. A detailed study of technology used for the LA is presented in [146]. In the present context, LA is applied to visually display the correlations between the terms extracted by text analysis from PPR based unstructured data and further investigate these correlations to identify key knowledge.

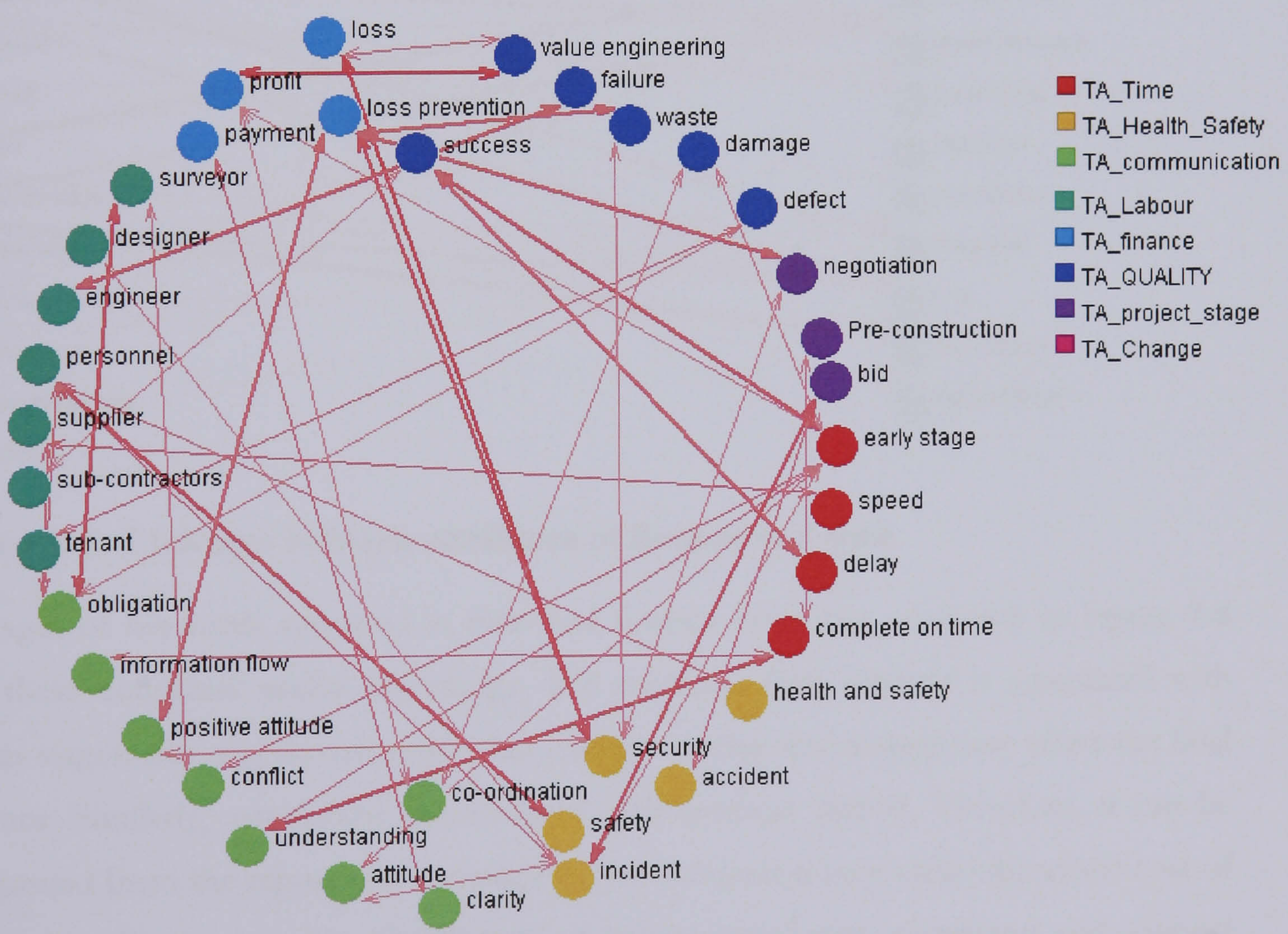


Figure 9-7: Link analysis between Keywords and their correlation

Figure 9-7 shows an example of the application of LA on the PPRs. LA has been applied on the ontology developed for companies. Each colour code represents a high level hierarchy of the ontology developed in section 9.4.2. It shows how a particular group of words are associated with another group of words. The strength of the link shows the correlation between keywords or phrases. In Figure 9-7, "value engineering" is directly linked to "profit" with strong support. Using PolyAnalyst, the cluster of reports where value engineering has affected the success of project can be shown by clicking the link line. Other key words and phrases can be explored in a similar manner. This analysis provides a way of information retrieval and enables relevant subsets of the original PPR data to be collected for further exploration based on the particular topics (or terms) of interest. Exploration of these links provides the user with a set of knowledge where one keyword or phrase has effect on the other. In a similar manner, link analysis between each high level hierarchy has been developed to show one to one relationships between ontology groups as shown in Figure 9-8, where linkages between finance and time are shown, i.e. various attributes of "Time" are linked with attributes of "finance".

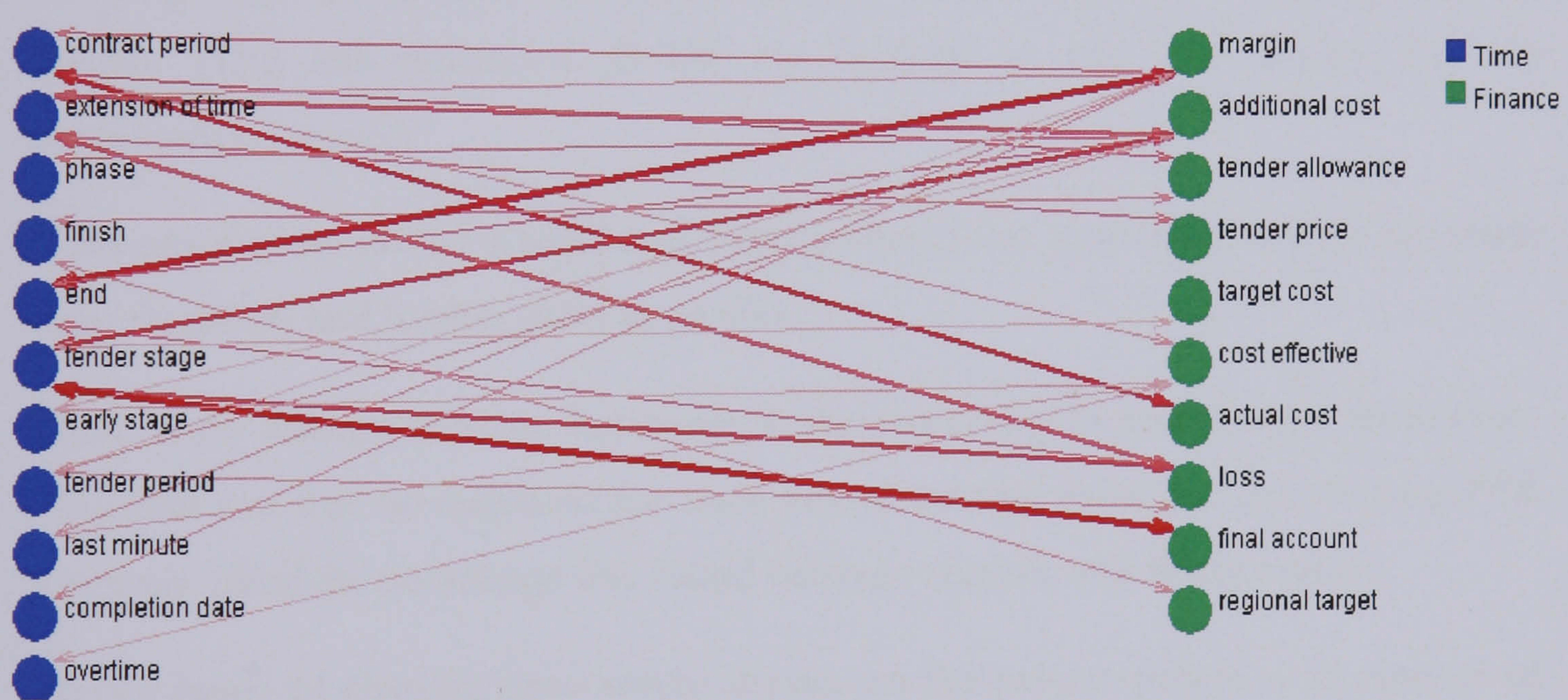


Figure 9-8: Linkages between attributes of finance and time

Linkages of keywords appeared in time and finance categories as shown in Figure 9-8 and these constitute useful knowledge. For example, final account is associated with tender stages; this may further imply that changes during tender stage may affect the final account. Similarly, actual cost is associated with contract period. Therefore, it can be interpreted from the report that a change in contract period may affect the actual cost of the project. To measure the significance of linkages minimum correlation and support are used a tuning parameters. Maximum and minimum number of links can be varied based on tuning the correlation and support. Correlation and support are the statistical measure to identify the validity of the linkages. Linkages with high support and correlation are less likely to be misleading than the linkages with low correlation and support. Linkages should be eliminated which are not statistically valid and may mislead the results. Exhaustive experimentation has been carried out by changing the minimum support and correlations. Some of the knowledge found can be interpreted as follows:

- The correlations identified in three reports between "maintain a good relationship" and "profit" shows that maintaining a good relationship with construction parties such as suppliers leads to a substantial discount and cheap rates which affect the profit on a project.
- The key issues which affect profit are quality, project handover, health and safety, additional cost, extension of time, changes during various stages of projects etc.
- The actual cost of a project is affected by the contract period - this was identified in a number of projects.
- Working with certain sub-contractors resulted in good quality work and a good margin. (The sub-contractor details are missing in the PPR report due to confidentiality issues)
- Two reports mentioned a small loss being incurred due to confusion between client team members and further delay in project.
- There is no linkage between keywords in the environment and finance hierarchies. This indicates that no emphasis has been put on relating these two areas during PPR meetings. Similarly no linkage was found between security and finance issues.
- Several kinds of changes have severe impact on the project progress. A number of projects faced changes during planning stage. Out of 11, 3 projects faced a loss due to change of some kind such as personnel working, plan and design etc, and 2 projects showed that a loss occurred due to design change and expensive redesign at late stages of the project.
- Adoption of "Value Engineering" resulted in profit in 2 projects.
- Changes at various stages of the project have affected the Lead-in-times in several projects.

- A set of good practices in terms of communications has been observed in 7 of the PPR reports.

In this manner, one can see that the results obtained using LA are very promising as they show that some useful (previously unidentified) knowledge can be extracted from the PPR reports. The processes presented in the preceding sections are examples provided to explore the potential of text mining of PPR reports, and based on the results from these processes, it is apparent that there is a lot of scope for using text mining to discover useful knowledge from PPR reports. These knowledge can then be transferred in the form of IF-THEN rules by the user and added to the expert module manually for further use by Moderators. In order to better explain the scenario, let us consider 2 "item of interest" such as design change and accident. The knowledge can be represented as follows:

> *IF "there is a change in design" THEN "alert the contracting company for possible loss" AND "negotiate with designing team for redesign".*
>
> *IF "accident happens" THEN "alert the contracting company for extension of project".*

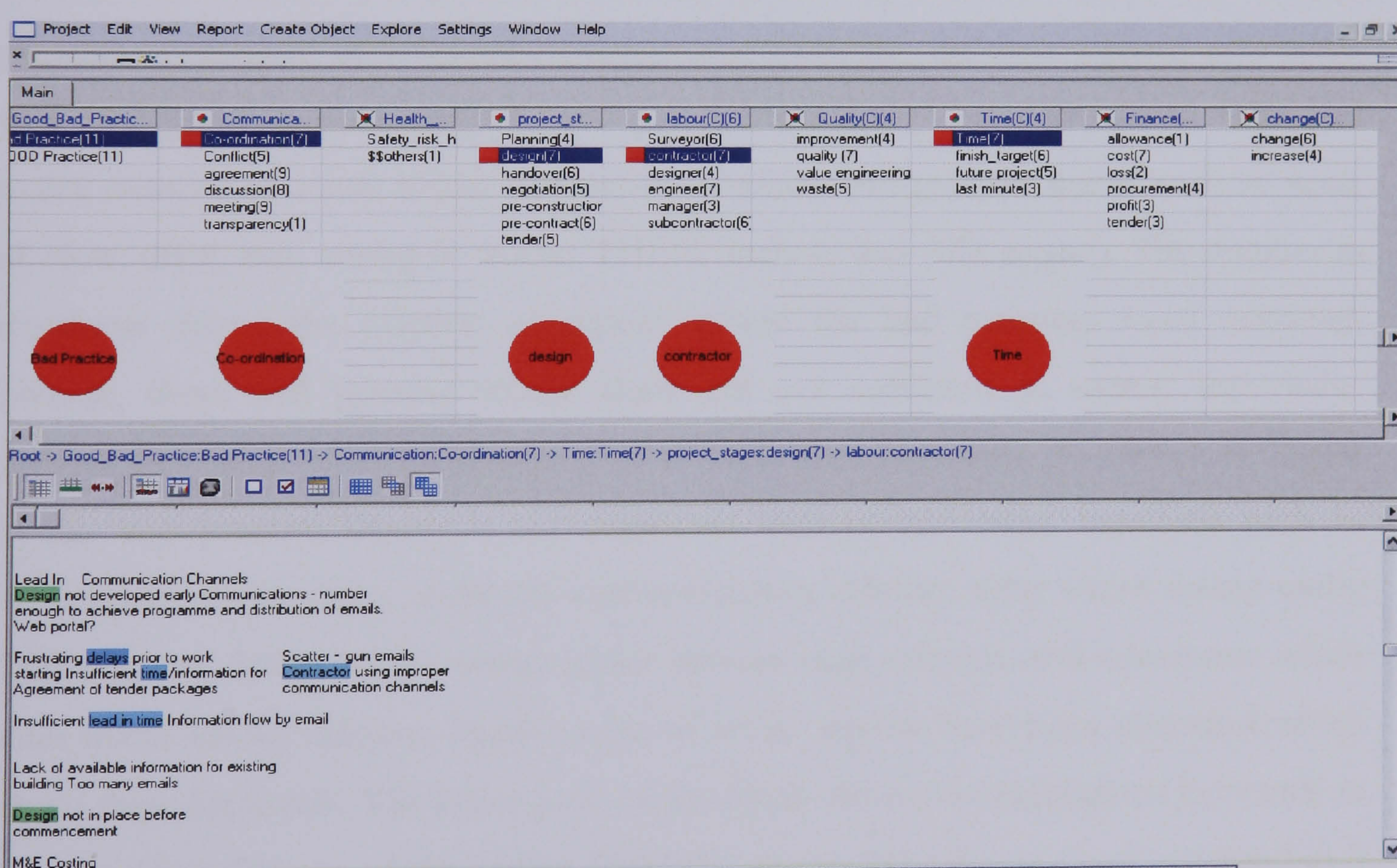


Figure 9-9: Dimensional Matrix representing key knowledge areas.

In this manner, the task of the Moderator will be to identify these rules and perform corresponding action if there is a change in the item of interests or the condition part of the rule is satisfied. However, one of the limitations of LA is that it only relates two keywords at a time. In order to overcome this, the next section discusses the use of a dimensional matrix or Text Online Analytical Process (OLAP) as a way to discover key knowledge based on a combination of keywords.

9.4.4.3 Text OLAP (Dimension Matrices)

A Text OLAP or dimensional matrix having 6 columns representing each knowledge area with several key words and rules was created to analyse the PPR reports. More detail of this approach is presented in section 5.5.1. The Dimensional Matrix uses the OLAP - On-Line Analytical Processing feature which provides the user with the capability of performing multi-dimensional analysis of the data. Each column consists of different cells where each cell (block) represents the key word(s) to be searched for within the PPR Reports. While working with the OLAP report, the user defines the values of one or more cells and then browses the subset of records, belonging to the selected cell which represents a keyword or combination of keywords. One of the advantages of this approach is that the dimensional matrix can be exported to new project and reused for new dataset.

The dimensional matrix is pictorially shown in the Figure 9-9, where every column represents a category of keywords/phrases. For example, the first column represents Good_Bad_practices consisting of good practice and bad practice keywords. Here bad practice consists of a rule IF(the PPR report contains bad practice words such as poor, bad, slow, delay, late, wrong or worse) THEN (include that PPR report). The number in parenthesis shows the number of reports where the bad practices word occurred. However, these bad practice words alone are not sufficient to extract knowledge therefore they need to be combined with several other keywords. As shown in Figure 9-9, the bad practice keywords are combined with several other keywords such as coordination which comes under the communication column, delay which comes under TIME column, design which comes under project stage column, and contractor which comes under labour column. Finally, a set of seven reports have been identified which contain these keywords. The lower pane of the figure shows the highlighted keywords in different colour for one of the report. Similarly, phrase "Additional Cost" occurred in 6 reports. While combining with "Extension of time" in the column TIME, both together appeared in 4 reports. The knowledge derived can be interpreted as due to extension of time in 4 projects resulting in the company incurring additional cost. Furthermore, these two can be combined with Quality to identify how extension of time and additional cost affected the quality of the product. A dimension for "change" has also been created to identify the changes that might happen during the project and identify its impact on the over all success of project. In this manner, it can be seen that knowledge can be derived from PPRs based on the domain expertise using LA.

9.4.4.4 Summary Statistics

Basic statistics can be generated for the PPR text at various stages of the TM to compare its attributes, key words, or generated rules. These include means, standard deviation, frequencies, frequency chart for each category, strings, and yes and no variable, etc. In the present example, these statistics help in finding the frequencies of frequently used words in the PPR reports. In this way, summary statistics are useful in identifying which reports are more important in the context of a particular issue and an example of their use is given in Figure 9-10. It shows that for company 1 out of 27 projects a profit has been achieved in 17 projects and 10 suffered through loss or no profit. Similarly, out of 26 projects, some additional cost has been incurred in 8 projects. This information can be used to extract useful knowledge relating to these events by the domain experts. These types of reports can further be explored to discover useful knowledge relating to profit or loss of company. However, in the present context due to a limited number of reports it was difficult to further explore those reports.

Logical (yes/no) attributes:	Values	N of 1,	%	N of 0,	%
margin	27	14	51.85185%	13	48.14815%
"additional cost"	27	10	37.03704%	17	62.96296%
profit	27	17	62.96296%	10	37.03704%
"tender allowance"	27	8	29.62963%	19	70.37037%
"target cost"	27	4	14.81481%	23	85.18519%
"actual cost"	27	18	66.66667%	9	33.33333%
loss	27	9	33.33333%	18	66.66667%
"final account"	27	21	77.77778%	6	22.22222%
price	27	24	88.88889%	3	11.11111%
"regional target"	27	5	18.51852%	22	81.48148%

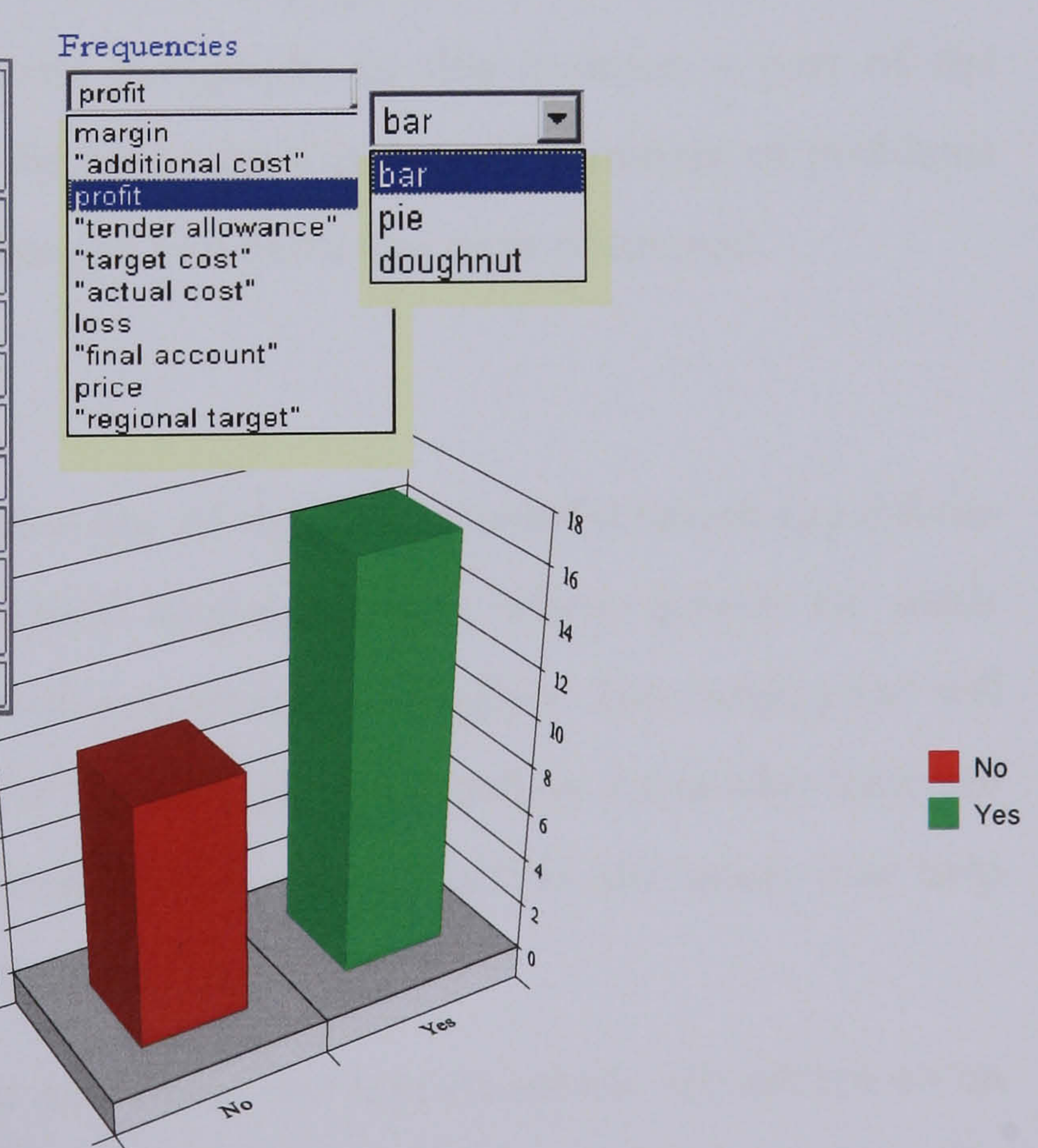


Figure 9-10: Application of summary statistics on PPRs.

9.4.4.5 Summarization

PPRs are commonly neglected and under exploited because of their length and the consequent time and effort that subsequent project teams need to apply in order to read through them to identify any knowledge and lessons that are relevant to their new project. Summarization techniques can reduce the PPR reports to a fraction of their original size whilst still retaining significant content in the summaries. Summarization techniques determine the semantic weight of sentences written in the PPRs and only those sentences whose semantic weight is higher than the threshold are kept. The final summary then lists the most important sentences written in the PPR report. The size of the summary can be changed by changing the semantic threshold. Useful knowledge can be retained within very short summaries, as shown in the following example.

No profit gain, with extension of time. It is understood that the **Framework Manager** *(at Head* **Office***) has made an alternative concession to Company "X". The people giving instructions to "ABCD" were not in a position to explain and* ***"X"*** *middle* **managers** *had no wish to assist "ABCD", as they did not want to see the work outsourced. The X* ***Framework Manager*** *negotiated this Contract and the arrangements were beyond the control of this Business Unit. "ABCD" have produced them, but the "X"* **Manager** *concerned, ----, continues to object to minor items and will not sign then off.*

In this case an original document of approximately 12 pages has been reduced to a useful set of summarized information within one paragraph. In this manner, a part of the Moderator's action might be to provide the user with this type of summary of problems that had occurred in the past related to changes to a particular item of interest.

9.4.4.6 Semantic Search

The semantic search engine of TM provides one of the most powerful search capabilities for PPR reports. It is very similar to natural language query, where queries are made using natural language, typing a question in conversational English. The result pane will display all the related answers to the query. Results are displayed as a tree-like structure based on the question. This sub-tree of concepts that are related to the query may help the user in simulating a better answer to the asked query.

For example, during this study, a search was made for reports which are related to an extension of time for the project. The search result pulls all the sentences from the original PPR reports that best respond to the query and places then in the result pane. It shows all the text where the project report mentions the extension of time, or extension of period as shown in Figure 9-11. In this manner, the user can use semantic search to identify the information relating to their area of interest and further consolidating these information in the form of knowledge applicable to moderator.

Here it is important to indicate that text mining can identify apparent relationships and new knowledge but the final judgements on the value and usefulness of the identified knowledge must be down to human experience and judgements. However, a further benefit of the Text Mining approach is that it identifies which reports or sections of reports need to be read when making these judgements.

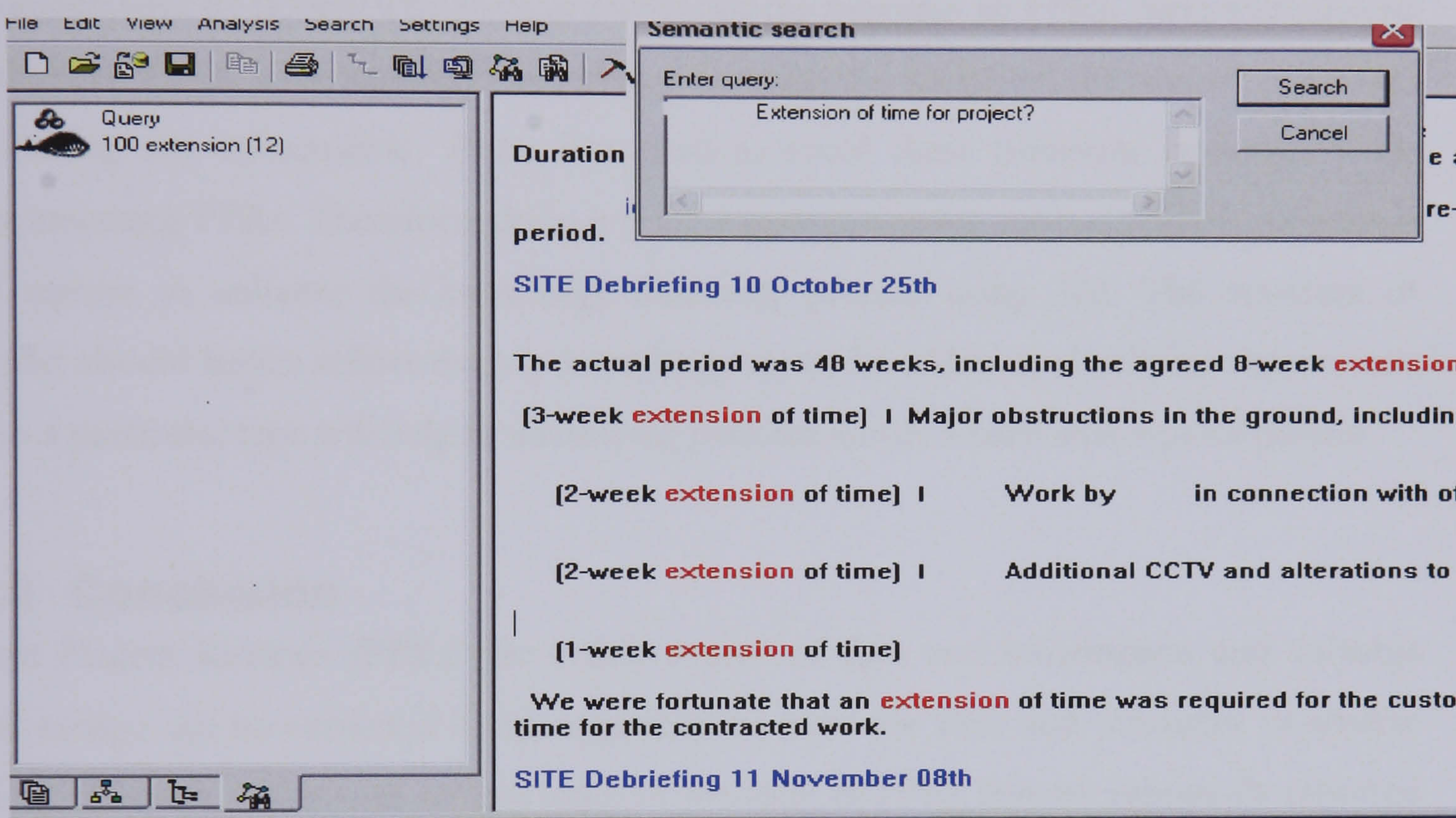


Figure 9-11: Semantic search on PPR reports.

9.5 Discussion and Limitations

In this manner, it can be seen that a variety of TM techniques have good potential to derive knowledge from PPRs. However, several limitations are identified with the application of the proposed approaches. In the past, text mining has been applied to cases where there are plenty of data available for analysis. In the present context, one of the limitations of this approach includes availability of data. The results obtained are on the basis of a very small set of data. Text mining is more useful to large set of reports preferably more than 100, as with large set of data there is a good potential to find patterns across a number of project. In present context, small set of data that is only 29, it was difficult to identify consistent patterns in the occurrence of key words and phrases.

The analysis indicated that although LA may show a strong correlation between keywords/phrases, it is not always the case that there is a direct relationship between these words. It is mainly due to the lengthy nature of the PPR reports. Generally LA is more effective when the length of the text based file is 1-2 pages. Whereas in present context, the length of PPR reports vary from 12-15 pages. Sometimes due to the distance between appearances of keywords/phrases within in the report, it was not possible to identify useful knowledge as those keywords have been used in different context. In addition, some of the reports are written in such a manner that it was difficult to understand by people who have not attended the meeting. In PPRs, there were several sentences which were irrelevant and thus increasing the length of the reporting without providing any information. There is a need to avoid these irrelevant sentences while documenting PPRs. Therefore, there is a need to improve the representation and format of reports to enhance the knowledge discovery process using TM. The structure of report should better reflect the key knowledge areas. In addition, classifying the projects into a particular type will help in identifying patterns across a particular type of project.

9.6 Conclusion

Post Project Reviews (PPRs) are a rich source of data and information and valuable knowledge can be extracted if the organisations have the time and resources to analyse them. However, there are several types of problem in PPRs that are repeatedly reported in research literatures as discussed in section 9.4.3. They relate to issues of how knowledge from PPRs and the recorded lessons learnt from projects can be quickly and effectively identified and disseminated to maximise the benefits gained by the organization and new projects. A hypothetical example of a construction project moderator has been presented and emphasis has been placed on how a variety of knowledge can be extracted form PPRs using a variety of techniques from TM. This knowledge can then be interpreted by the domain experts and presented to the use for addition or update of an expert module.

KDT and TM are fairly new research areas which address problems of information overload and provide many tools and techniques to help identify useful relevant information and present it to users in a concise or easily searchable form. The main focus of this chapter has therefore been to determine the potential benefits of the application of TM on PPR processes to extract useful knowledge for Moderator's knowledge update. The amount of knowledge generated during PPRs makes them ideal candidates to explore the use of KDT techniques, and this study has helped to show that TM has an enormous potential for use and benefits for future project improvement, avoiding mistakes, improving customer service, and making the organization aware of previously unknown facts and problem areas. The case study discussed in the section 9.4 has shown that TM could prove to be a very useful tool to discover patterns, trends and hidden relationships between various issues, topics and keywords used in PPR reports.

10

Knowledge Miners for Tendering Opportunities

This chapter proposes that the knowledge discovery Module of Moderator can be used to raise the awareness of possible business opportunities and corresponding partners by examination of invitation to tender documents. A case study of UK based SME is considered to illustrate the concept. It has been shown that how various text mining techniques can be used to identify possible business opportunities and business partners using their area of interest and competencies stored in the shared information of the collaboration pool.

Chapter Outline

10.1 Introduction

In order to compete with the market leaders, enterprises are concentrating on their core competencies and collaborating with other enterprises that complement their skills and core activities. Collaboration between companies can facilitate both strategic and operational foci, allowing individual members to exploit their core competencies to fulfil the mission of a project or service.

Market challenges are more intense for small and medium enterprises (SMEs) may be held back by lack of resources, finances, and new information and communication technology systems. Therefore, to remain competitive, SMEs are looking beyond their self centred passive environments and are forming inter-enterprise networks to achieve their business goals. This is much more significant in the context of SMEs sector within the UK, as according to the Department of Trade and Industry Small Business Services [178, 179], the SME sector accounts for approximately 60% of UK GDP and 58% of employment. Crucially the SME sector plays an important role by supporting the large business sectors via the supply of products or services. A large variety of collaborative networks (CN) have emerged during the last few years as a result of the challenges faced by both the business and scientific world. A detailed study of these types of collaborations and differentiation between them is detailed in chapter 3. In this chapter, two manifestations of CN are considered as VO and VBE (Collaboration pool).

Core competency centric companies within the pool can extend their business processes by teaming with other member of the collaboration pool having complementary competencies and having similar focus. However, one of the major challenges is the identification of appropriate business opportunities and selection of right partners to accomplish the goal. When a business opportunity is identified by one member, which is acting as a broker, a subset of the organizations from within a collaboration pool can be selected to form a virtual organization [180].

Therefore, this chapter proposes that the knowledge miners of moderators can be used to identify the possible business opportunities based on the competencies and areas of interest of the members of the pool. It could also prompt the broker by identifying possible collaborating partners from the pool to provide the complementary competencies required to fulfil the requirements of the business or service.

10.2 Problem overview of Tendering Process: A case study of UK based SME

In order to understand the problems and requirement of moderator service, let's consider a scenario associated with an SME AKC Ltd as follows:

> *AKC Ltd is an SME and a member of consortium XYZ Net. XYZ Net is a network cluster of SMEs, whose members have potential and wish to collaborate with each other based on agreed terms and conditions, if a possible business opportunity arises. AKC Ltd generally deals in areas related to in-house or onsite training in the areas of health and safety, environmental issues, IT, management and web based learning. They also provide services related to the design of software related to web based learning and e-commerce, design and development of web services etc. Its main strength lies in the customer focussed staff that understand the current and future market requirements and provide customers with very competitive price and service. Being an SME with 40 full time employees, the company always seeks for new business opportunities and projects that suit their competencies and area of interest. They often do so by identifying appropriate requests for tender proposals which appear in various sources in the market. In order to ease their search, the company is registered with different tender alert websites such as Tender Electronics Daily, sell2wales etc. Every morning, one of the managers of the company receives notification for tenders in the form of emails or RSS feeds with a list of suitable tender opportunities for the firm. These notifications are based on the criteria which the manager has chosen on tender alert websites. These emails are generally a page long containing the short description of the project. Then job of the manager is to manually scan each email and RSS feed and identify possible tender opportunities based on it's relevance with respect to the company's profile, requirement and its suitability in current market situation. He identifies a list of possible tender opportunities and then passes them to other members of the management team for discussion and selection.*

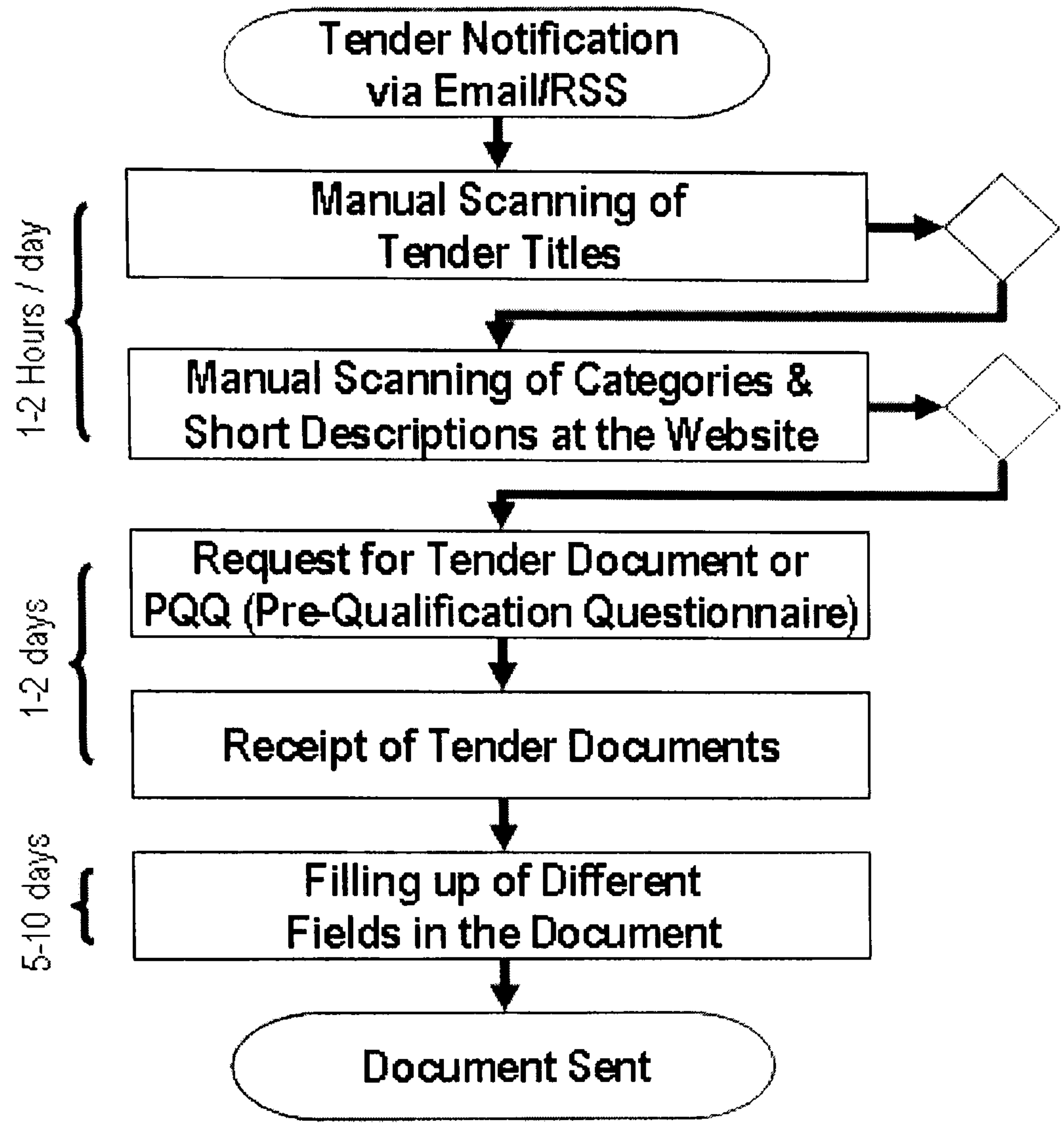


Figure 10-1: Process of tendering

It can be seen from the abovementioned scenario that the process of tendering involves several stages starting from notification of tenders in the form of emails, through identification of partners to the submission of bids as shown in Figure 10-1.

Online tender scanning and selection of suitable appropriate tender opportunities is very important for an SME as loosing bids may damage the reputation of the company and waste valuable time and money for the organization. Thus no SMEs wish to commit their resources on projects where there is only minimal chance of success. SME staff may spend 1-2 hour daily to find suitable tendering opportunity that might exist in the market. Although the company is registered with tender alert services which list the tenders based on their resemblance to the company's profile and tender requirements, some opportunities may be overlooked as they can not undertake the project on their own, but may be able to do so in collaboration with other SMEs. There are also many tender opportunities received from tender notification services which upon examination fall low in the manager's priority list, as they would not provide a good business fit for his company. In addition, since this process is manual, the search space is limited to only a couple of tender notification services. In this manner, a company may miss many opportunities as the exhaustive search of databases is not possible due to manual nature of the job and the restricted search space.

10.3 Function of Moderator for Tendering Opportunity

This research proposes that the integration of knowledge discovery module and application of corresponding knowledge miner can identify appropriate business opportunity based on the core competencies and the areas of interests of the company. In this context, the major function of the moderator is to:

- Firstly capture the knowledge associated with the team members of collaboration pool. This includes profile, areas of interest, competencies (preferable with evidence of past performance), and the type of project they are interested in.
- Secondly, identify the sources of tender data such as tender alert services and track the appropriate tenders based on required competencies and subject areas.
- Identify possible collaboration partners by categorizing the pool and matching it with the requirement of tender invitation documents.
- Notify the user with possible tenders and collaborating partners.

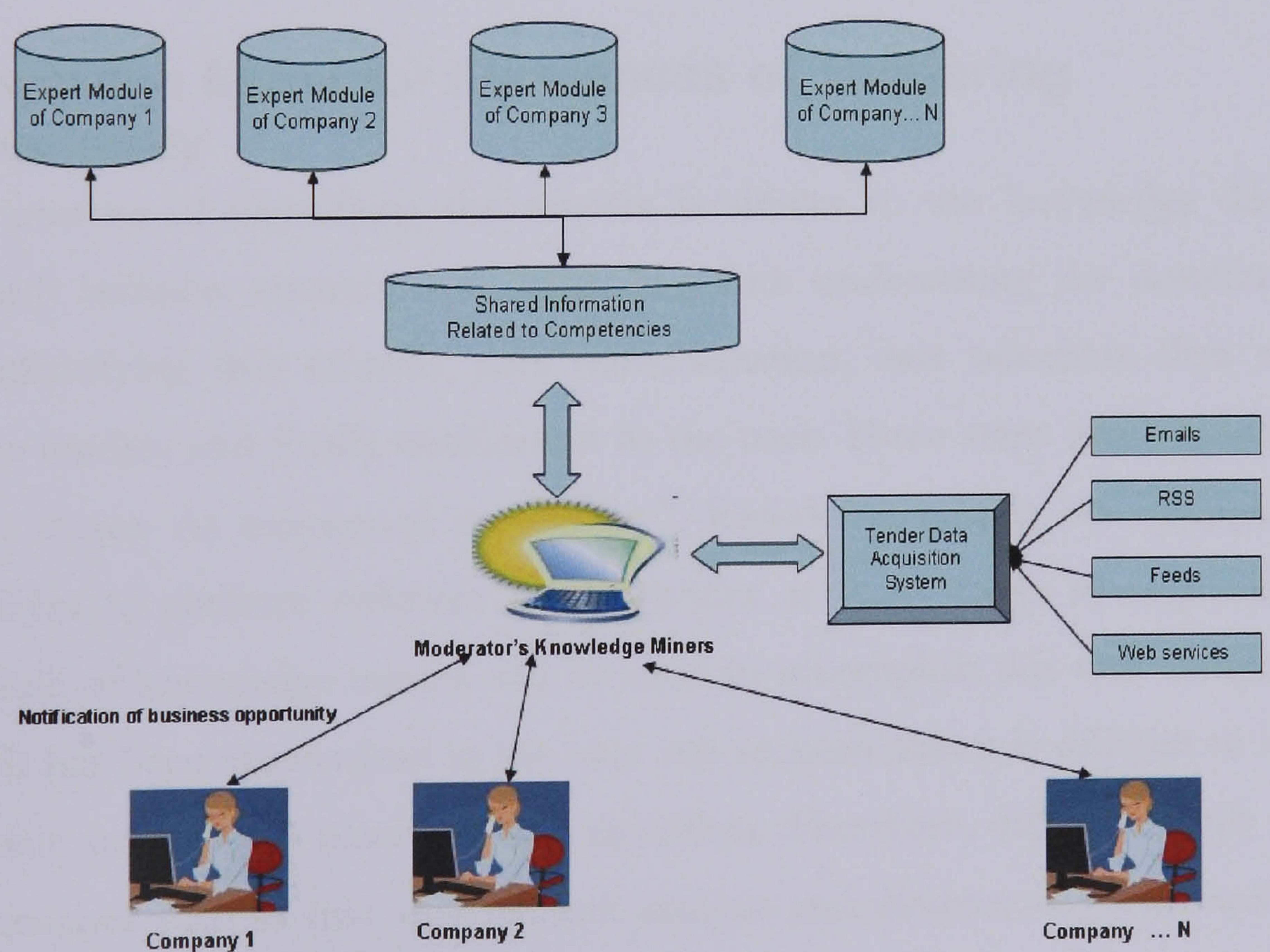


Figure 10-2: Functioning of the proposed approach in context of tendering study

To accomplish these goals, Moderator should have access to a variety of knowledge about the members of its collaboration pool. Figure 10-2 illustrates the functioning of the proposed approach in the context of this case study. In virtual collaboration, enterprises temporarily share competencies, areas of interests or resources in a collaborative rather than competitive manner. The basic structure and the content of the Moderator remains the same as mentioned in chapter 7. Every company within the collaborating pool is associated with an expert module, in which the KAM stores knowledge about the company including company's profile details such as its name, contact details, knowledge related to its areas of interests and competencies etc. Some of the knowledge in the expert modules such as competencies, capabilities and areas of interest need to be shared and therefore should be available from a shared database.

As shown in Figure 10-2, the knowledge discovery module has access to the shared information database, as this information will be used to guide the knowledge discovery module to search of appropriate tender documents. The invitation to tender data can be found in several formats such as RSS, Feeds, emails, and websites. It is assumed that the tender data acquisition system has access to these data sources and is capable of collecting invitations to tender related to data from these sources. In the present scenario, the knowledge miners of the knowledge discovery module are used to extract the possible tender documents using the information stored in the shared info.

10.4 Knowledge Miner for Awareness of Tendering Opportunity

The whole process of identifying the tenders is similar to the knowledge discovery process, which includes several stages beginning with understating the domain of an enterprise, identifying data sources, data transformation, data selection, data mining, matching the tenders and finally notification to the user. These steps can be iterative at any of these stages. As mentioned in chapter 7, knowledge miners are equipped with several modules to perform different functionalities at these stages as required. Data mining module of knowledge miners will be used to accomplish this task using several algorithms. It has been emphasized in the next sub sections that it is difficult to identify the appropriate tenders just based on one algorithm. Therefore, different kinds of text mining approaches such as text analysis; link analysis and dimensional techniques have been applied. It also discusses the shortcomings associated with the individual approaches and the need of others. In this research, PolyAnalyst 5.0 software has been used as an instance of knowledge miner to perform the text mining tasks.

10.4.1 Domain Understanding and Tender Data Collection

The functionality of the knowledge miner in the present context is dependent on the knowledge available to the Moderator. This knowledge will be used to guide the search of appropriate tenders and corresponding partners. As this research is exploratory in nature, an iterative approach has been adopted to achieve the goal. Firstly, domain experts are consulted with a view to understand the key terms they are likely to be see in a tender document. Based on this information a set of ontologies is developed to guide the search process. Secondly, tender data sources are identified and corresponding invitations to tender documents are collected. Thirdly, several text mining algorithms are applied to identify the appearances of key words in the invitation of tender documents. Finally, identified invitations to tender documents are communicated to the user for further discussion.

World Wide Web and internet technology has enabled various organizations such as companies, governmental agencies, academic institutes and service providers etc., to publish their tender needs electronically. Data acquisition system should be capable of capturing these new business opportunities floating in the market and pass them to the Moderator. One of the prevailing methods for collecting data on the web involves *crawling* the web, downloading the content of websites, and creating a searchable index. Web crawling refers to the process of selecting a set of URL (Universal Resource locators) and downloading their content from the corresponding web servers. The e-documents are generally available in various natural language formats such as .htm, .doc, .pdf etc. These tender documents include a wide range of information about the possible business opportunity such as basis of the tender, background information of the work, user experiences, specification of tender, criteria of selection and validity period etc.

In the present context, for the experimentation purpose, the data are collected from various data sources such as www.sell2wales.com and tender electronic direct in the form of emails. As shown in Figure 10-3, these documents contain a brief description of the project, place of performance along with CPV codes. As the purpose of the research is to illustrate how a knowledge miner can be used for identifying business opportunities, an example set of 40 such tender documents have been collected.

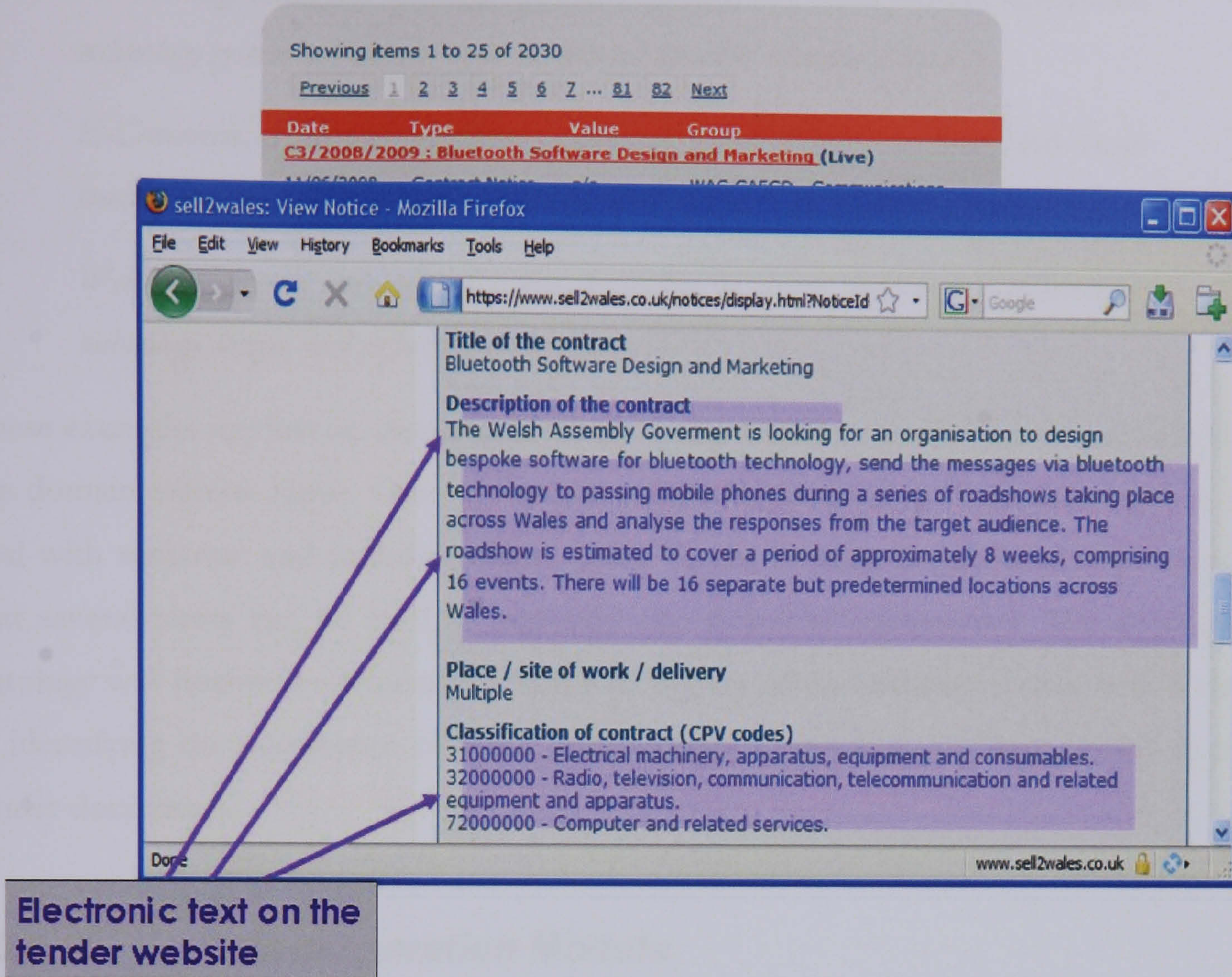


Figure 10-3: Invitation of Tender document.

Before the application of text mining it is necessary to develop an ontology corresponding to the competencies of the company.

10.4.2 Ontology Development

Development of an ontology guides the text mining process towards identifying specific appropriate tenders. It is likely possible that two or more terms can be used to represent the same context. For example, in the present context, e-learning, web based learning, internet based learning, and online learning can all mean the same context. Therefore, based on domain expert knowledge, an ontology has been developed and extrapolated relating to the core competencies and areas of interest of the company. These ontology are basically major factors on the basis of which tenders need to be filtered and can be broadly categorized in terms of technical requirements, financial positions, deadlines, competitiveness, past relationships and geographical location etc. An ontology needs to be developed for each category. In the present context for the company AKC ltd, a set of ontologies has been developed for the technical requirements as follows:

E-Learning, Electronic learning, web based learning, web based training, information technology system in distant/ flexible/ virtual/ blended/ education/ learning.

E-Commerce, Electronic commerce, e-business, internet based business, web based trading etc.,

Web development, web services, online technology in design, web design, information technology design, and software system design.

These examples are just an instance of the several possible examples which are based on the domain experts input. One of the major reasons to develop these ontologies is to deal with semantic and multilingualism issues. It can be seen from the above example that several terms can be used to represent one technical competency. The developed ontology was further implemented into the dictionary of the software system with a view to identifying the occurrence of these competencies in the form of "key words" in the tender documents.

10.4.3 Data preparation Module

The Data preparation module performs several steps such as transformation of data, data cleaning and other pre-processing tasks. It can be used to transform the invitation to tender related data received from various sources into a format which is suitable for text mining. For example, .pdf or .htm files need to be converted into .txt to make it homogeneous and provide better knowledge representation after the text mining process. In the present context, the emails received from various sources or web based contents are generally one page in length and in the form of .pst or .htm. All the textual data related to the tendering requirements are converted into .txt format to ease the process of knowledge search. These files are loaded into the PolyAnalyst software which is basically acting as a knowledge miner to illustrate the functioning of the Knowledge Discovery Module of the Moderator.

In the present context it has been identified that there is no need to perform pre-processing except to modify the dictionary. This is due to the fact that the algorithms used for this case study such as link analysis and Text OLAP do not require pre-processing of data. The dictionary is updated based on the domain expert's knowledge and developed ontology for key word.

10.4.4 *Text Mining to Tender Documents*

Text mining of invitation to tender documents is performed by the modelling module of the knowledge miner and it is detailed in section 7.2.2. The modelling module should apply an appropriate text mining algorithm on the pre-processed data to extract key competencies required in the tendering documents and further associate a linkage between the competencies of different partners to indicate the collaboration required for possible business opportunity. A range of functions such as text analysis, link analysis, OLAP-Dimensional Matrix are applied in this case study to demonstrate the process of knowledge identification.

10.4.4.1 Text Analysis

Text Analysis extracts and counts the important keywords or a combination of keywords from the tender based textual data and stores them in the form of rules. In this case study, the text analysis is more focussed on identifying the competencies and area of interest of the companies in the form of keywords. Therefore, guided text analysis is used to identify the tendering documents which are based on the information available in shared info of the pool. This approach identifies the possible invitation of tender documents which illustrate the competencies and area of interests of the pool. These competencies and area of interest are captured in the form of keywords or phrases and list the number of textual documents where they are appeared. Similar approach is also adopted by several tender alerting services. However, their capability is limited to broad areas and therefore does not focus on the need of a particular company. For example, sell2wales offers a similar service to identify the tenders where software development is needed. However, it is unable to identify the tenders which relates to software development for e-learning, which is more focussed on the competency of the company. That results in the number of invitation to tenders and therefore increase the time for manual inspection. In addition, current web services do not offer service of the partner selection problem.

The keyword/phrases are extracted in the form of rules. A combination of keywords can then further be used to create a new rule using OR, AND or XOR operator. Based on the competencies, a set of rules can be created corresponding to each company. These rules can then be further applied to the textual database to identify the relevant tenders where such types of competencies are required and area of interest match with the tender requirement. For example, to identify the tenders relating to software development in domain of E learning in the Cardiff or Wales area, a rule can be generated in the form of IF the document contains "(Software development AND (E-Learning OR Web based learning OR Internet based learning) AND (Cardiff OR Wales)) THEN Select the Tender for corresponding company. The application of this rule enabled 3 tenders' invitation documents to be found out of 40 documents where software development for e-learning was required in the Cardiff or Wales area. Rule application is an area where domain expertise is required as expert's input is needed to identify a set of phrases based on the past experience. And subsequently identify the relevancy of the identified business opportunities and other issues of importance. The process of formation of a rule from generated keywords is shown in the Figure 10-4.

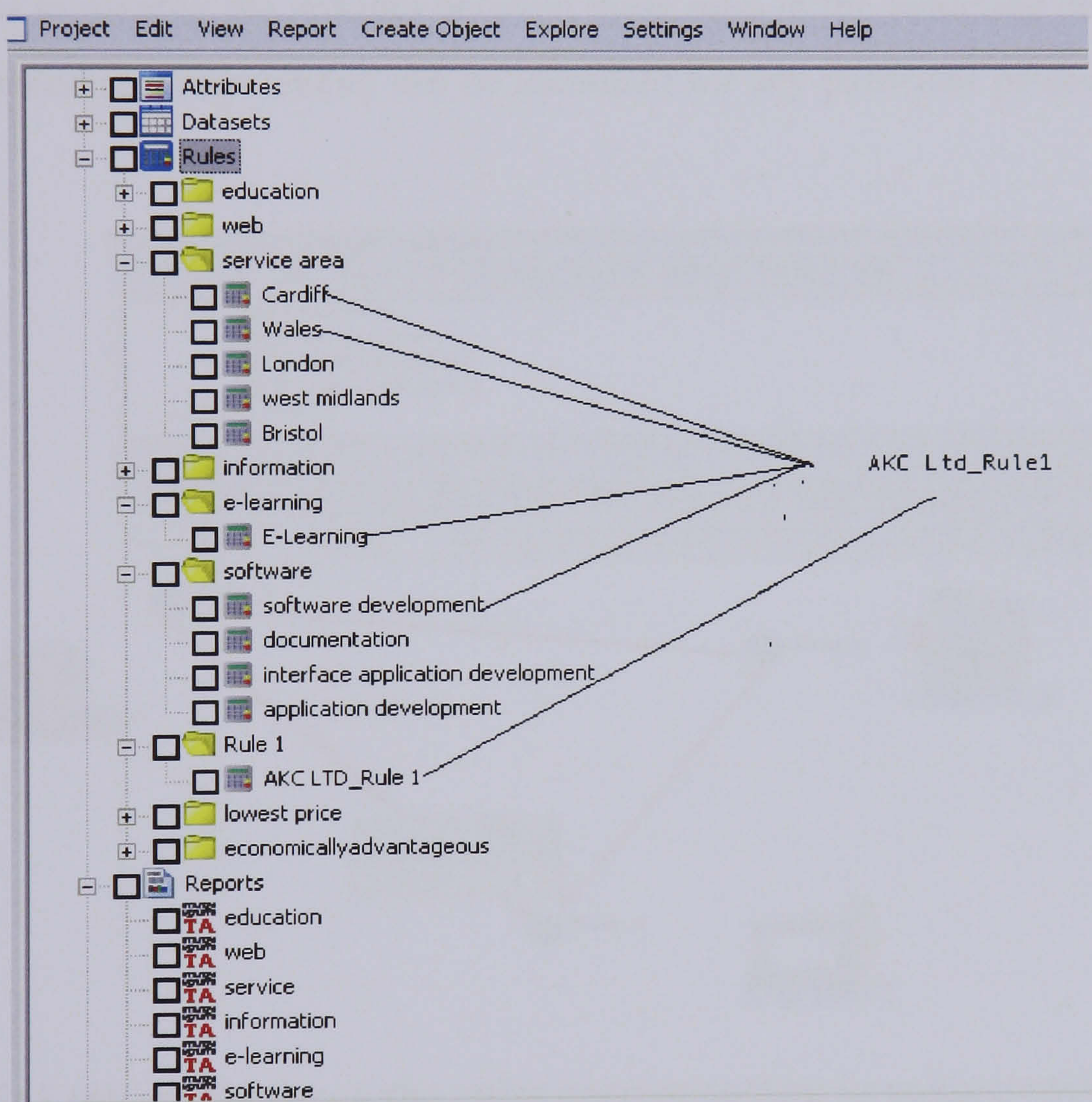


Figure 10-4: Formation of rules from generated keywords/phrases

In this manner, text analysis as an initial step of text mining process can be used to identify the tenders which are specific to a company and correspondingly the Moderator can make the company aware of the business opportunities specific to them. The outcome of the text analysis can be used for further application of other algorithms such as link analysis or OLAP. However, application of this approach is limited to the identification of tenders for a specific company only and therefore does not establish the collaboration between companies. In order to establish the collaboration between a pair of companies, link analysis can be used as described in the next section.

10.4.4.2 Link Analysis

In this case study, link analysis has been applied to visually display the correlations between the competencies of companies in the form of keywords extracted by the text analysis from invitation to tender documents and further investigate this linkage to identify possible collaboration partners for business opportunities. In this case study, a set of rules are generated using the keywords identified in the text analysis process for each company. In this manner, a rule satisfies the requirement of a particular company. The motive is to identify the linkages between these rules in the tendering documents, so that complementary competencies can be identified for any particular project or a set of projects.

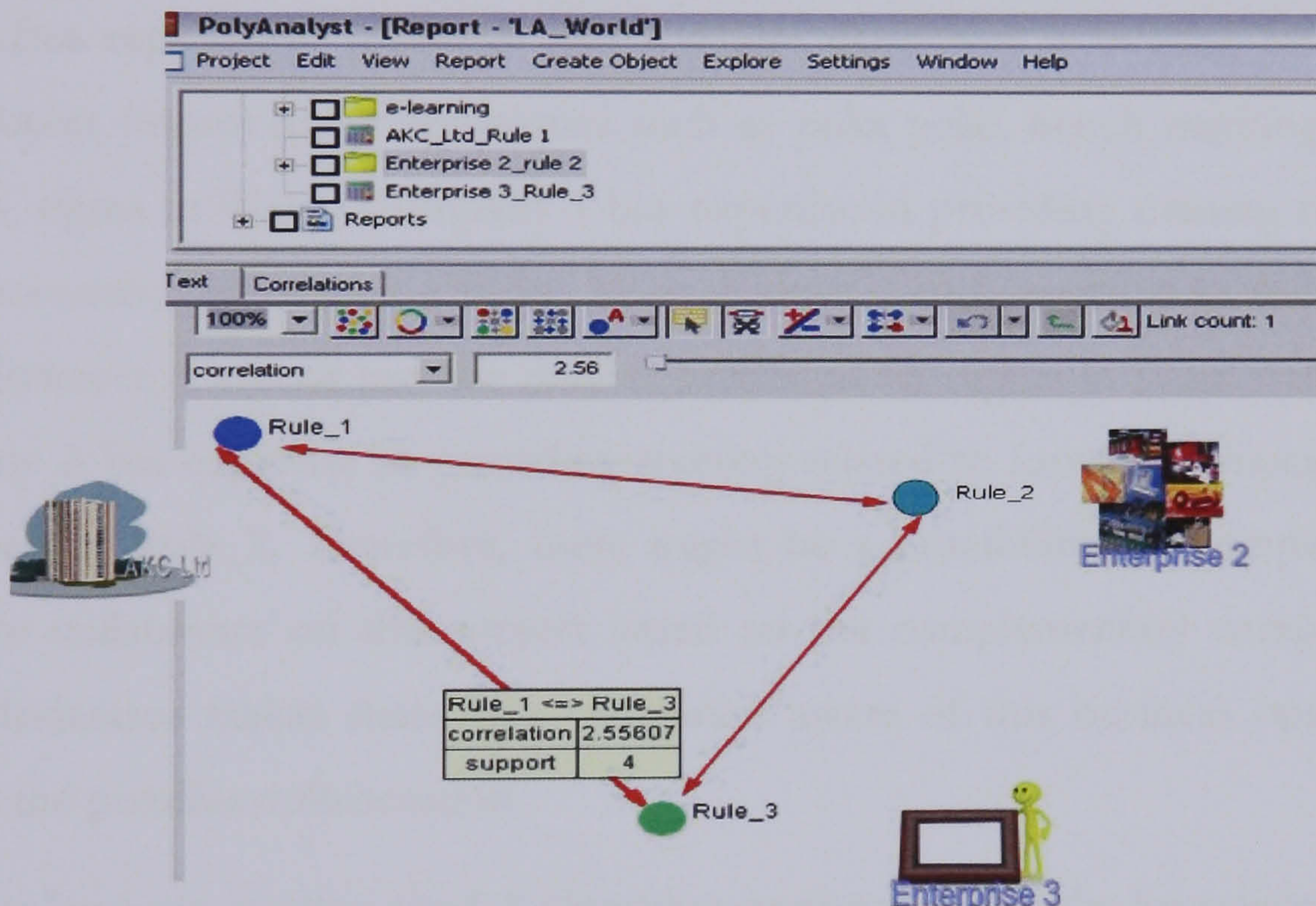


Figure 10-5: Linkages between the rules corresponding to competencies of the companies.

Figure 10-5 shows an example of the application of LA on a set of invitation to tender documents. LA has been applied on the rules developed for the individual company. Each colour code represents a different rule (a set of competencies and areas of interests) for a different company. The strength of link shows the correlation between rules. For example, Rule_1 which represents the competencies and area of interest of AKC Ltd is associated with Rule_3 which represents the competencies and area of interest of company 3. It indicates that four invitations to tender documents required the competencies or expertise from both the company 1(AKC Ltd) and Company 3. Using

PolyAnalyst, the cluster of reports which mentions the details of projects can be shown by clicking on the link. For example, one of the reports identified required 5 major key words /phrases, including web based design, software, education, Cardiff and web hosting. Rule_1 indicates that it has the capability to provide services relating to education related software design and web services; however, it does not meet the requirements of web hosting. One of the competencies mentioned in Rule_3 is web hosting. It means that Company 3 has an expertise in web hosting. In this manner it can be seen that, this project can be beneficial for both the company 1(AKC Ltd) and company 3. Based on the link, these sets of invitation to tender documents can be forwarded to company 1 and company 3 and to make them aware of the possible business opportunities.

Similarly, the correlation between company 1 and company 2 shows that there is a linkage between the competencies of both the companies on 1 of the invitation of tender document. After exploration it reveals that a tender document is looking for training related to process improvement techniques such as poka yoke, bench marking, kaizen, Lean and six- sigma in Wales. Company 1 has expertise in providing training related to process improvement areas such a Poka, Yoke, Kaizen, bench marking as indicated in the rule_1. However, it cannot provide training related to Six-Sigma or Lean. At the same time, company 2 has expertise in providing training related to Lean and Six-sigma and this is mapped in Rule_2. Therefore, there might be a possibility for company1 and company 2 to collaborate on this project based on the complementary competencies. Therefore, Moderator makes these two companies aware of this business opportunity and indicates the possible collaboration.

In this manner, one can see that the LA algorithm implemented in the knowledge miners of the Moderator can be used to raise the awareness of possible business opportunities and indication of collaboration between two companies. This approach however is limited to the collaboration between two companies only. However, there might be several situations when there is a need to collaborate with two or more companies to meet the demands of the project and competencies required. In order to overcome this limitation, the next section discusses the application of Dimensional matrix or Text OLAP as a way to establish collaboration between multiple enterprises.

10.4.4.3 Text OLAP (Dimensional matrix)

The Text OLAP feature of the knowledge miners can provide the Moderator with the capability to perform multi dimensional analysis of textual data in order to identify the possible business opportunity. In this case study, two different types of approaches are adopted to analyze the invitation to tender documents. In the first approach, the analysis is focussed on the individual company and the main aim is to identify business opportunities which are more likely to satisfy the criteria of an individual company for a new project. The second approach involves identification of business opportunities as well as the business partners for a specific project. The process of the application of dimensional matrix starts with the creation of a matrix. Each column of the matrix consists of different cells where each cell represents a specific competency or a rule comprising of competencies and areas of interest to be searched within the tender based data. While working with Text OLAP, the user defines the value of each cell and then browses the subset of records, belonging to the keyword or selected cell. The advantage of this approach lies in reusability of model. Once the matrix is created, it can be exported to new project and reused for future datasets.

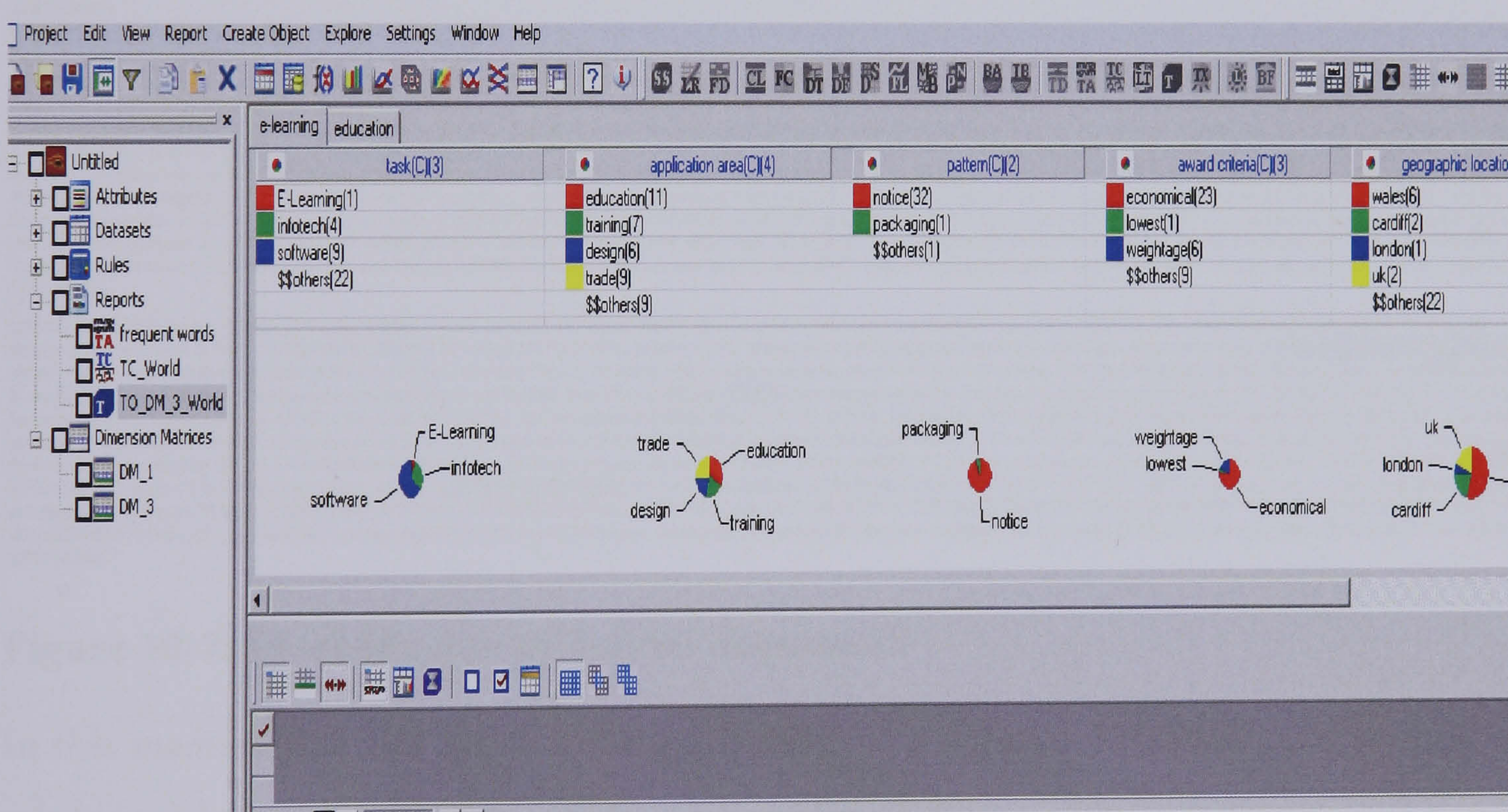
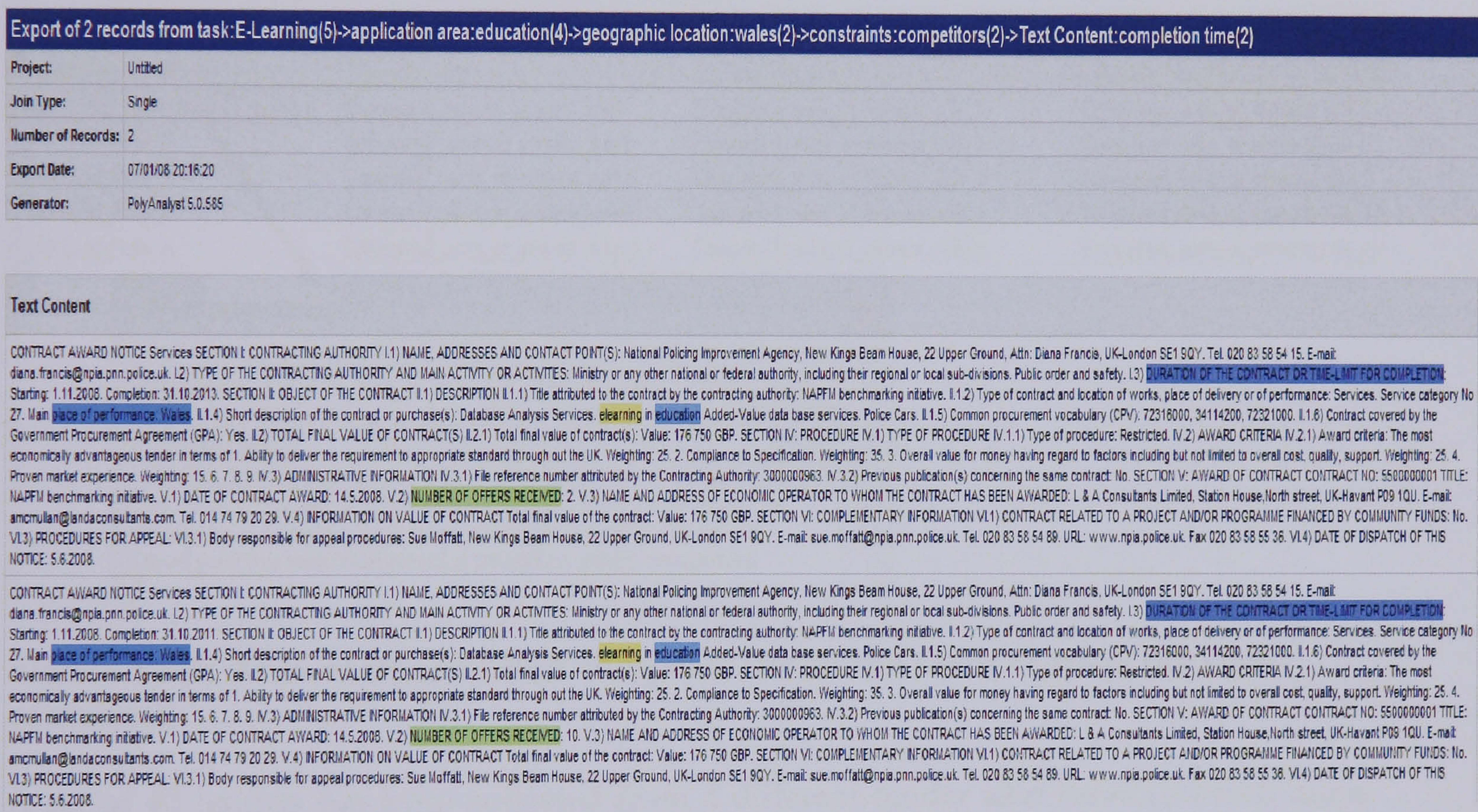


Figure 10-6: Dimensional matrix representing the competencies and areas of interest for AKC Ltd.

The dimensional matrix is pictorially shown in Figure 10-6. This dimensional matrix is created based on the competency and area of interest of the AKC Ltd. Each column of the dimensional matrix shows the area of interest of the company. As shown in the Figure 10-6, the matrix is divided into 5 different columns relating to task, application areas, patterns, award criteria, and geographical location of the project. Each column consists of multiple cells representing the area of interest of the company. It can be seen from the figure that in the category of geographical location the company is mainly interested in finding business opportunities in the nearby area of Wales, Cardiff or London. Therefore, search will mainly focus on the tender documents which are based in these areas only. The number in the parenthesis shows the number of tender documents where they appeared. However, one column is not sufficient to identify appropriate business opportunity. Therefore, the task column is combined with other keywords to identify a pattern of business opportunities. In this manner, E learning, education, geographical location Wales, competitors and completion time resulted in identifying 2 invitations to tender documents. These documents can be exported as .htm file and emailed to the user for further exploration as shown in Figure 10-7.



Export of 2 records from task:E-Learning(5)->application area:education(4)->geographic location:wales(2)->constraints:competitors(2)->Text Content:completion time(2)

Project:	Untitled
Join Type:	Single
Number of Records:	2
Export Date:	07/01/08 20:16:20
Generator:	PolyAnalyst 5.0.585

Text Content
CONTRACT AWARD NOTICE Services SECTION I: CONTRACTING AUTHORITY I.1) NAME, ADDRESSES AND CONTACT POINT(S): National Policing Improvement Agency, New Kings Beam House, 22 Upper Ground, Attn: Diana Francis, UK-London SE1 9QY. Tel. 020 83 58 54 15. E-mail: diana.francis@npia.pnn.police.uk. I.2) TYPE OF THE CONTRACTING AUTHORITY AND MAIN ACTIVITY OR ACTIVITIES: Ministry or any other national or federal authority, including their regional or local sub-divisions. Public order and safety. I.3) DURATION OF THE CONTRACT OR TIME-LIMIT FOR COMPLETION: Starting: 1.11.2008. Completion: 31.10.2013. SECTION II: OBJECT OF THE CONTRACT II.1) DESCRIPTION II.1.1) Title attributed to the contract by the contracting authority: NAPFM benchmarking initiative. II.1.2) Type of contract and location of works, place of delivery or of performance: Services. Service category No 27. Main place of performance: Wales. II.1.4) Short description of the contract or purchase(s): Database Analysis Services. elearning in education Added-Value data base services. Police Cars. II.1.5) Common procurement vocabulary (CPV): 72316000, 34114200, 72321000. II.1.6) Contract covered by the Government Procurement Agreement (GPA): Yes. II.2) TOTAL FINAL VALUE OF CONTRACT(S) II.2.1) Total final value of contract(s): Value: 176 750 GBP. SECTION IV: PROCEDURE IV.1) TYPE OF PROCEDURE IV.1.1) Type of procedure: Restricted. IV.2) AWARD CRITERIA IV.2.1) Award criteria: The most economically advantageous tender in terms of 1. Ability to deliver the requirement to appropriate standard through out the UK. Weighting: 25. 2. Compliance to Specification. Weighting: 35. 3. Overall value for money having regard to factors including but not limited to overall cost, quality, support. Weighting: 25. 4. Proven market experience. Weighting: 15. 6. 7. 8. 9. IV.3) ADMINISTRATIVE INFORMATION IV.3.1) File reference number attributed by the Contracting Authority: 3000000963. IV.3.2) Previous publication(s) concerning the same contract: No. SECTION V: AWARD OF CONTRACT CONTRACT NO: 5500000001 TITLE: NAPFM benchmarking initiative. V.1) DATE OF CONTRACT AWARD: 14.5.2008. V.2) NUMBER OF OFFERS RECEIVED: 2. V.3) NAME AND ADDRESS OF ECONOMIC OPERATOR TO WHOM THE CONTRACT HAS BEEN AWARDED: L & A Consultants Limited, Station House,North street, UK-Havant P09 1QU. E-mail: amcmullan@landaconsultants.com. Tel. 014 74 79 20 29. V.4) INFORMATION ON VALUE OF CONTRACT Total final value of the contract: Value: 176 750 GBP. SECTION VI: COMPLEMENTARY INFORMATION VI.1) CONTRACT RELATED TO A PROJECT AND/OR PROGRAMME FINANCED BY COMMUNITY FUNDS: No. VI.3) PROCEDURES FOR APPEAL: VI.3.1) Body responsible for appeal procedures: Sue Moffatt, New Kings Beam House, 22 Upper Ground, UK-London SE1 9QY. E-mail: sue.moffatt@npia.pnn.police.uk. Tel. 020 83 58 54 89. URL: www.npia.police.uk. Fax 020 83 58 55 36. VI.4) DATE OF DISPATCH OF THIS NOTICE: 5.6.2008.
CONTRACT AWARD NOTICE Services SECTION I: CONTRACTING AUTHORITY I.1) NAME, ADDRESSES AND CONTACT POINT(S): National Policing Improvement Agency, New Kings Beam House, 22 Upper Ground, Attn: Diana Francis, UK-London SE1 9QY. Tel. 020 83 58 54 15. E-mail: diana.francis@npia.pnn.police.uk. I.2) TYPE OF THE CONTRACTING AUTHORITY AND MAIN ACTIVITY OR ACTIVITIES: Ministry or any other national or federal authority, including their regional or local sub-divisions. Public order and safety. I.3) DURATION OF THE CONTRACT OR TIME-LIMIT FOR COMPLETION: Starting: 1.11.2008. Completion: 31.10.2011. SECTION II: OBJECT OF THE CONTRACT II.1) DESCRIPTION II.1.1) Title attributed to the contract by the contracting authority: NAPFM benchmarking initiative. II.1.2) Type of contract and location of works, place of delivery or of performance: Services. Service category No 27. Main place of performance: Wales. II.1.4) Short description of the contract or purchase(s): Database Analysis Services. elearning in education Added-Value data base services. Police Cars. II.1.5) Common procurement vocabulary (CPV): 72316000, 34114200, 72321000. II.1.6) Contract covered by the Government Procurement Agreement (GPA): Yes. II.2) TOTAL FINAL VALUE OF CONTRACT(S) II.2.1) Total final value of contract(s): Value: 176 750 GBP. SECTION IV: PROCEDURE IV.1) TYPE OF PROCEDURE IV.1.1) Type of procedure: Restricted. IV.2) AWARD CRITERIA IV.2.1) Award criteria: The most economically advantageous tender in terms of 1. Ability to deliver the requirement to appropriate standard through out the UK. Weighting: 25. 2. Compliance to Specification. Weighting: 35. 3. Overall value for money having regard to factors including but not limited to overall cost, quality, support. Weighting: 25. 4. Proven market experience. Weighting: 15. 6. 7. 8. 9. IV.3) ADMINISTRATIVE INFORMATION IV.3.1) File reference number attributed by the Contracting Authority: 3000000963. IV.3.2) Previous publication(s) concerning the same contract: No. SECTION V: AWARD OF CONTRACT CONTRACT NO: 5500000001 TITLE: NAPFM benchmarking initiative. V.1) DATE OF CONTRACT AWARD: 14.5.2008. V.2) NUMBER OF OFFERS RECEIVED: 10. V.3) NAME AND ADDRESS OF ECONOMIC OPERATOR TO WHOM THE CONTRACT HAS BEEN AWARDED: L & A Consultants Limited, Station House,North street, UK-Havant P09 1QU. E-mail: amcmullan@landaconsultants.com. Tel. 014 74 79 20 29. V.4) INFORMATION ON VALUE OF CONTRACT Total final value of the contract: Value: 176 750 GBP. SECTION VI: COMPLEMENTARY INFORMATION VI.1) CONTRACT RELATED TO A PROJECT AND/OR PROGRAMME FINANCED BY COMMUNITY FUNDS: No. VI.3) PROCEDURES FOR APPEAL: VI.3.1) Body responsible for appeal procedures: Sue Moffatt, New Kings Beam House, 22 Upper Ground, UK-London SE1 9QY. E-mail: sue.moffatt@npia.pnn.police.uk. Tel. 020 83 58 54 89. URL: www.npia.police.uk. Fax 020 83 58 55 36. VI.4) DATE OF DISPATCH OF THIS NOTICE: 5.6.2008.

Figure 10-7: Identification of Tender documents

In this manner, one can see that the application of a dimensional matrix for knowledge miners can be used to identify and raise awareness of the possible business opportunities.

10.4.4.4 Dimensional matrix for possible partner selection

This approach involves the development of a dimensional matrix in such a manner that each column represents a single company and each cell of that column represents the area of interest or competencies in the form of rules as shown in Figure 10-8. This will facilitate the search process of tenders across multiple companies' areas of interests.

Selection of a particular rule in one column will highlight the other companies' cell where the complementary competencies are found to meet the requirements of the project. This will indicate which partners should collaborate to meet the demands of the project. This concept has been demonstrated using an illustrative example of SMEs. Let us again consider the scenario mentioned in 10.2 about company AKC Ltd and their network XYZ Net. In practical terms, there can be "n" number of members in the collaborative pool as long as the members agree with the terms and conditions with the network. However, to illustrate the concept, this example considers 4 companies including AKC Ltd as a member of the network. These companies are willing to share their areas of interest and competencies which are stored in the shared info. Table 10-1 shows the areas of interests related to these companies.

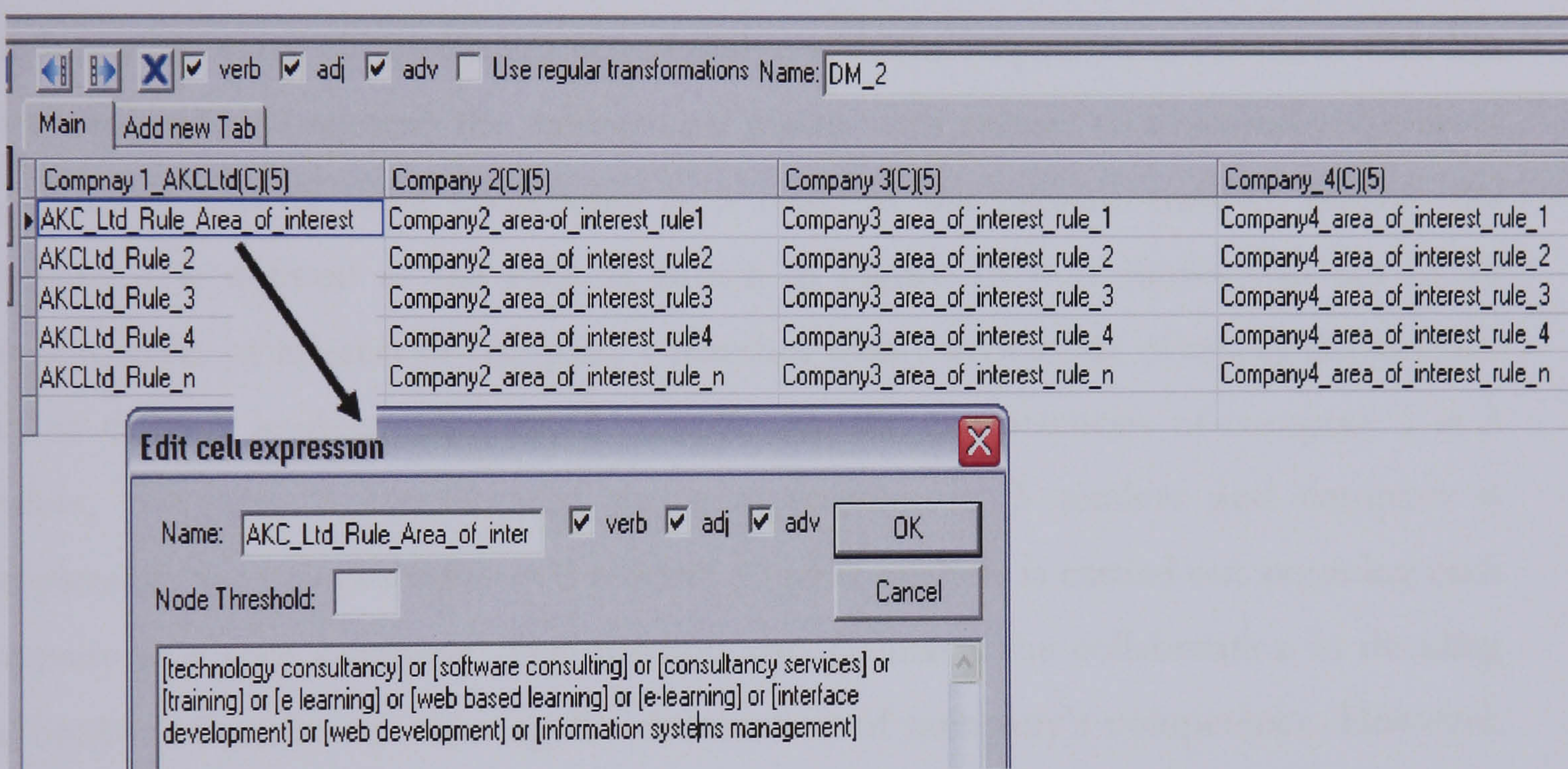


Figure 10-8: Area of interest and competencies in the form of rules for each company.

A set of 40 tenders has been collected from 3 different tender alert services which quote these tenders under the area of information technology. Based on the domain expertise and developed ontology, a set of rules has been created for each company as shown in Figure 10-8. The created dimensional matrix is applied on the dataset in the PolyAnalyst software with a view to analyze the tender based data.

The application of the dimensional matrix shows that the area of interest of company 1 (AKC Ltd) appeared in 7 of the invitation to tender documents out of 40. Similarly for company 2, 3 and 4 their area of interest or competencies appeared 4, 8 and 3 times respectively.

Table 10-1 Areas of interest of each company

Company 1	Company 2	Company 3	Company 4
Technology Consultancy, Software Consulting, Consultancy services, Training services, IT Training, E learning, web based learning, Interface development, Web development, Information Systems management etc.	Platform Integration, Integration support, Software integration, System integration, Service oriented architecture, ERP, Business to business solution for Supply chain etc.	Software Maintenance, IT Services Maintenance, IT Maintenance, Software Development, Application Development, IT Implementation, Medical Software package etc.	Hardware maintenance, supply of Laptops, computers, printers and hardware, web hosting, IT Monitor services etc.

However, at this point it is the knowledge identified is only useful for each individual company. It might be possible that among these tenders there will be some tenders which require extra competencies not available with the respective companies. This has been achieved by analyzing the dimensional matrix with respect to a particular company i.e. defining any of these companies as a root. In the present context, initially the company 1 is defined as the root as shown in Figure 10-9. It shows that out of 40 Tenders, areas of interest of company 1 matches with 7 invitations of tender documents. Out of those 7 tenders, company 2 complement the competencies of company 1 in 3 tenders, company 3 complement the competencies in 3 tenders and company 4 complement the competencies in 2 tenders. Further analysis is carried out assuming each company as a root company. This method also facilitates the collaboration in deciding the broker of tender by showing the domination of company's competence. However, practically it can only be decided during the next stage of tendering process therefore not the scope of this research.

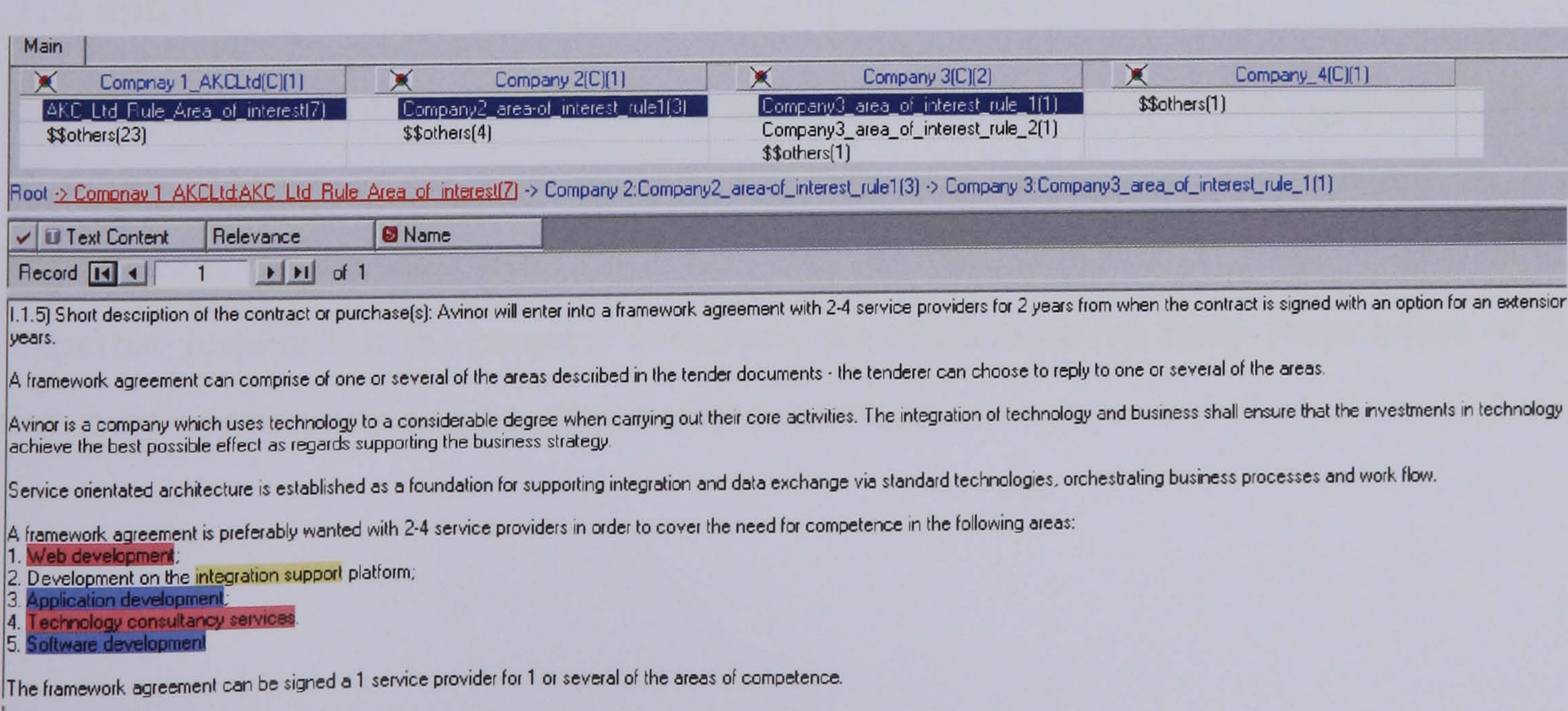


Figure 10-9: Indication of suitable tenders and partners for company 1, 2 and 3.

Figure 10-9 shows a snapshot of the application of the dimensional matrix where company 1 has been chosen as root. This indicates a possible tender for company1 with suitable collaborator as company 2 and company 3. In the body of text of the invitation of tender documents, it can be seen that the project requires 5 different types of expertise. Among these company 1 has expertise in technology consulting services and web development. It will not be possible for company 1 to handle this project due to shortfall of expertise. Company 2 provides services in the area of integration support platform and alone company 2 cannot handle the project. Similarly, company 3 can provide services related to software development and application development however cannot handle the project alone. As shown in the figure, there is a possibility to establish collaboration between company 1, 2 and 3 to meet the requirements of the project by providing all 5 competencies together. Knowledge miners can therefore pass this information to the Moderator to raise the awareness of possible business opportunities and possible partners to company 1, 2 and 3.

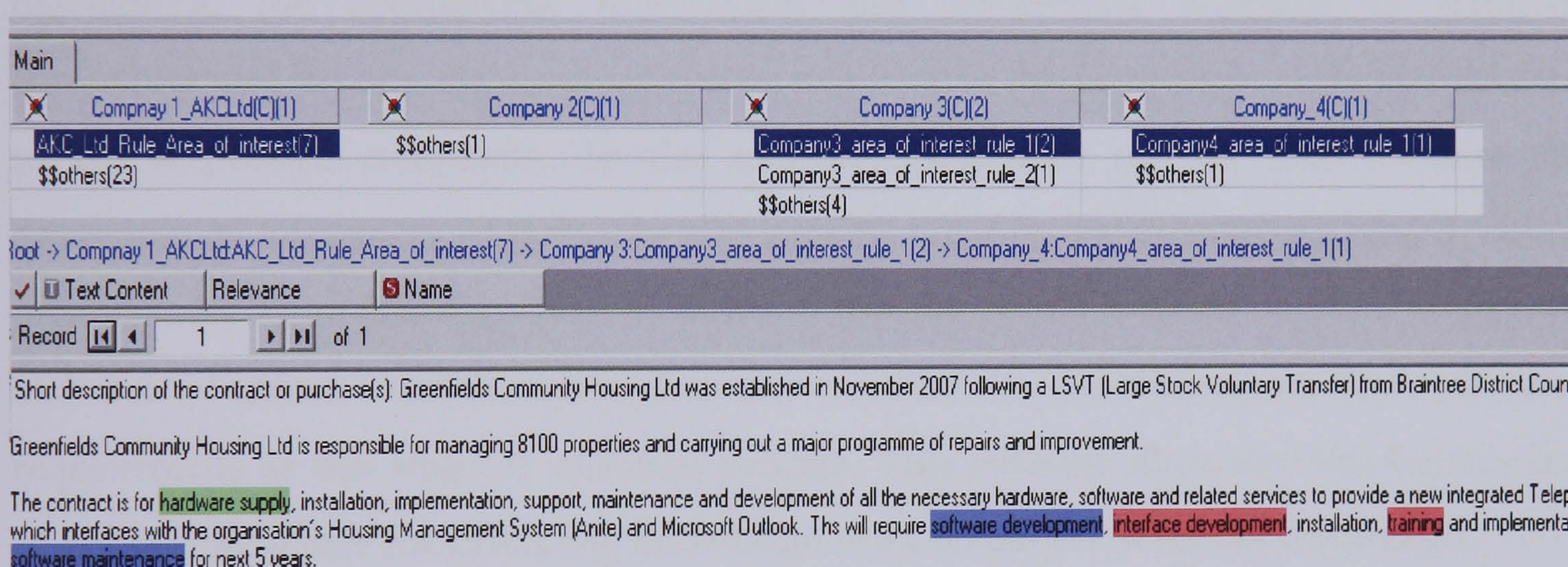


Figure 10-10: Possible business opportunity and suitable partners for company 1, 2 and 4.

Similarly Figure 10-10 indicates the possible business opportunities for company 1, 2 and 4, where alone they can only partially fulfil the requirement of project based on their expertise. It can be seen from the figure that together company 1, 2 and 4 meet all the expertise required of the project. Therefore, the Moderator can make them aware of this business opportunity and possible business partners.

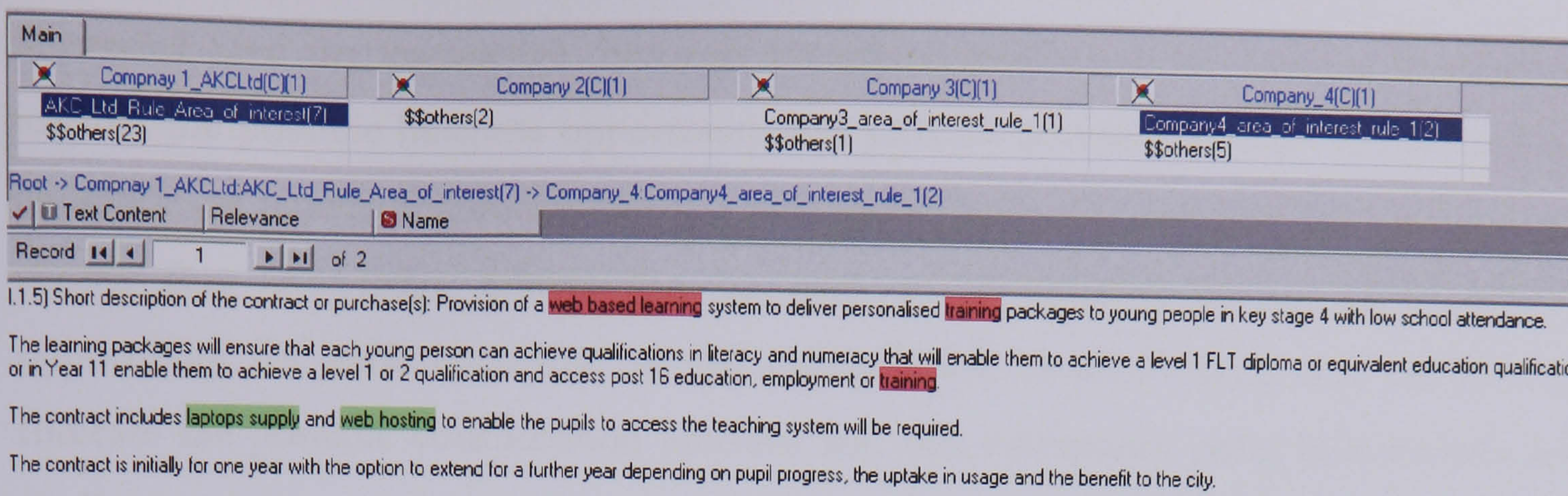


Figure 10-11: Possible business opportunity and suitable partners for company 1 and 4.

Figure 10-11 shows the possible business opportunity and partnership for company 1 and company 4. Similarly, Figure 10-12 shows the possible business opportunity for company 3 and company 4. In this case, company 4 has been chosen as the root. It indicates that company 4 can collaborate with company 3 on two projects and with company 1 on one project as shown in the parentheses. All the possible tenders highlighting the competence requirement and competence supplied by different companies and list of collaborations can be exported in the form of .htm file and emailed to the respective companies. In this manner, Moderator can be used to raise the awareness of a business opportunity and suitable business partners to establish a virtual organization. Here it is important to mention that the search is based on the competence and area of interest and therefore does not consider the financial constraints and other requirements at this stage of selection process. This example illustrated the functioning of knowledge miner for extracting knowledge from invitation of tender documents.

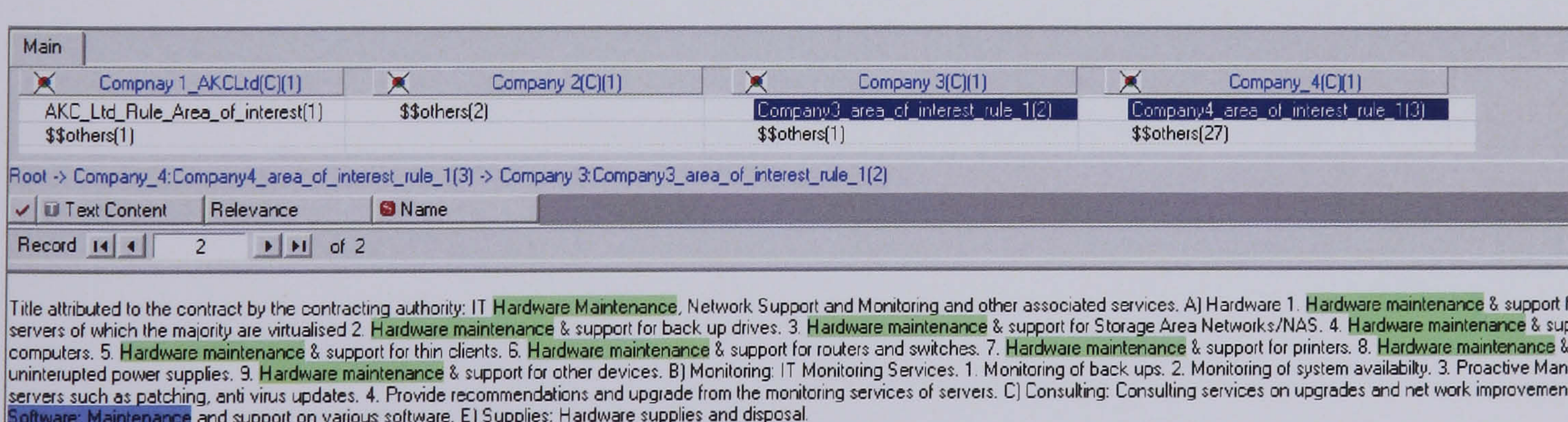


Figure 10-12: Possible business opportunity and suitable partners for company 3 and 4.

10.5 Conclusion from Tendering Case Study.

One of the major challenges faced by SMEs working in a collaborative environment is identifying appropriate business opportunities and simultaneously determining the possible partners to fulfil the requirements of the business opportunity. This chapter illustrated how the knowledge discovery integrated moderator services can be used to identify the possible business opportunity and possible partners to meet the demand of invitation of tender documents using a case study from UK based SMEs. It has been shown that knowledge miners can be used to (1) identify the possible business opportunity for each enterprise in the collaborative network using text analysis (2) Indicate the possible collaboration between any two enterprises using link analysis and (3) Raise awareness of enterprises by indicating the possible business opportunities and possible business partners for multi-enterprise collaboration using dimensional matrix.

11

Conclusions

This chapter summarizes this thesis and discusses the key aspects of the proposed research. This outlines the novelty and contributions of this research and recommends direction for future research.

Chapter Outline

11.1 Conclusions and Discussion

Velocity, visibility, scalability, innovation and cost govern competitive advantage for organizations viewing the entire world as their market. Knowledge provides power in many of these contexts enabling and facilitating the preservation of valuable heritage, learning new things, solving intricate problems, and creating core competencies. Therefore, knowledge needs to be acquired, extracted and integrated into the system to improve its performance. The use of Moderators to support collaborative projects offers the opportunity for relevant knowledge held by individual partners to be shared in moderating the behaviour of the collaborative projects. The quality of the support that any Moderator can provide is limited by the knowledge about the project's partners as collected by the KAM and stored in the EMs. The question addressed by this research was "how to enhance the Moderator's functionality and capability by integrating knowledge discovery based semi-automated knowledge acquisition thereby enabling moderators to "learn" and "update" their relevant expert modules from knowledge discovered in the existing operational databases of companies?".

Knowledge discovery and data mining has been identified as a way of addressing the needs for regular update of knowledge and ongoing learning for Moderator. The literature review undertaken showed an increasing use of data mining to solve knowledge acquisition problems in a wide variety of industrial contexts. However, this review revealed that there is a need to fully exploit, reuse and integrate the discovered knowledge with the existing knowledge based systems to realize the true benefit of KDD. This review also indicated the scope of data mining in the areas related to collaborative projects, supply chain and virtual enterprises since projects also record information in the form of text (e.g. project reports and minutes). Knowledge discovery in Text and Text mining have been studied with a view to extracting knowledge from unstructured text based data.

A KOATING framework with data mining capability has been proposed for Moderators for semi-automate the process of knowledge acquisition. This approach provided a mechanism for integration of data mining system with the existing Moderators and reuses the knowledge stored in expert module for moderation process. In addition, the KOATING framework has been integrated with state-of-art Moderators called Universal Knowledge Moderator. As substantial research had been carried out in to UKM in earlier projects, this research has been exploited to test and demonstrate the integration of

KOATING framework. An illustrative example showed the functioning of the proposed KOATING framework in the context of e- supply chain. It has been shown that KOATING framework is capable of embedding the updating of knowledge and learning capability in the Moderator system.

To design, develop and document the proposed system, Unified Modelling Language has been used in chapter 8. UML has been considered as a de facto standard for modelling software applications and it visually capture the design of software. UML has been employed to capture the static structure and dynamic behaviour of the system. Six different types of visual diagrams have been detailed to capture the system analysis, system design and system development aspects of the proposed KOATING framework. This enabled the design of the framework to be clarified and greater detail added to the design. This increased the usefulness and applicability of this research. To demonstrate the proof of the proposed KOATING framework, a case study related to post project reports of construction project supply chain has been presented. It has been shown that KOATING framework is capable of extracting different types of knowledge from unstructured text based PPR related data to update the expert module of construction project moderator.

Knowledge in the expert module must be accurate and therefore a semi-automated approach is adopted instead of fully automated system. It is important to emphasize that due to the wide variety and complexity of different types of data in a wide range of contexts that knowledge miner must deal with it is highly unlikely that knowledge miner of KOATING framework could operate in a fully automatic manner in near future. In addition, domain expertise and human interaction will be needed to guide the process of knowledge evaluation, selection and verification in the context of Moderator.

It has also been shown using case study to the SME of the UK that KOATING framework can be used to enhance Moderators awareness capability by raising awareness of possible business opportunity and business partners in order to establish a virtual organization. In this manner, this research enables a Moderator to learn and continuously update its knowledge in the expert module from various sources of operational or past data using knowledge discovery and data mining technology. In addition, proposition of knowledge miners enhances the Moderators capability by raising awareness of business opportunity and possible partners.

11.2 Novelty and Contributions

This research looked into tasks and technologies needed to support the Moderator's knowledge acquisition process. This research is novelty and major contributions are summarized as follows:

- ***Automated method of identifying the research gaps***: The novelty of the research started with literature review of this research. A novel text mining approach has been adopted to identify the research gaps from 150 literatures in the domain of data mining applications in manufacturing. This is illustrated in section 5.3 of chapter 5. A paper based on the findings has been published and attached as appendix 2.

- ***Integrating the knowledge discovery and data mining based framework***: A knowledge discovery and data mining framework has been developed and integrated with the Moderator system. Proposed framework enabled the Moderator to semi-automatically update the expert module's knowledge and learn. The detailed discussion is presented in chapter 7. A part of the proposed approach is published in [62, 157, 158] and their abstracts are attached as appendix 7, 8 and 9.

- ***Knowledge discovery for virtual-e-supply chain***: An illustrative example is presented in section 7.6 of chapter 7, where it has been shown how knowledge can be extracted from numerical data relating to suppliers in a virtual–e chain environment.. This research is novel in a sense that using an example a proposition is made that how knowledge can be discovered from Virtual–e–supply chain to support the Moderator's requirement and function.

- ***UML for Semi-Automated Knowledge Acquisition system development***: According to the best of author's knowledge this research is first of its kind where UML has been used to model and develop a semi-automated knowledge acquisition system specifically in the context of Moderator technology. It is presented in chapter 8 of this thesis.

- ***Capturing Tacit Knowledge into Implicit form and Applying Text Mining on PPRs to Discover Knowledge***: In this research, PPRs have been used as a means to capture tacit knowledge in its implicit form. Furthermore, various text mining technologies have been used to update the Moderator's knowledge in the expert module. A detailed study of knowledge discovery from PPRs is discussed in chapter

9 of this thesis. This research has also been published in [175, 176, 181] and their abstracts are attached as appendix 3,4 and5.

- ***Enhancing the Capability of Moderators by Identifying Possible Business Opportunities and Partners using Text Mining*.** This research proposes that the knowledge miners can be used to identify the possible business opportunity from invitation of tender documents floating through various sources. Based on the area of interest and competency in the shared info, it also identifies possible business partners to form a VO. This is illustrated in chapter 10 using a case study of UK based SME.

11.3 Recommendation for Future Work

. This future scope of this research is outlined as follows:

- Knowledge discovery on semantic web for automated knowledge creation is further challenge for Moderators research. Integration of Ontological engineering with data mining for ontology based information extraction and ontology learning in context of virtual enterprise and e-SCM is a major challenge. Current UKM research partially addresses this challenge; however, a considerable effort is required in this research area.
- Semantic Web technologies and tools require considerable technical expertise, and are thus not well suited for users outside the field of computer science. This makes it hard for domain experts and ontology engineers to work together on e-manufacturing tasks. Another challenge for the Moderators research is to enhance and improve the ease of interaction and operation between participants and thereby enable mass collaboration and knowledge sharing. The rapid rise in popularity of the Semantic Wiki mechanisms, shows co-creating corporate knowledge for partners and clients via the Extranet has becomes the main challenge to realizing the vision of seamless business interaction across the boundaries of e-manufacturing chains. These topics are therefore recommended for future investigation.
- World Wide Web (WWW) is the great source of information extraction. Web mining is an important tool to deal with the discovery of useful, novel and crucial information on web. Web mining will enhance the knowledge acquisition

capability of moderator technology. Therefore, application of web mining in context of Moderator will be another area of future research.

- To date, all the applications of Moderators are limited to Manufacturing or related areas. The following are unconnected with manufacturing, but still considered to have potential moderator applications and therefore should be examined as future research areas:
  - Scheduling applications (e.g. Air traffic control, railway timetabling (planning and operations)
  - Agent moderation in autonomous agent-based systems.
  - Pharmacology (e.g. Identifying conflicts in medication)

References

[1] J. L. M. Lastra and M. Delamer, "Semantic web services in factory automation: fundamental insights and research roadmap," *Industrial Informatics, IEEE Transactions on,* vol. 2, pp. 1-11, 2006.

[2] ManuFuture 2005, "Strategic Research Agenda", http://www.manufutureuk2005.org.uk/ last updated 21st May 2009.

[3] N. Ikujiro and H. Takeuchi, *The Knowledge-Creating Company.* New York: Oxford University Press, 1995,

[4] MISSION Consortium, "Final report from MISSION Project—Deliverable D24.K," in ***M**odelling and **S**imulation Environments for De**Sign**, Planning and Operation of Gl**o**bally Distributed E**n**terprises (IMS/ESPRIT29656)* K. Popplewell, J. A. Harding and M. Rabe, Eds. 2001,

[5] MOSES. (1995, Model oriented simultaneous engineering system (EPSRC project number GR/H24273). *2008(7/2),*

[6] H. K. Lin, "Manufacturing System Engineering Ontology Model for Global Extended Project Team" PhD Thesis, Loughborough University, 2004.

[7] T. Skjoett-Larsen, C.Thernoe and C. Andresen, "Supply chain collaboration: Theoretical perspectives and empirical evidence," *International Journal of Physical Distribution & Logistics Management,* vol. 33, pp. 531-549(19), 18 July 2003.

[8] B. R. Katzy and M. Dissel, "A toolset for building the virtual enterprise," *J. Intell. Manuf.,* vol. 12, pp. 121-131, 04/01. 2001.

[9] K. Popplewell and J. A. Harding, "Impact of simulation and moderation through the virtual enterprise life cycle," in *Proceedings of the 11th ASIM Dedicated Conference,* 2004, pp. 299.

[10] E. Y. Li and H. Lai, "Collaborative work and knowledge management in electronic business," *Decision Support Systems,* vol. 39, pp. 545-547, 6. 2005.

[11] B. R. Gaines, D. H. Norrie and A. Z. Lapsley, "Mediator: an intelligent information system supporting the virtual manufacturing enterprise," *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century.* , *IEEE International Conference on,* vol. 1, pp. 964-969 vol.1, 1995.

[12] J. A. Harding and K. Popplewell, "Driving concurrency in a distributed concurrent engineering project team: a specification for an Engineering Moderator," *International Journal of Production Research,* vol. 34, pp. 841, 1996.

[13] J. Browne and J. Zhang, "Extended and virtual enterprise-similarities and differences," *International Journal of Agile Management Systems,* vol. 1, pp. 30-36, 1999.

[14] Frécon, E. and Nöu, A. A., "Building distributed virtual environments to support collaborative work," in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology,* 1998, pp. 105-113.

[15] L. Zhou and R. Nagi, "Design of distributed information systems for agile manufacturing virtual enterprises using CORBA and STEP standards," *Journal of Manufacturing Systems,* vol. 21, pp. 14-31, 2002.

[16] A. J. Slade and A. F. Bokma, "Ontologies within extended enterprises," *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on,* pp. 541-550, 2002.

[17] F. Shafiei and D. Sundaram, "Multi-enterprise collaborative enterprise resource planning and decision support systems," *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on,* pp. 10 pp., 2004.

[18] N. Panteli and S. Sockalingam, "Trust and conflict within virtual inter-organizational alliances: a framework for facilitating knowledge sharing," *Decision Support Systems,* vol. 39, pp. 599-617, 6. 2005.

[19] H. J. Ahn, H. J. Lee, K. Cho and S. J. Park, "Utilizing knowledge context in virtual collaborative work," *Decision Support Systems,* vol. 39, pp. 563-582, 6. 2005.

[20] I. Cil, O. Alpturk and H. R. Yazgan, "A new collaborative system framework based on a multiple perspective approach: InteliTeam," *Decision Support Systems,* vol. 39, pp. 619-641, 6. 2005.

[21] S. Misono, S. Koide, N. Shimada, M. Kawamura and S. Nagano, "Distributed collaborative decision support system for rocket launch operation," *Advanced Intelligent Mechatronics. Proceedings, 2005 IEEE/ASME International Conference on,* pp. 1318-1323, 2005.

[22] S. B. Shum, D. D. Roure, M. Eisenstadt, N. Shadbolt and A. Tate, "CoAKTinG: Collaborative advanced technologies in the grid," in *Proc. Second Workshop on Advanced Collaborative Environments,* 2002, pp. 1-8.

[23] B. Ramesh and A. Tiwana, "Supporting Collaborative Process Knowledge Management in New Product Development Teams," *Decision Support Systems,* vol. 27, pp. 213-235, 11. 1999.

[24] G. Q. Huang, J. Huang and K. L. Mak , "Agent-based workflow management in collaborative product development on the Internet," *Computer-Aided Design,* vol. 32, pp. 133-144, 2/1. 2000.

[25] M. Ulieru, D. Norrie, R. Kremer and W. Shen, "A multi-resolution collaborative architecture for web-centric global manufacturing," *Information Sciences,* vol. 127, pp. 3-21, 8. 2000.

[26] K. Rodriguez and A. Al-Ashaab, "Knowledge web-based system architecture for collaborative product development" *Computers in Industry.,* vol. 56, issue 1, pp. 125-140, 2005.

[27] B. Scotney and S. McClean, "Knowledge discovery from databases on the semantic Web," *Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on,* pp. 333-336, 2004.

[28] Yue-jie Wen, Ling Tian and Bing-shu Tong, "Knowledge discovery based on Web robots in collaborative design," *Computer and Information Technology, 2005. CIT 2005. the Fifth International Conference on,* pp. 1040-1044, 2005.

[29] J. Numata, K. Hane, Bangyu Lei and Y. Iwashita, "Knowledge discovery and sharing in an information system," *Innovation in Technology Management - the Key to Global Leadership. PICMET '97: Portland International Conference on Management and Technology,* pp. 713-716, 1997.

[30] ECOLEAD. (2008, June 25, 2008). European collaborative networked organisations leadership initiative. *2008(7/14),*

[31] ECOLEAD, "ECOLEAD project public deliverable 21.1: Characterization of key components, features, and operating principles of the virtual breeding environment, " 2005.

[32] L. Camarinha-Matos and H. Afsarmanesh, "A comprehensive modeling framework for collaborative networked organizations," *J. Intell. Manuf.,* vol. 18, pp. 529-542, 10/01. 2007.

[33] M. Pechoucek, J. Vokrinek and P. Becvar, "ExPlanTech: multiagent support for manufacturing decision making," *Intelligent Systems, IEEE,* vol. 20, pp. 67-74, 2005.

[34] L. M. Camarinha-Matos, "Results of ECOLEAD: D11.4b, Set of Publications (Period Apr 04 - Sep 06)," http://www.ve-forum.org/default.asp?obj=B3-P=284&P=284&L=2#achieved , last accesses Jan 2009.

[35] W. D. Li and Z. M. Qiu, "State-of-the-art technologies and methodologies for collaborative product development systems," *Int J Prod Res,* vol. 44, pp. 2525, 2006.

[36] J. A. Harding, K. Popplewell and H. K. Lin, "A generation of moderators from single product to global E-supply," in *Knowledge and Technology Management in Virtual Organizations: Issues, Trends, Opportunities and Solutions* G. D. Putnik and M. M. Cunha, Eds. Portugal: IGI Publishing, 2007, pp. Chapter 5.

[37] K. Popplewell and J. A. Harding, "Engineering Moderation - Supporting Concurrency in Engineering Using Hybrid Knowledge Representation", in *Proceedings of IFIP Working Group 5 - 7 Working Conference on Managing Concurrent Manufacturing to Improve Industrial Performance, International Federation for Information Processing, Managing Concurrent Manufacturing to Improve Industrial Performance*, Seattle, U.S.A., 1995, pp 347-357.

[38] F. L. Krause, F. Kimura, T. Kjellberg and S. C. Y. Lu, "Product Modelling," *Annals of CIRP,* vol. 42, pp. 695-706, 1993.

[39] A. McKay, M. S. Bloor and A. de Pennington, "A framework for product data," *IEEE Trans. Knowled. Data Eng.,* vol. 8, pp. 825-838, 1996.

[40] A. Molina and R. Bell, "A Manufacturing Model representation of a flexible manufacturing facility. " *Journal of Engineering Manufacture: Proceedings of the Institution of Mechanical Engineers,* vol. 213, pp. 225-246, 1999.

[41] J. A. Harding and B. Yu, "Information-centred enterprise design supported by a factory data model and data warehousing," *Computers in Industry,* vol. 40, pp. 23-36, 9. 1999.

[42] J. A. Harding and K. Popplewell, "Information modelling: an integration of views of a manufacturing enterprise," *International Journal of Production Research,* vol. 37, pp. 2777, 1999.

[43] J. A. Harding, "A knowledge representation model to support concurrent engineering teamwork." *PhD Thesis, Loughborough University,* 1996.

[44] R. Knaus and C. Jay, "Transporting knowledge bases: a standard," *AI Expert,* vol. 5, pp. 34-39, November 1990. 1990.

[45] J. Arroyo, B. Eberhard, M. Fischer, G. García de Gurtubai, J. A. Harding, F. Jaekel, J. Morazo, K. Nonnenmacher, K. Popplewell, M. Rabe and D. Weinelt. (2001, Final report from mission project, deliverable D24.K. http://www.ims-mission.de,

[46] X. Zhang, J. Edwards, and J. Harding, "Personalised online sales using web usage data mining", *Comput. Ind.* vol 58, issue 8-9, pp 772-782, 2007

[47] Esprit Project, "CIMOSA: Open system architecture for CIM vol 1," Springer-Verlag, Berlin, 1993.

[48] A. W. Scheer, *Architecture of Integrated Information Systems: Foundations of Enterprise Modelling.* Berlin: Springer-Verlag, 1992,

[49] D. Delen, "A holistic approach to manufacturing systems modelling" *International Journal of Simulation and Process Modelling,* vol. 5, no.1 pp. 54, 2009.

[50] B. Yu, J. A. Harding and K. Popplewell, "Supporting enterprise design through multiple views," *International Journal of Agile Management Systems,* vol. 2, pp. 71-83, 2000.

[51] J. A. Harding, K. Popplewell and D. Cook, "Manufacturing systems engineering moderator: an aid for multidiscipline project teams." *Int. J. Prod. Res.,* vol. 41, pp. 1973-1986, 2003.

[52] K. Popplewell and J. A. Harding, "A manufacturing systems engineering moderator: Identifying conflicts in the design of distributed manufacturing system," in *The 4Th International EUROSIM: Shaping Future with Simulation,* 2001,

[53] N. M. Sadeh, D. W. Hildum and D. Kjenstad. "Agent-based e-Supply Chain Decision Support", *Journal of Organizational Computing and Electronic Commerce,* Vol. 13, No. 3, 2003.

[54] H. K. Lin and J. A. Harding, "An ontology driven manufacturing system engineering moderator for global virtual enterprise teams," in *Advances in Manufacturing Technology XVI. Proceedings of the International Conference on Manufacturing Research,* 2003, pp. 365-370.

[55] H. K. Lin, J. A. Harding and M. Shahbaz, "Manufacturing system engineering ontology for semantic interoperability across extended project teams," *Int. J. Production. Research,* vol. 42, pp. 5099-5118, 2004.

[56] H. K. Lin and J. A. Harding, "A manufacturing system engineering ontology model on the semantic web for inter-enterprise collaboration," *Computers in Industry,* vol. 58, pp. 428-437, 6. 2007.

[57] H. K. Lin, J. A. Harding and P. C. Teoh, "An inter-enterprise semantic web system to support information autonomy and conflict moderation," *Proc. Inst. Mech. Eng. Pt. B: J. Eng. Manuf.,* vol. 219, pp. 903-911, 04/01. 2005.

[58] A.R. Omar, J.A. Harding and K. Popplewell, "Design for Customer Satisfaction: An Information Modelling Approach", *Integrated Manufacturing Systems: The International Journal of Manufacturing Technology Management,* 10(4), 1999, pp 199-209.

[59] K. Kosanke, J. Roland, J. G. Nell and A. O. Bas, *Enterprise Inter- and Intra-Organizational Integration: Building International Consensus (International Federation for Information Processing).* Massachusetts, USA: Kluwer Academic Publishers, 2003,

[60] J. M. Dorador and R. I. M. Young, "Application of IDEF0, IDEF3 and UML methodologies in the creation of information models," *International Journal of Computer Integrated Manufacturing,* vol. 13, pp. 430, 2000.

[61] SYNERGY Project, http://www.synergy-ist.eu/, 2009, Last accessed 19th May 2009.

[62] A. K. Choudhary, J. A. Harding and K. Popplewell, "Knowledge discovery for moderating collaborative projects," in *IEEE International Conference on Industrial Informatics,* 2006, pp. 519-524.

[63] S. Liao, "Knowledge management technologies and applications—literature review from 1995 to 2002," *Expert Syst. Appl.*, vol. 25, pp. 155-164, 2003.

[64] M. Polanyi, *The Tacit Dimension.* London: Routledge and Kegan Paul, 1996,

[65] J. Liebowitz, *Knowledge Management Handbook.* New York, USA: CRC Press, 1999,

[66] E. Turban, *Expert System and Applied Artificial Intellignce.* Macmillan, 1992,

[67] J. Sowa, *Conceptual Structures.* Addision Wesley, 1984,

[68] K. Wiig, *Knowledge Management Foundation.* Schma Press, 1993,

[69] R. Van der spek and A. Spijkervet, "Knowledge management: Dealing intelligently with knowledge," in *Knowledge Management and its Integrative Elements* Liebowitz & Wilcox, Ed. CRC Press, 1997,

[70] T. Beckman, "A methodology for knowledge management," in *International Association for Development (IASTED, AI and Soft Computing Conference,* 1997,

[71] I. Spiegler, "Technology and knowledge: bridging a "generating" gap," *Information & Management,* vol. 40, pp. 533-539, 2003.

[72] A. Aamodt and M. Nygard, "Different roles and mutual dependencies of data, information, and knowledge - an AI perspective on their integration," *Data and Knowledge Engineering,* vol. 16, pp. 191-222, 1995.

[73] J. C. Spender, "Competitive advantage from tacit knowledge? Unpacking the concept and its strategic implications," *Academy of Management Best Papers Proceedings 1993,* pp. 37-41, 1993.

[74] K. H. Vat, "Toward an Actionable Framework of Knowledge Synthesis in the Pursuit of Learning Organization," *Information Science,* pp. 1085-1100, 2003.

[75] F. Kappe, "Hyperwave white paper - version 1.2," Hyperwave Research and Development Centre, Graz, Austria, 1999.

[76] I. Rus, M. Lindvall and S. S. Sinha, "Knowledge management in software engineering, DACS state-of-the-art-report," Data and Analysis Center for Software, 2001.

[77] G. Fischer and J. Ostwald, "Knowledge management: Problems, promises, realities, and challenges," *IEEE Intell. Syst.*, vol. 16, pp. 60-72, 2001.

[78] K. M. Wiig, "What future knowledge management users may expect," *J. Knowl. Manag.*, vol. 3, pp. 155-165, 1999.

[79] R. Dieng, O. Corby, A. Giboin. and M. Ribière "Methods and tools for corporate knowledge management" *J. Hum. -Comput. Stud.*, vol. 51, pp. 567-598, 1999.

[80] A. Satyadas, U. Harigopal and N. P. Cassaigne, "Knowledge management tutorial: An editorial overview," *IEEE Trans. Syst. , Man, Cybern. C, Appl. Rev.* , vol. 31, pp. 429-437, 2000.

[81] D. Chauvel and C. Despres, "A review of survey research in knowledge management: 1997–2001 " *J. Knowl. Manag.*, vol. 6, pp. 207-223, 2000.

[82] B. Rubenstein-Montano, J. Liebowitz, J. Buchwalter, D. McCaw, B. Newman and K. Rebeck, "A systems thinking framework for knowledge management," *Decision Support Systems,* vol. 31, pp. 5-16, 2001.

[83] J. Liebowitz, "Knowledge management and its link to artificial intelligence," *Expert Systems with Applications,* vol. 20, pp. 1-6, 2001.

[84] S. H. Liao, "Problem solving and knowledge inertia," *Expert Systems with Applications,* vol. 22, pp. 21-31, 2002.

[85] I. Nonaka and H. Takeuchi, *The Knowledge Creating Company; how Japanese Companies Create the Dynamics of Innovation.* London: Oxford University Press, 1995,

[86] K. M. Wiig, *Knowledge Management. the Central Management Focus for Intelligent-Acting Organization.* Arlington, USA: Schema Press, 1994,

[87] J. S. Dhaliwal and I. Benbasat, "The use and effects of knowledge-based system explanations: theoretical foundations and a framework for empirical evaluation," *Information Systems Research,* vol. 7, pp. 342-362, 1996.

[88] K. C. Laudon and J. P. Laudon, *Essential of Management Information Systems.* ,5th ed.New Jersey, USA: Prentice Hall, 2002,

[89] P. J. Graham-Jones and B. G. Mwllor, "Expert and knowledge based systems in failure analysis" *Expert Systems with Applications,* vol. 2, pp. 137-149, 1995.

[90] N. N. Dawood, "A strategy of knowledge elicitation for developing an integrated bidding/production management expert system for the precast industry. " *Advances in Engineering Software,* vol. 25, pp. 225-234, 1996.

[91] R. M. Keefe and A. D. Preece, "The development, validation and implementation of knowledge-based systems. " *European Journal of Operational Research,* vol. 92, pp. 458-473, 1996.

[92] S. Dutta, "Strategies for implementing knowledge-based systems. " *IEEE Transactions on Engineering Management,* vol. 44, pp. 79-90, 1997.

[93] R. S. Mitra and A. Basu, "Knowledge representation in MICKEY: An expert system for designing microprocessor-based systems. " *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans,* vol. 27, pp. 467-479, 1997.

[94] R. J. Mockler, D. G. Dologite and M. E. Gartenfeld, "Talk with the experts: learning management decision-making using CAI. " *Cybernetics and Systems: An International Journal,* vol. 31, pp. 431-464, 2000.

[95] G. Acosta, C. A. Gonzalez and B. Pulido, "Basic tasks for knowledge-based supervision in process control. " *Engineering Applications of Artificial Intelligence,* vol. 14, pp. 441-455, 2001.

[96] K. Mills and H. Gomaa, "Knowledge-based automation of a design method for concurrent systems. " *IEEE Transactions on Software Engineering, 28, 228–255.,* vol. 28, pp. 228-255, 2002.

[97] A. Skoglund, A. Brundin and C. F. Mandenius, "Applying process monitoring with multivariate analysis through a knowledge-based systems approach to a paperboard machine," *Computers in Industry,* vol. 56, pp. 472-478, 2005.

[98] S. M. Sapuan, M. S. D. Jacob, F. Mustapha and N. Ismail, "A prototype knowledge-based system for material selection of ceramic matrix composites of automotive engine components," *Materials & Design,* vol. 23, pp. 701-708, 2002.

[99] H. W. Law, H. Y. Tam, A. H. S. Chan and I. K. Hui, "“Object-oriented knowledge-based computer-aided process planning system for bare circuit boards manufacturing ”," *Computers in Industry,* vol. 45, pp. 137-153, 2001.

[100] J. P. T. Mo and C. Menzel, "An integrated process model driven knowledge based system for remote customer support," *Computers in Industry,* vol. 37, pp. 171-183, 11. 1998.

[101] Y. B. Moon, C. K. Divers and H. Kim, "AEWS: an integrated knowledge-based system with neural networks for reliability prediction," *Computers in Industry,* vol. 35, pp. 101-108, 1998.

[102] R. J. V. Lee and R. I. Young, "Information Supported Design for Manufacture of Injection-Moulded Rotational Products," *International Journal of Production Research,* vol. 36, pp. 3347-3366, 1998.

[103] S. W. Changchien and L. Lin, ""A knowledge-based design critique system for manufacture and assembly of rotational machined parts in concurrent engineering " " *Computers in Industry,* vol. 32, pp. 117-140, 1996.

[104] Q. Tian, J. Ma and O. Liu, "A hybrid knowledge and model system for R&D project selection," *Expert Systems with Applications,* vol. 23, pp. 265-271, 2002.

[105] Srinivas, A. K. Choudhary, R. Swarnkar, B. P. Das, R. D. Alcock and J. A. Harding, "A knowledge driven decision support framework to aid autonomous vehicle systems' mission objectives," in *DTC Annual Conference,* 2006,

[106] P. Tyndale, "A taxonomy of knowledge management software tools: origins and applications. " *Evaluation and Program Planning,* vol. 25, pp. 183-190, 2002.

[107] H. Chen, J. Schroeder, R. V. Hauck, L. Ridgeway, H. Atabakhsh, H. Gupta, C. Boarman, K. Rasmussen and A. W. Clements, "COPLINK Connect: information and knowledge management for law enforcement. " *Decision Support Systems,* vol. 34, pp. 271-285, 2002.

[108] H. Cunningham, D. Maynard, K. Bontcheva and V. Tablan. GATE: A framework and graphical development environment for robust NLP tools and applications. Presented at Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), 2002. [http://gate.ac.uk/sale/acl02/acl-main.pdf].

[109] M. S. Abdullah, R. Paige, C. Kimble and I. Benest, "A UML Profile for Knowledge-Based Systems Modelling," *Software Engineering Research, Management & Applications, 2007. SERA 2007. 5th ACIS International Conference on,* pp. 871-878, 2007.

[110] F. Afshar, "Capturing consensus knowledge from multiple experts" in *Proceedings of ES-2002, the Twenty-Second SGAI International Conference on Knowledge Based Systems and Applied Artificial Intelligence,* 2002, pp. 253-263.

[111] A. Hakansson, "UML as an approach to modelling knowledge in rulebased systems," in *Proceedings of ES2001, the Twenty-First SGAI International Conference on Knowledge Based Systems and Applied Artificial Intelligence,* 2001, pp. 187-200.

[112] A. J. Rhem, *UML for Developing Knowledge Management System.* Auerbach publications, Taylor & Francis, 2006,

[113] H. H. Huang, X. Hao and M. Benjamin, "Automated Knolwedge Acqistion for design and Manufacturing: The case of micromachine atomizer," vol. 12, pp. 377-391, 2001.

[114] M. Shahbaz, Srinivas, J. A. Harding and M. Turner, "Product Design and Manufacturing Process Improvement using Association Rules," *IMechE Part B Journal of Engineering Manufacture,* vol. 220, pp. 243-254, 2006.

[115] G. Piatetsky-Shapiro, "Knowledge discovery in real databases: A report on the IJCAI-89 Workshop," *AI Mag.*, vol. 11, pp. 68-70, 1991.

[116] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From data mining to knowledge discovery: An overview," in *Advances in Knowledge Discovery and Data Mining* Anonymous American Association for Artificial Intelligence, 1996, pp. 1-34.

[117] C. Glymour, D. Madigan, D. Pregibon and P. Smyth, "Statistical inference and data mining," *Commun ACM,* vol. 39, pp. 35-41, 1996.

[118] D. J. Hand, "Statistics and data mining: intersecting disciplines," *SIGKDD Explor.Newsl.,* vol. 1, pp. 16-19, 1999.

[119] S. Mitra, S. K. Pal and P. Mitra, "Data mining in soft computing framework: a survey," *Neural Networks, IEEE Transactions on,* vol. 13, pp. 3-14, 2002.

[120] A. A. Freitas, "A survey of evolutionary algorithms for data mining and knowledge discovery," in *Advances in Evolutionary Computing: Theory and Applications* Anonymous Springer-Verlag New York, Inc., 2003, pp. 819-845.

[121] D. J. Hand, *Construction and Assessment of Classification Rules.* John Wiley & Sons, 1997,

[122] S. Liao and C. Wen, "Artificial neural networks classification and clustering of methodologies and applications – literature analysis from 1995 to 2005," *Expert Systems with Applications,* vol. 32, pp. 1-11, 1. 2007.

[123] R. Xu and D. Wunsch, "Survey of clustering algorithms" *Neural Networks, IEEE Transactions on,* vol. 16, pp. 645-678, 2005.

[124] R. Agrawal, T. Imielinski and A. Swami, "Mining association rules between sets of items in large databases" *Proceedings of the 1993 ACM SIGMOD,* Washington, D.C., United States pp. 207-216, 1993.

[125] J. Hipp, U. Guntzer and G. Nakhaeizadeh, "Algorithms for association rule mining-a general survey and comparison" *SIGKDD Explor.Newsl.,* vol. 2, pp. 58-64, 2000.

[126] D. Chiang, L. R. Chow and Y. Wang, "Mining time series data by a fuzzy linguistic summary system," *Fuzzy Sets Syst.,* vol. 112, pp. 419-432, 2000.

[127] P. Giudici, *Applied Data Mining: Statistical Method for Business and Industry.* West Susex, England: John Wiley & sons ltd, 2003,

[128] W. Pedrycz, "Fuzzy set technology in knowledge discovery" *Fuzzy Sets and Systems,* vol. 98, pp. 279-290, 9/16. 1998.

[129] J. Yu, S. Zhou, L. Wang and J. Lei, "Advances in fuzzy sets and knowledge discovery" *Computers & Mathematics with Applications,* vol. In Press, Corrected Proof,

[130] E. Hüllermeier, "Special issue on fuzzy sets in knowledge discovery", *Fuzzy Sets and Systems,* vol. 149, pp. 1-3, 1/1. 2005.

[131] J. R. Quinlan, "Induction of Decision Trees," *Mach.Learn.,* vol. 1, pp. 81-106,

[132] J. R. Quinlan, *C4.5: Programs for Machine Learning.* Morgan Kaufmann Publishers Inc., 1993, pp. 302.

[133] D. T. Pham and A. A. Afify, "Machine-learning techniques and their applications in manufacturing " *Proceedings of the I MECH E Part B Journal of Engineering Manufacture,* vol. B5, pp. 395-412, 2005.

[134] S. S. Haykin, *Neural Networks: A Comprehensive Foundation.* Newyork, Oxford: Maxwell Macmillan International, Newyork and Macmillan College publishing, Oxford, 1994.

[135] S. Srinivas and J. A. Harding, "A data mining integrated architecture for shop floor control," *Proc. Inst. Mech. Eng. Pt. B: J. Eng. Manuf.*, vol. 222, pp. 605-624, 05/01. 2008.

[136] Z. Michalewicz, *Genetic Algorithms Plus Data Structures Equals Evolution Programs.* Springer-Verlag New York, Inc., 1994, pp. 340.

[137] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley Longman Publishing Co., Inc., 1989, pp. 372.

[138] A. K. Choudhary, M. K. Tiwari and J. A. Harding, "Part Selection and Operation-Machine Assignment in a Flexible Manufacturing System Environment: A Genetic Algorithm with Chromosome Differentiation-Based Methodology," *Proc. Inst. Mech. Eng. Pt. B: J. Eng. Manuf.*, vol. 220, pp. 677-694, 01/01. 2006.

[139] Z. Pawlak, "Rough set approach to knowledge-based decision support," *European Journal of Operational Research,* vol. 99, pp. 48-57, 5/16. 1997.

[140] A. Kusiak, "Rough set theory: a data mining tool for semiconductor manufacturing," *Electronics Packaging Manufacturing, IEEE Transactions on,* vol. 24, pp. 44-50, 2001.

[141] R. Agrawal, Imieli, T. Ski and A. Swami, "Mining association rules between sets of items in large databases," pp. 207-216, 1993.

[142] S. Kotsiantis and D. Kanellopoulos, "Association Rule Mining: A Recent Overview," *GESTS International Transactions on Computer Science and Engineering,* vol. 32, pp. 71-82, 2006.

[143] J. A. Harding, M. Shahbaz, Srinivas and A. Kusiak, "Data Mining in Manufacturing: A Review," *American Society of Mechanical Engineers (ASME): Journal of Manufacturing Science and Engineering,* vol. 128, pp. 969, 2006.

[144] J. Han, *Data Mining: Concepts and Techniques.* Morgan Kaufmann Publishers Inc., 2005,

[145] A. K. Choudhary, J. A. Harding and M. K. Tiwari, "Data mining in manufacturing: a review based on the kind of knowledge," *J. Intell. Manuf.* 2009 (Published online)

[146] R. Feldman, *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.* Cambridge University Press, 2006,

[147] K. Bontcheva, V. Tablan, D. Maynard and H. Cunningham, "Evolving GATE to meet new challenges in language engineering," *Nat.Lang.Eng.,* vol. 10, pp. 349-373, 2004.

[148] M. V. Kiselev, S. M. Ananyan and S. B. Arseniev, "PolyAnalyst Data Analysis Technique and Its Specialization for Processing Data Organized as a Set of Attribute Values," pp. 352-360, 1998.

[149] N. Uramoto, H. Matsuzawa, T. Nagano, A. Murakami, H. Takeuchi and K. Takeda, "A text-mining system for knowledge discovery from biomedical documents," *IBM Syst J,* vol. 43, pp. 516-533, 2004.

[150] A. Cohen and W. Hersh, "A survey of current work in biomedical text mining," *Brief Bioinform,* vol. 6, pp. 57-71, 03. 2005.

[151] Y. Tseng, C. Lin and Y. Lin, "Text mining techniques for patent analysis," *Inf.Process.Manage.,* vol. 43, pp. 1216-1247, 2007.

[152] A. Kloptchenko, T. Eklund, J. Karlsson, B. Back, H. Vanharanta and A. Visa, "Combining data and text mining techniques for analysing financial reports: Research Articles," *Int.J.Intell.Syst.Account.Financ.Manage.*, vol. 12, pp. 29-41, 2004.

[153] J. Gerdes, "EDGAR-Analyzer: automating the analysis of corporate data contained in the SEC's EDGAR database," *Decision Support Systems,* vol. 35, pp. 7-29, 4. 2003.

[154] J. Dorre, P. Gerstl and R. Seiffert, "Text mining: finding nuggets in mountains of textual data," pp. 398-401, 1999.

[155] R. Menon, L. H. Tong, S. Sathiyakeerthi, A. Brombacher,C. Leong, "The Needs and Benefits of Applying Textual Data Mining within the Product Development Process" *Qual. Reliab. Eng. Int.,* vol. 20, pp. 1-15, 2004.

[156] S. Loh, J. Oliveira, Palazzo M.De and M. A. Gameiro, "Knowledge Discovery in Texts for Constructing Decision Support Systems," *Appl. Intell.,* vol. 18, pp. 357-366, 2003.

[157] H. K. Lin, J. A. Harding and A. K. Choudhary, "The Universal Knowledge Moderator for globally distributed and collaborative e-manufacturing," *Industrial Informatics, 2008. INDIN 2008. 6th IEEE International Conference on,* pp. 1227-1231, 2008.

[158] A. K. Choudhary, J. A. Harding and H. K. Lin, "Engineering moderator to universal knowledge moderator for moderating collaborative projects," *Global Journal of e-Business & Knowledge Management,* vol. 3, pp. 5-13, 2007.

[159] M. Holsheimer, "Data mining by business users: integrating data mining in business processes," pp. 266-291, 1999.

[160] R. Pechter, "Data mining standards, services and platforms 2005 workshop report," *SIGKDD Explor.Newsl.,* vol. 7, pp. 137-138, 2005.

[161] C. Shearer, "The CRISP-DM model: the new blueprint for data mining," *Journal of Data Warehousing,* vol. 5, pp. 13-22, 2005.

[162] F. F. Ren, M. J. Zhang and Q. Bai, "A fuzzy-based approach for partner selection in multi-agent systems." in *6th IEEE/ACIS International Conference on Computer and Information Science,* 2007,

[163] W. Shen, Q. Hao, S. Wang, Y. Li and H. Ghenniwa, "An agent-based service-oriented integration architecture for collaborative intelligent manufacturing," *Robotics and Computer-Integrated Manufacturing,* vol. 23, pp. 315-325, 6. 2007.

[164] M. Merdan, V. Kordic, A. Zoitl and A. Lazinica, "Knowledge-based Multi-agent Architecture," *International Conference on Computational Intelligence for Modelling Control and Automation,and International Conference on Intelligent Agents,Web Technologies and Internet Commerce (CIMCA-IAWTIC'06),* 2006.

[165] W. Shen, Q. Hao, H. J. Yoon and D. H. Norrie, "Applications of agent-based systems in intelligent manufacturing: An updated review," *Advanced Engineering Informatics,* vol. 20, pp. 415-431, 10. 2006.

[166] Y. Li, B. Huang, W. Liu, C. Wu and H. Gou, "Multi-agent system for partner selection of virtual enterprise," in *Proceedings of 16Th IFIP World Computer Congress,* 2000,

[167] H. Wang, "Intelligent Agent-Assisted Decision Support System: Integration of Knowledge of Discovery, Knowledge Analysis, and Group Decision Support," *Expert System with Application,* vol. 12, pp. 323-335, 1997.

[168] J. H. Lee and S. C. Park, "Agent and data mining based decision support system and its adaptation to a new customer-centric electronic commerce," *Expert Systems with Applications,* vol. 25, pp. 619-635, 11. 2003.

[169] N. Huang and S. Diao, "Ontology-based enterprise knowledge integration," *Robotics and Computer-Integrated Manufacturing,* vol. 24, pp. 562-571, 8. 2008.

[170] W. W. C. Chung and J. J. F. Pak, "A case study: using UML to develop a knowledge-based system for supporting business systems in a small financial institute," *Int. J. Comput. Integr. Manuf.,* vol. 19, pp. 59-68(10), Number 1/January-February 2006.

[171] Zhou Xiaoye, Liu Zuofeng and Xie Xiuyi, "Virtual Enterprise Inventory Management Model Based on UML," *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on,* pp. 4954-4957, 2007.

[172] W. Wang, X. Liu and W. Liu, "Development of Knowledge-based System for Adaptive Scheduling Using Unified Modeling Language," *Control and Automation, 2007.ICCA 2007.IEEE International Conference on,* pp. 914-918, 30 2007-June 1. 2007.

[173] G. Booch, *Object Oriented Design with Applications.* Redwood City: Benjamin Cummings, 1991,

[174] M. S. Abdullah, R. Paige, C. Kimble and I. Benest, "A UML Profile for Knowledge-Based Systems Modelling," pp. 871-878, 2007.

[175] A. Choudhary, J. Harding, P. Carrillo, P. Oluikpe and N. Rahman, "Text mining post project reviews to improve the construction project supply chain design," in *DMIN,* 2008, pp. 391-397.

[176] A. K. Choudhary, P. Oluikpe, J. A. Harding and P. M. Carillo, "The Needs And Benefits Of Text Mining Applications On Post Project Reviews," *Computers in Industry,* vol. Accepted for publication, 2009.

[177] S. Newell, M. Bresnen, L. Edelman, H. Scarbrough and J. Swan, "Sharing Knowledge Across Projects: Limits to ICT-led Project Review Practices," *Management Learning,* vol. 37, pp. 167-185, June 1. 2006.

[178] L. Kagal, J. Hendlerb and T. Berners-Leea, "The Semantic Web and Policy" *Semantics: Science, Services and Agents on the World Wide Web,* vol. 7, pp. vii-ix, 1. 2009.

[179] Department of Trade and Industry Small Business Services, "Statistical Press Release 20th December", 2006, Last updated Jan 2009.

[180] N. G. Sánchez, D. Apolinar, G. Zubiaga, J. Atahualpa, I. González and A. Molina, "Virtual breeding environment: A first approach to understanding working and sharing principles," in *In Proceedings of InterOp-ESA'05,* 2005,

[181] P. M. Carillo, P. Oluikpe, A. K. Choudhary and J. A. Harding, "Text mining of post project reviews," in *CIB W102 Conference on Information and Knowledge Management,* 2008, pp. 70-81.

[182] B. R. Katzy, G. Schuh, K Millarg, "Die Virtuelle Fabrik - Produzieren in Netzwerken", Technische Rundschau, (43),1996, 30-34.

[183] TELEflow, a telematics engineering project, funded under the TELEMATICS APPLICATIONS Program (4th EU Framework Program 194-98).

[184] Luis M. Camarinha-Matos and Hamideh Afsarmanesh, "A framework for virtual organization creation in a breeding environment", *Annual Reviews in Control,* Vol. 31, Issue 1, pp. 119-135, 2007.

List of Publications

The following papers have been published or accepted for publication related to this research. Abstracts of each paper are attached as appendix.

1. A. K. Choudhary, J. A. Harding and K. Popplewell, "Knowledge discovery for moderating collaborative projects," in *IEEE International Conference on Industrial Informatics,* 2006, pp. 519-524.
2. A. K. Choudhary, J. A. Harding and M. K. Tiwari, "Data mining in manufacturing: a review based on the kind of knowledge,"*J. Intell. Manuf, 2009, accepted for publication.*
3. H. K. Lin, J. A. Harding and A. K. Choudhary, "The Universal Knowledge Moderator for globally distributed and collaborative e-manufacturing," *Industrial Informatics, 2008. INDIN 2008. 6th IEEE International Conference on,* pp. 1227-1231, 2008.
4. A. K. Choudhary, J. A. Harding and H. K. Lin, "Engineering moderator to universal knowledge moderator for moderating collaborative projects," *Global Journal of e-Business & Knowledge Management,* vol. 3, pp. 5-13, 2007.
5. A. K. Choudhary, J. A. Harding, P. Carrillo, P. Oluikpe and N. Rahman, "Text mining post project reviews to improve the construction project supply chain design," in International Conference of Data Mining, *DMIN,* 2008, Lasvegas, pp. 391-397.
6. A. K. Choudhary, P. Oluikpe, J. A. Harding and P. M. Carillo, "The Needs And Benefits Of Text Mining Applications On Post Project Reviews," *Computers in Industry,* Accepted for publication, 2009.
7. N. Khilwani, J. A. Harding and A. K. Choudhary, "Semantic Web in Manufacturing", 2009, Accepted for publication in Journal of Engineering Manufacture, I Mech E Part-B.
8. P. M. Carillo, P. Oluikpe, A. K. Choudhary and J. A. Harding, "Text mining of post project reviews," in *CIB W102 Conference on Information and Knowledge Management,* 2008, pp. 70-81.

Appendix - 1

Semantic web in manufacturing

N Khilwani, J A Harding*, and A K Choudhary
Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, UK

The manuscript was received on 15 October 2008 and was accepted after revision for publication on 11 March 2009.

DOI: 10.1243/09544054JEM1399

Abstract: Advances in manufacturing systems include attempts to create collaborative networks for enterprise integration and information interoperability. To achieve collaboration and sharing effectively, various networking technologies have been proposed in the literature. The web has emerged as a basic entity for interconnecting man and machine and almost all parts of the enterprise community are being reshaped to exploit the opportunities that it offers. Apart from web technology, there are various other tools and techniques that have attracted research communities for representing data in ways that both machines and humans can understand. Semantic web, the second-generation web technology, is one enriched by machine-processable information to support the users in their tasks. This paper presents the vision of the semantic web and describes ontologies and associated metadata as the building blocks of the semantic web. It reviews the literature dealing with the application of the semantic web and ontology in the broad domain of manufacturing. First, brief details about key enablers, i.e. web services, semantic web, semantic services, and ontology, are presented. Then the implementation of these approaches in different sectors of manufacturing is discussed. A knowledge base for all the information resources concerned with the manufacturing domain is also built up in this paper. An ontology model for a knowledge base of information resources is designed in Protégé software, which can be used for storing and searching information about authors, journals, blogs, newspapers, and many other sources of information.

Keywords: manufacturing, web technology, ontology, semantic web, protégé

1 INTRODUCTION

Amid, the immense pressure of continuous improvement in productivity, responsiveness and flexibility, manufacturing firms are facing huge challenges from their consumer market to satisfy individual customer requirements. Manufacturers are striving to meet these requirements by focusing their core competencies to enable them to compete effectively and efficiently to the best of their ability. The traditional view of an enterprise, with clear boundaries, limited relationships with others and a strong internal focus on efficiency and quality is no longer adequate [1]. Today, firms aim to improve their competitive edge by focusing on core competencies and outsourcing all other functions. In such environments, organizational barriers break down, partnerships with suppliers, clients, and even competitors are common place and efficiency and quality are considered beyond enterprise boundaries. The current trend among manufacturing firms is to form collaborative networks to succeed and achieve business goals [2].

The 'automotive industry' is one of the best examples to show this transition. In order to compete with the global giants, the automotive sector has concentrated on core competencies and collaborations between organizations that compliment their skills and core activities [3]. Companies adopt this new policy by sharing designs, developments, and platforms and thus reduce the burden of costs associated with the development of non-core activities [4]. This revolution started in the recession of the early 1990s, with the massive move towards outsourcing of non-core activities such as information technology, public relations, human resources, etc. This step was an impressive move in the automotive sector that resulted in massive reductions of fixed

**Corresponding author: Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Ashby Road, Loughborough LE11 3TU, UK.*
email: j.a.harding@lboro.ac.uk

Appendix - 2

J Intell Manuf
DOI 10.1007/s10845-008-0145-x

Data mining in manufacturing: a review based on the kind of knowledge

A. K. Choudhary · J. A. Harding · M. K. Tiwari

Received: 14 December 2007 / Accepted: 2 June 2008

Abstract In modern manufacturing environments, vast amounts of data are collected in database management systems and data warehouses from all involved areas, including product and process design, assembly, materials planning, quality control, scheduling, maintenance, fault detection etc. Data mining has emerged as an important tool for knowledge acquisition from the manufacturing databases. This paper reviews the literature dealing with knowledge discovery and data mining applications in the broad domain of manufacturing with a special emphasis on the type of functions to be performed on the data. The major data mining functions to be performed include characterization and description, association, classification, prediction, clustering and evolution analysis. The papers reviewed have therefore been categorized in these five categories. It has been shown that there is a rapid growth in the application of data mining in the context of manufacturing processes and enterprises in the last 3 years. This review reveals the progressive applications and existing gaps identified in the context of data mining in manufacturing. A novel text mining approach has also been used on the abstracts and keywords of 150 papers to identify the research gaps and find the linkages between knowledge area, knowledge type and the applied data mining tools and techniques.

A. K. Choudhary · J. A. Harding (✉)
Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK
e-mail: J.A.Harding@lboro.ac.uk

A. K. Choudhary
e-mail: a.k.choudhary@lboro.ac.uk

M. K. Tiwari
Department of Industrial Engineering and Management, Indian Institute of Technology, Kharagpur 721302, India
e-mail: mkt09@hotmail.com

Keywords Knowledge discovery · Data mining · Manufacturing · Text mining · Literature review

Introduction

Knowledge provides power in many manufacturing contexts enabling and facilitating the preservation of valuable heritage, new learning, solving intricate problems, creating core competencies and initiating new situations for both individuals and organizations now and in the future (Choudhary et al. 2007). In most sectors, manufacturing is extremely competitive and the financial margins that differentiate between success and failure are very tight, with most established industries needing to compete, produce and sell at a global level. To master these trans-continental challenges, a company must achieve low cost production yet still maintain highly skilled, flexible and efficient workforces who are able to consistently design and produce high quality and low cost products. In higher-wage economies, this can generally only be done through very efficient exploitation of knowledge (Harding and Popplewell 2006; Choudhary et al. 2006). However knowledge can take many forms and it is necessary to identify the kind of knowledge to be mined when examining the huge amount of data generated during manufacturing.

In modern manufacturing, the volume of data grows at an unprecedented rate in digital manufacturing environments, using barcodes, sensors, vision systems etc. These data may be related to design, products, machines, processes, materials, inventories, maintenance, planning and control, assembly, logistics, performances etc., and may include patterns, trends, associations and dependencies. However, the use of accumulated data has been limited, which has led to the "rich data but poor information" problem (Wang and McGreavy

Computers in Industry xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

The needs and benefits of Text Mining applications on Post-Project Reviews

A.K. Choudhary [a], P.I. Oluikpe [b], J.A. Harding [a,*], P.M. Carrillo [b]

[a] Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK
[b] Department of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK

ARTICLE INFO

Article history:
Received 9 November 2007
Received in revised form 2 April 2009
Accepted 19 May 2009
Available online xxx

Keywords:
Text Mining
Knowledge discovery
Post-Project Reviews (PPRs)
Manufacturing and construction

ABSTRACT

Post-Project Reviews (PPRs) are a rich source of knowledge and data for organisations – if organisations have the time and resources to analyse them. Too often these reports are stored, unread by many who could benefit from them. PPR reports attempt to document the project experience – both good and bad. If these reports were analysed collectively, they may expose important detail, e.g. recurring problems or examples of good practice, perhaps repeated across a number of projects. However, because most companies do not have the resources to thoroughly examine PPR reports, either individually or collectively, important insights and opportunities to learn from previous projects, are missed. This research explores the application of knowledge discovery techniques and Text Mining to uncover patterns, associations, and trends from PPR reports. The results might then be used to address problem areas, enhance processes, and improve customer relationships. A case study related to two construction companies is presented in this paper and knowledge discovery techniques are used to analyse 50 PPR reports collected during the last three years. The case study has been examined in six contexts and the results show that Text Mining has a good potential to improve overall knowledge reuse and exploitation.

1. Overview

Post-Project Reviews (PPRs) are one of the most important and common approaches for the capture of project knowledge. They provide opportunities for project teams to share, discuss and even explain their experiences through face-to-face, facilitated interactions before a project is closed and the team is dissolved. PPRs therefore allow multi-disciplinary teams to critique a project to determine both positive and negative aspects, potentially capturing tacit knowledge as learning points to improve the planning and execution of future projects. Interactive debates between the project team members during PPRs may lead to greater innovation and better ideas than can be achieved from any individual. This shared communication is crucial as each individual contributor will inevitably have his or her own perspective or viewpoint and will only know part of the whole project story [1]. In the UK, major companies such as BP Amoco, BAA plc, National Grid Transco and construction companies such as Bovis Lend Lease, IPSL, Simons Design and Buro Happold, have adopted PPRs in an effort to learn from experience.

Conducting PPRs is time consuming, manpower intensive and expensive in terms of company overheads. Disterer [2] mentioned that after finishing the project, team members are spread all over the company and project documentation is stored in folders without retaining the essentials for later use. In his assessment of PPRs, Busby [3] highlighted that "PPRs were important learning mechanisms and their value seems to be underestimated by individuals who do not appreciate the need to disseminate insights throughout the organisation". PPRs are often conducted as a part of companies' quality systems, however a major problem lies in the fact that companies have insufficient resource to act on the outcome of PPRs [4]. Companies often have individuals with responsibility for creating PPRs, but they do not have individuals or teams responsible for subsequently analyzing the PPRs collectively to identify the good or bad practice and areas requiring improvements across a range of projects. In addition, Orange et al. [5] and Kamara et al. [6] also identified that PPRs have huge potential for much more thorough exploitation. If information and knowledge from PPRs can be extracted and analysed effectively, good and bad practices might be identified so that lessons are learnt from past projects and knowledge is reused or exploited to improve the quality and levels of success in future projects.

Knowledge Discovery in Text (KDT) and Text Mining (TM) [7] are very recent and increasingly interesting areas of research in computer science. KDT and TM are mostly automated techniques that aim to discover high-level information from huge amounts of textual data and present it in a useful form to the potential user, who might be an analyst, decision maker, project manager, etc. PPRs are generally recorded in reports or other text-based documentation so KDT or TM techniques might therefore be

* Corresponding author.
E-mail address: J.A.Harding@lboro.ac.uk (J.A. Harding).

0166-3615/$ – see front matter © 2009 Published by Elsevier B.V.
doi:10.1016/j.compind.2009.05.006

Text Mining Post Project Reviews to Improve the Construction Project Supply Chain Design

A. K. Choudhary[1], J. A. Harding[1*], P. M. Carrillo[2], P. Oluikpe[2] and N. Rahman[1]
[1]Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire, United Kingdom
[2]Department of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, United Kingdom
*Communicating Author: J.A.Harding@lboro.ac.uk

Abstract - *Post Project Reviews (PPR) capture good and bad practices, identify problems, waste, risks, missed opportunities, communication lag, financial issues, partner relationships etc., of a construction project supply chain (CPSC). They are huge sources of information, knowledge and experience from project managers, clients, suppliers and contractors, related to issues from every stage of the construction project. If these reports were analysed collectively, they may expose important detail, perhaps repeated across a number of projects. However, because most companies do not have the resources to thoroughly examine these PPRs, either individually or collectively, important insights are missed thereby leading to missed opportunities to learn from previous projects. This research shows that the hidden knowledge and experiences could be captured using knowledge discovery and text mining approaches to uncover patterns, associations, and trends in PPRs. The results might then be used to address specific problem areas, enhance processes and improve the design and planning for new construction projects.*

Keywords: PPRs, Construction Project, Supply Chain, Knowledge discovery and Text Mining

1 Introduction

The construction industry forms one of the most diverse and unstable sectors within the UK economy. It faces widely fluctuating demand cycles, project specific product demands, uncertain production conditions and has to combine a diverse range of specialist skills within geographically dispersed short term project environments[1]. A construction project supply chain (CPSC) may contain hundreds of firms, contractors, subcontractors, material and equipment suppliers, engineering and design teams, and consulting firms[2]. Collaborations between the various entities of the CPSC are temporary and may vary from project to project. The lifecycle of a construction supply chain is limited to a particular project only. Post Project Reviews (PPR) of construction projects are one of the most important and common approaches for the capture of knowledge and lessons learned from the operation of a CPSC. They provide opportunities for the project team members to share, discuss and even explain their experiences through face-to-face, facilitated interactions before a project is closed and the team is dissolved. PPRs therefore allow multi-disciplinary teams to critique a project to determine both positive and negative aspects, potentially capturing tacit knowledge as learning points to improve the planning, execution and design of new construction projects [3-4]. Debates between the project team members during PPRs may lead to greater innovation and better ideas than can be achieved from any individual. Orange *et al.* [5] and Kamara *et al.* [6] also identified that PPRs have a huge potential for much more thorough exploitation. If information and knowledge from PPRs can be extracted and analyzed effectively, good and bad practices might be identified so that lessons are learnt from past projects and knowledge reused or exploited to improve the quality and levels of success in redesigning the CPSC.

Knowledge Discovery in Text (KDT) and Text Mining (TM) [12] are very recent and increasingly interesting areas of research in computer science. KDT and TM are mostly automated techniques that aim to discover high level information from huge amounts of textual data and present it in a useful form to the potential user, who might be an analyst, decision maker or project manager etc. KDT or TM techniques might therefore be applied on PPR documents from multiple projects to potentially identify information relating to bad, good or even best practices. The benefits of mining PPR texts lie in the possibilities of discovering patterns, associations and linkages of processes, activities and terms occurring in the reports. The organization may then adjust its activities to reflect what is learned from the KDT and TM with the aim of improving processes, optimizing profit and improving client retention. In addition, whenever a new construction project is initiated, it would be very beneficial if lessons learned from previous projects could be quickly identified to reduce the chances of errors being repeated and increase the potential for savings in cost and time.

This paper examines a method for applying TM on PPRs as follows: Section II discusses how PPRs capture the knowledge and experiences of CPSCs; Section III briefly discusses Knowledge Discovery in Text and Text Mining

A. Choudhary, J. Harding, P. Carrillo, P. Oluikpe and N. Rahman, "Text mining post project reviews to improve the construction project supply chain design," in International Conference on Data Mining, *DMIN*, 2008, pp. 391-397

Text Mining of Post Project Reviews

Paul Oluikpe,
Loughborough University, UK (email: p.i.oluikpe@lboro.ac.uk)
Patricia Carrillo,
Loughborough University, UK (email: p.m.carrillo@lboro.ac.uk)
Jenny Harding
Loughborough University, UK (email:j.a.harding@lboro.ac.uk)
Alok Choudhary
Loughborough University, UK (email: a.k.choudhary@lboro.ac.uk)

Abstract

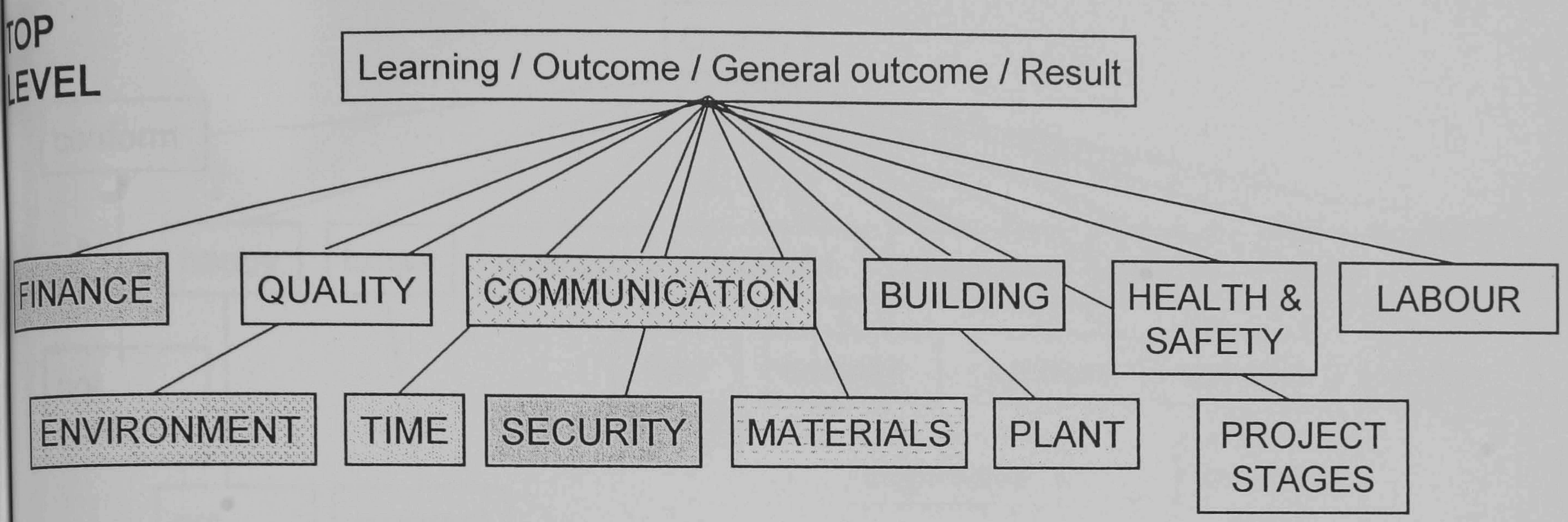
Post Project Reviews (PPR) are a rich source of knowledge and information for organisations - if they have the time and resources to analyse them. Too often such reports are stored, unread by many who can benefit from them. PPRs attempt to document the project experience – both good and bad. If these reports were analysed collectively, they may expose important detail, perhaps repeated between projects. However, because most companies do not have the resources to examine these PPR, either individually or collectively, important insights are missed thereby leading to a missed opportunity to learn from previous projects. Hidden knowledge and experiences can be captured by using knowledge discovery and text mining to uncover patterns, associations, and trends in data. The results might then be used to enhance processes, improve customer relationships, and identify specific problem areas to address.

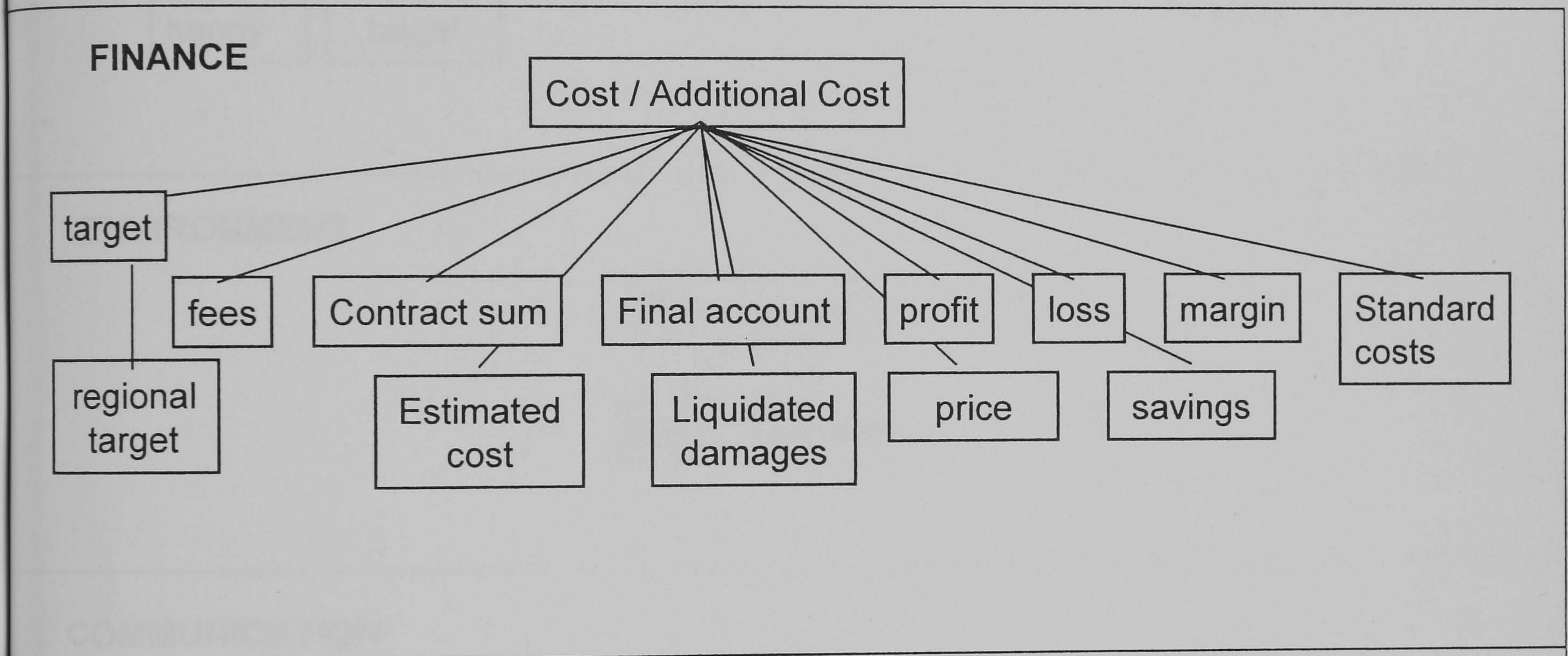
This paper outlines an ongoing research project that investigates the use of knowledge discovery and text mining on Post Project Reviews. An illustrative example will be presented using case studies from the construction sector. The PPR processes of two construction companies were mapped with the aim of understanding the context, format, terminologies used and key knowledge areas suitable for text mining. The textual examination of the PPR reports was complemented by semi-structured interviews and workshops to understand the production and content of the reports. Preliminary results highlight that although organisations have publicised, standard processes for PPR, there is a variance in how these are conducted and produced on a regional basis. These variances provide a number of challenges for organisations from a corporate perspective. Also, there is an over-reliance on key individuals with little attempt to make some of their knowledge more explicit and therefore easier to disseminate between project team members. This paper summarises the challenges in identifying the type of knowledge to be text mined, the format of PPR reports and the process of conducting PPR. It will also highlights the development of suitable

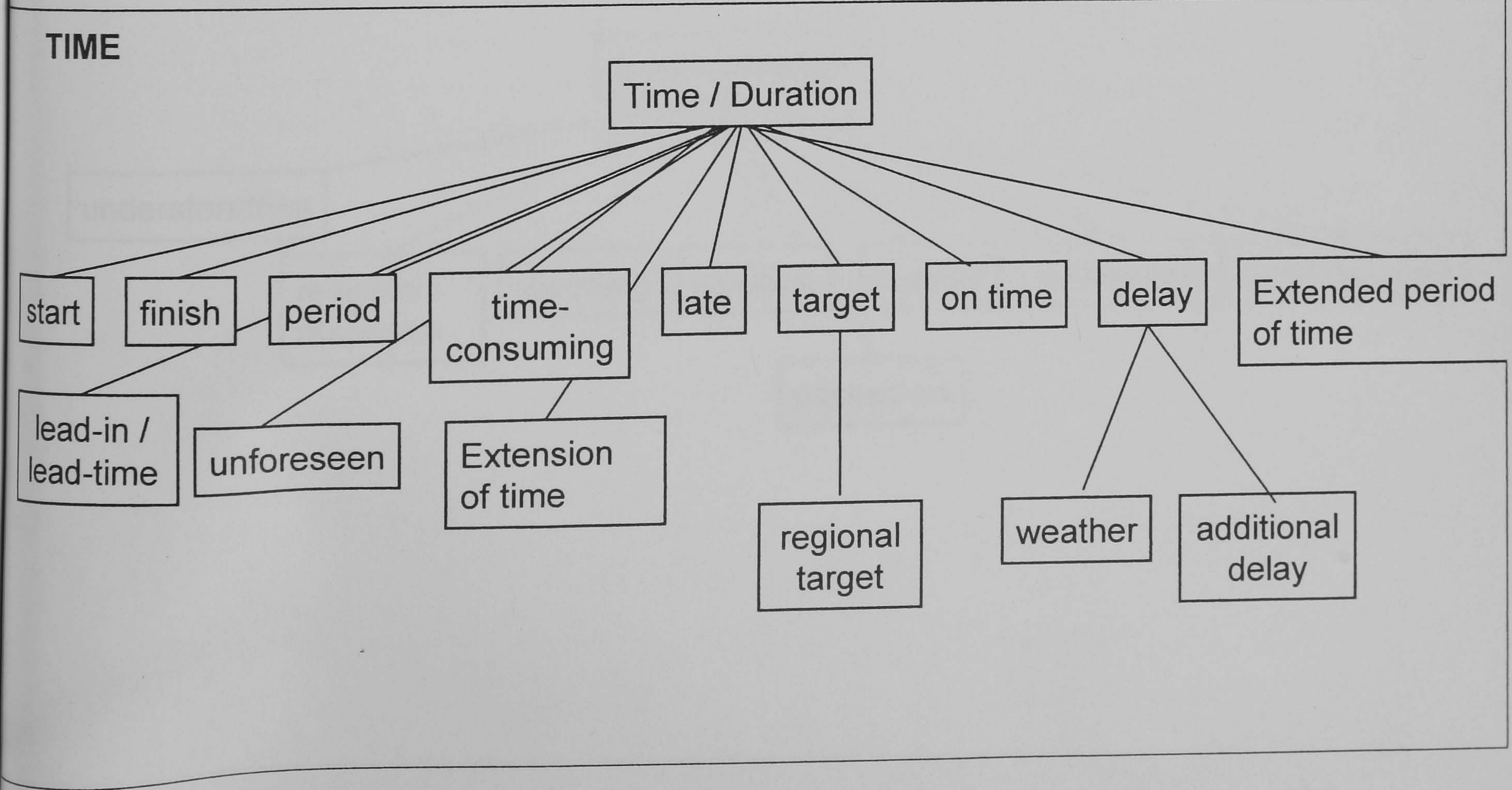
P. M. Carillo, P. Oluikpe, A. K. Choudhary and J. A. Harding, "Text mining of post project reviews," in *CIB W102 Conference on Information and Knowledge Management*, 2008, pp. 70-81.

Appendix - 6

Initial Ontology Development

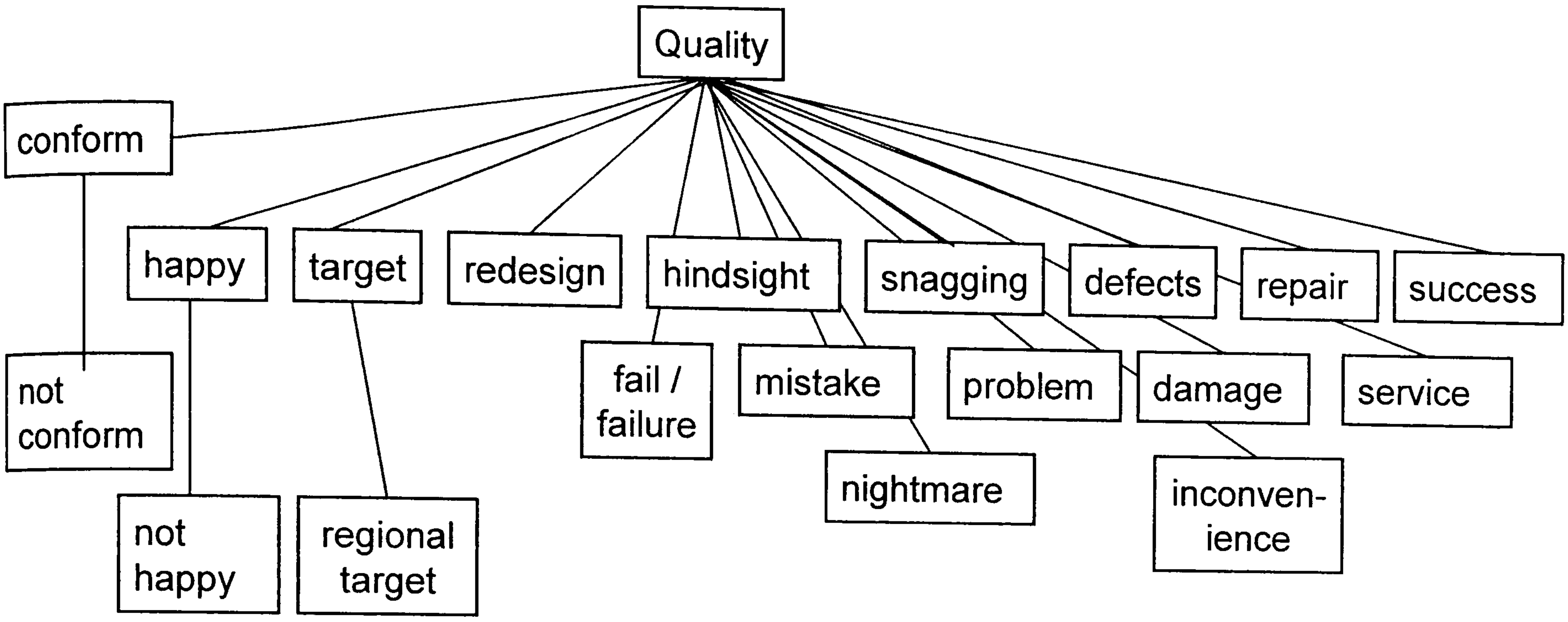




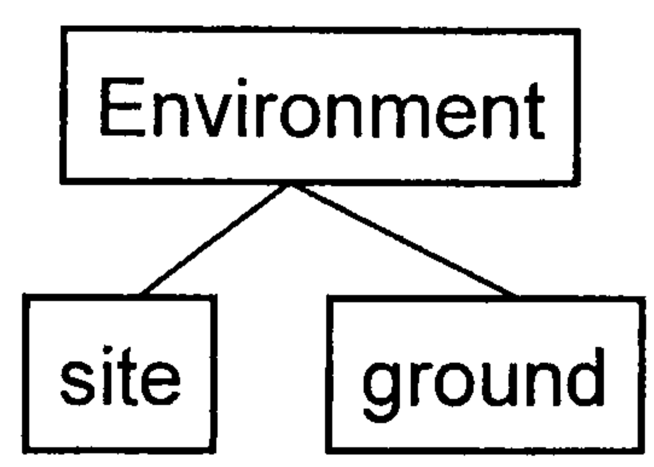


Appendix - 6

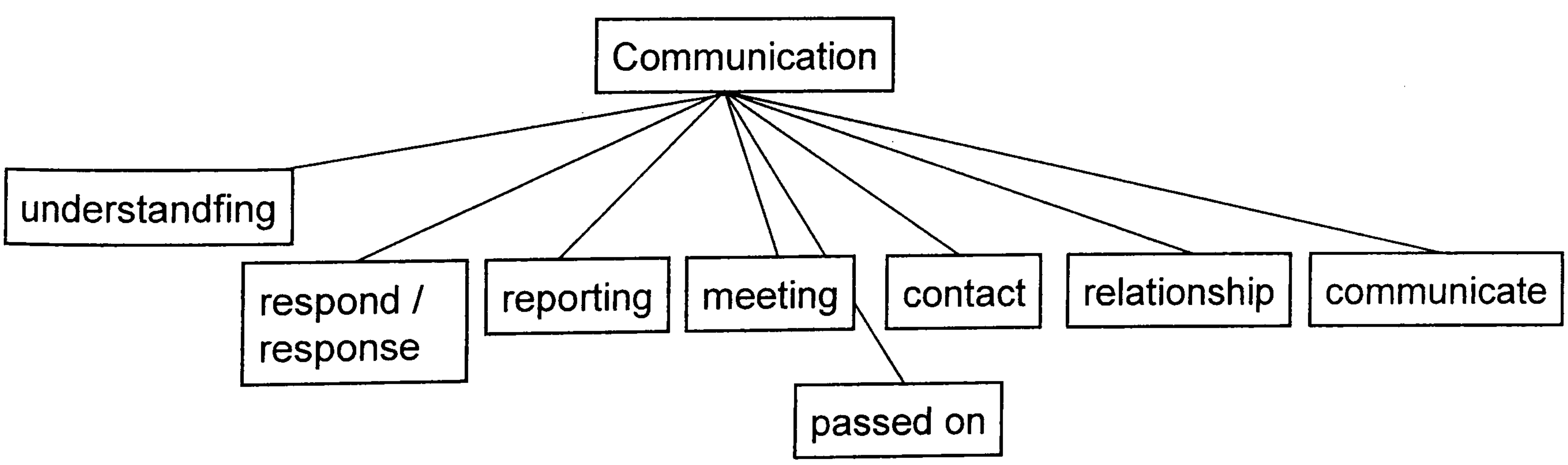
QUALITY



ENVIRONMENT

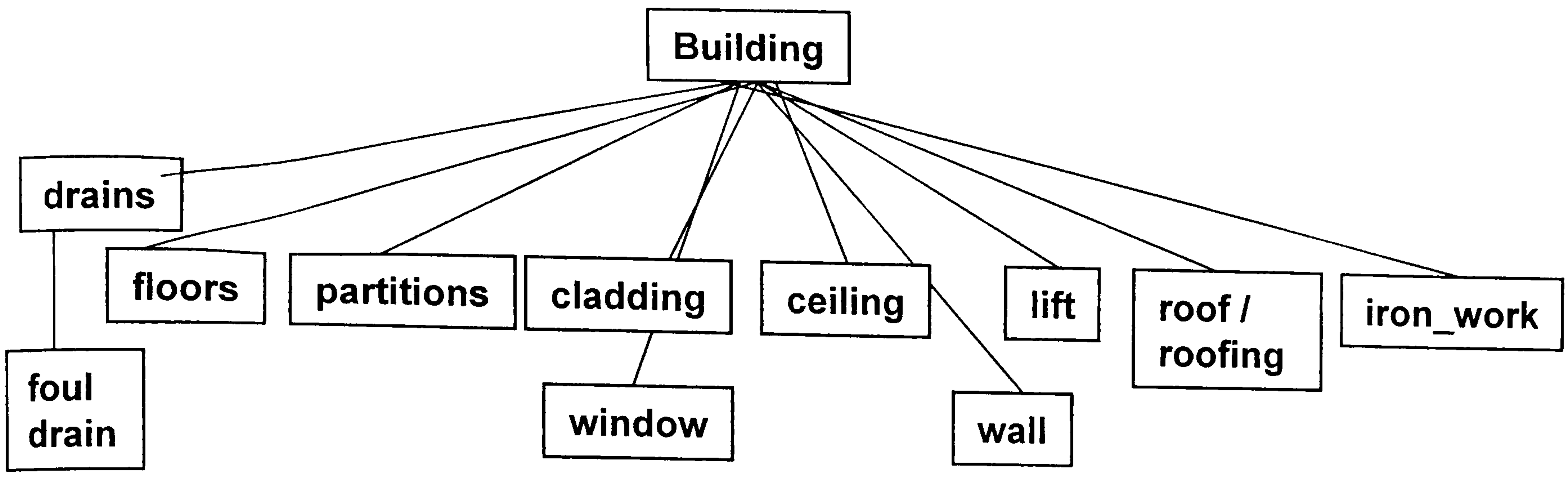


COMMUNICATION

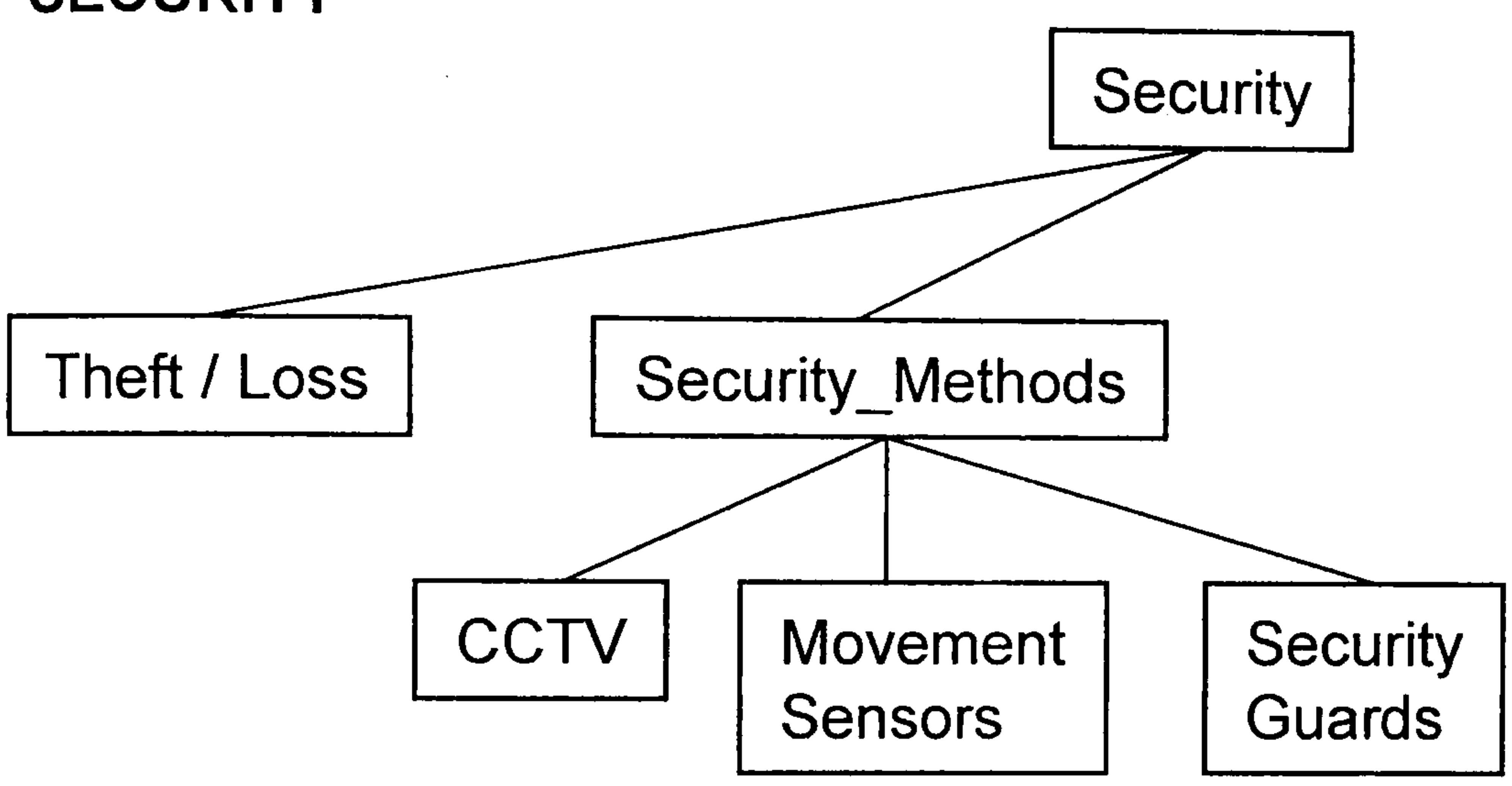


Appendix - 6

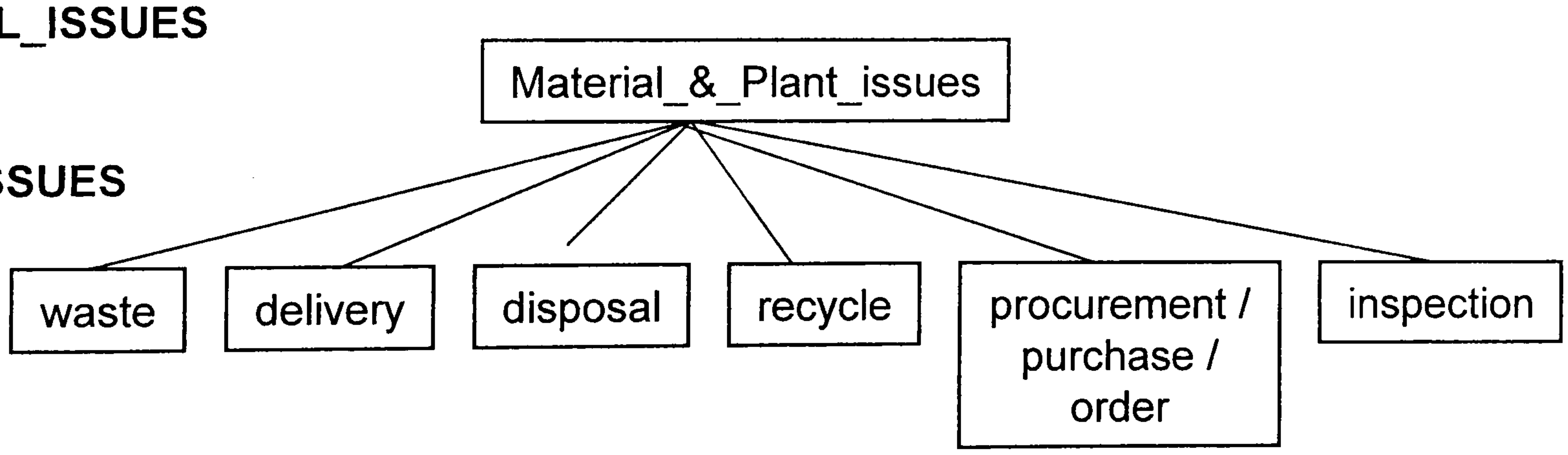
BUILDING



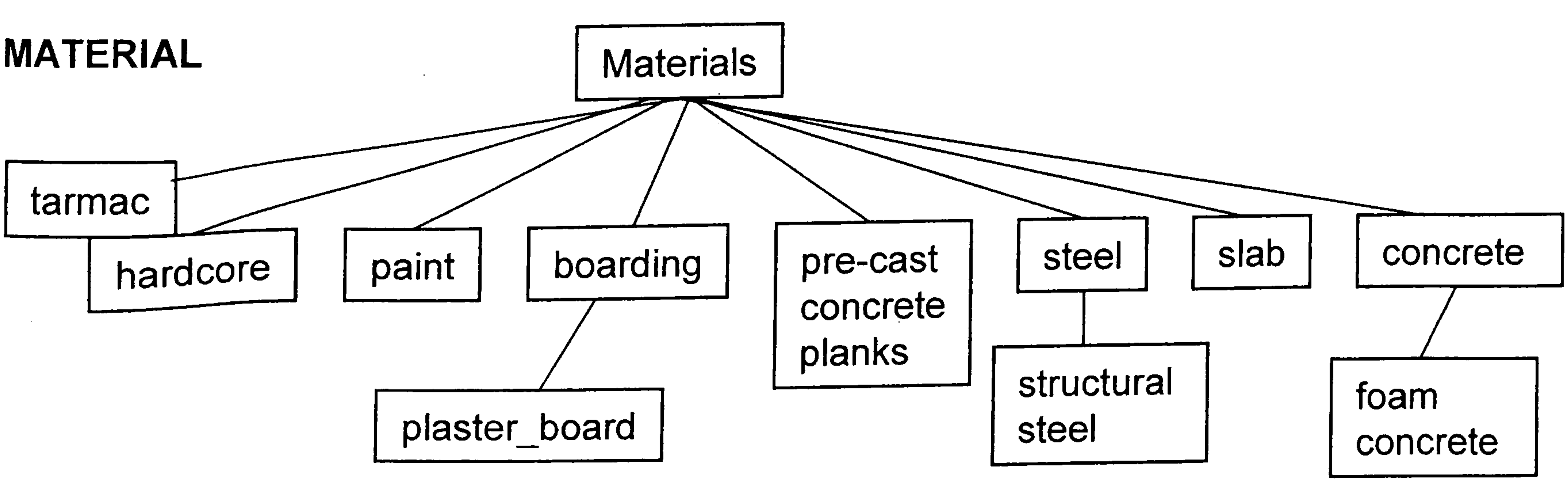
SECURITY



MATERIAL_ISSUES AND PLANT_ISSUES

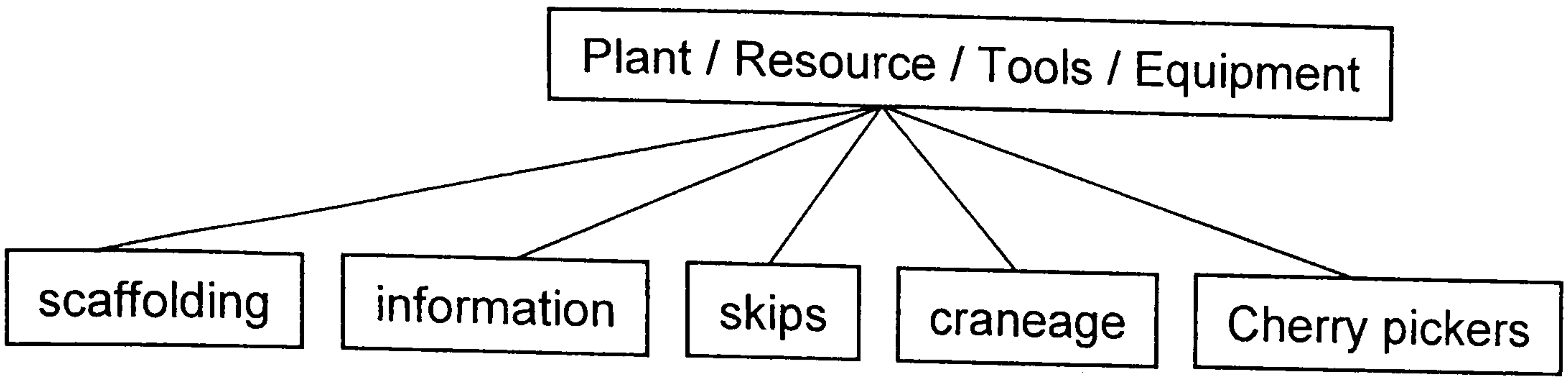


MATERIAL

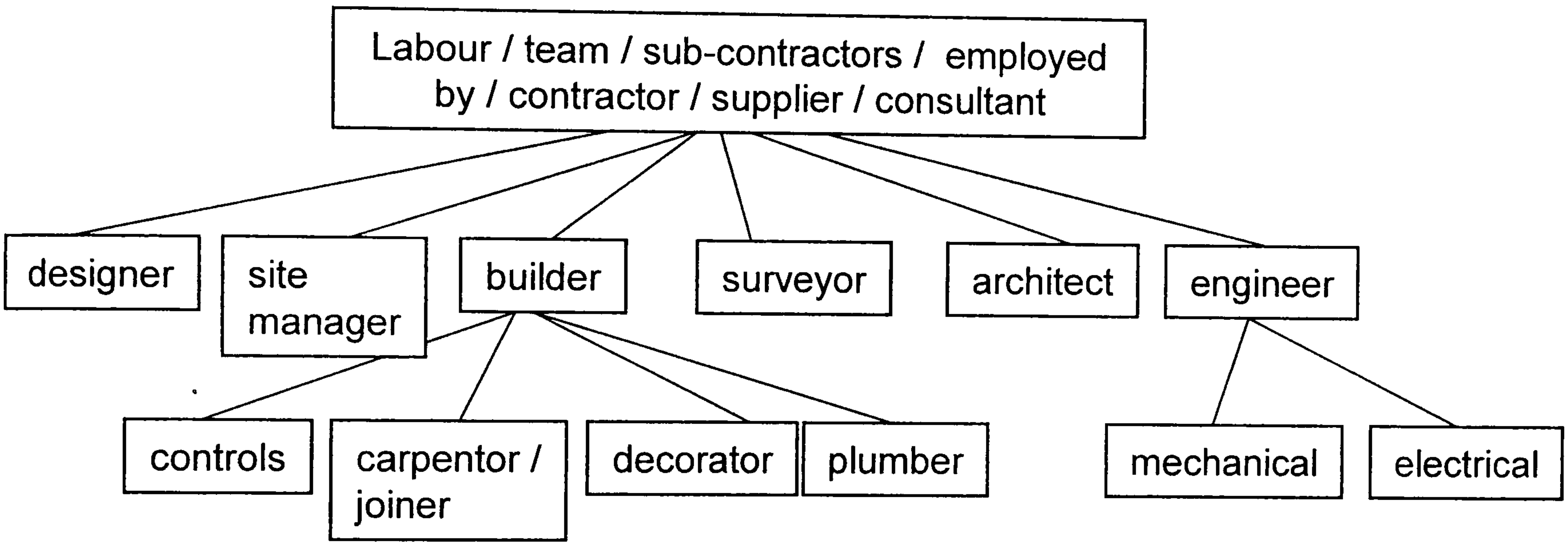


Appendix - 6

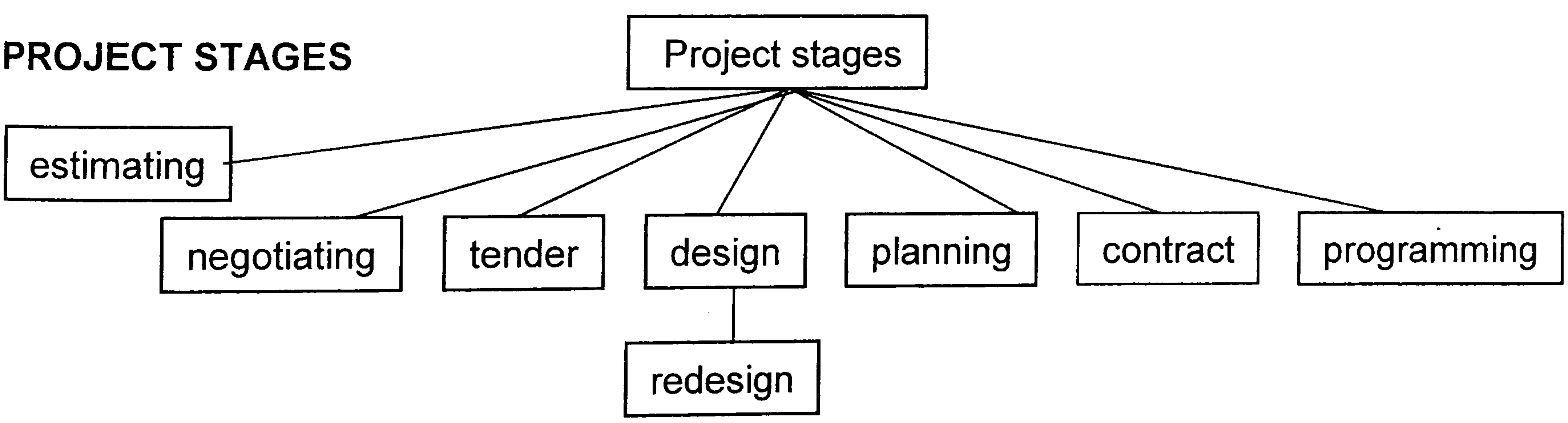
PLANT



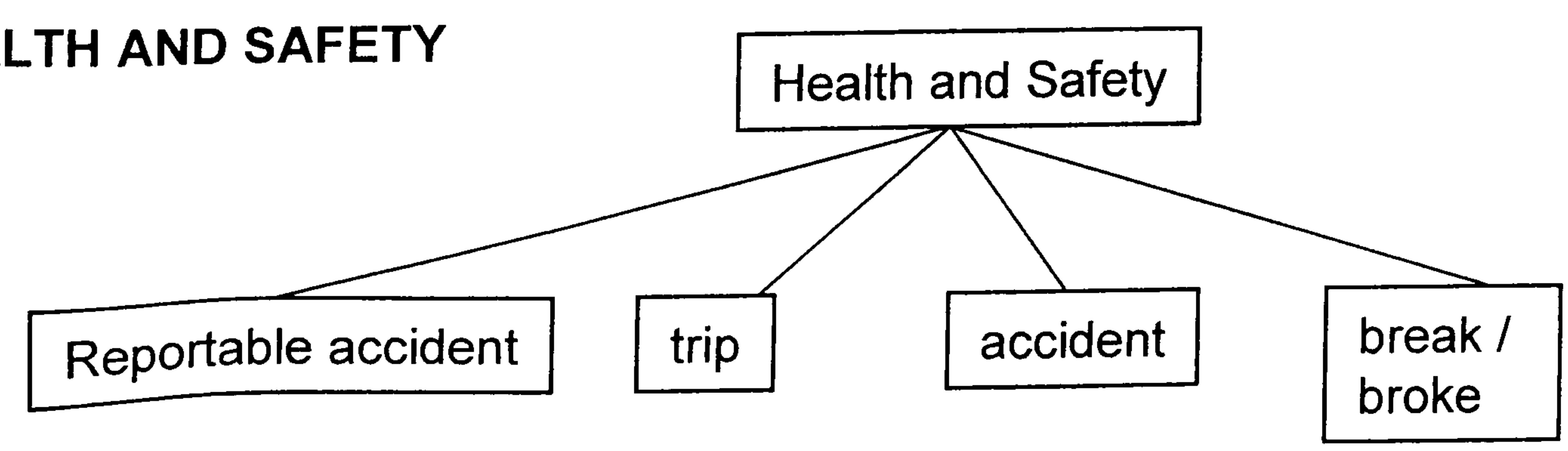
LABOUR



PROJECT STAGES



HEALTH AND SAFETY



Appendix - 7

Knowledge Discovery For Moderating Collaborative Projects

A. K. Choudhary, J. A. Harding*
Mechanical and Manufacturing Engineering
Loughborough University
Leicestershire; LE11 3TU
UNITED KINGDOM
A.K.Choudhary@lboro.ac.uk
*Communicating Author: J.A.Harding@lboro.ac.uk

K. Popplewell
Head of Engineering Manufacturing and Management
School of Engineering, Coventry University
Warwickshire. CV1 5FB.
UNITED KINGDOM
K.Popplewell@coventry.ac.uk

Abstract—**Competitive advantage can be gained through targeted exploitation of proprietary knowledge of all types, but especially of expertise and experience. One of the major issues in multidiscipline collaborative projects is how best to share and simultaneously exploit different types of expertise, without duplicating efforts or inadvertently causing conflicts or loss of efficiency through misunderstandings of individual or shared goals. Hence the challenges are enormous when considered in the context of distributed decision making from concept design, through product and manufacturing system design, volume production and on into obsolescence. Moderator Technology, in the form of knowledge based software support systems, have been successfully demonstrated in both the product and manufacturing system design domains. However, knowledge acquisition, learning and updating of knowledge is yet to be fully studied. This paper presents a knowledge discovery framework to support knowledge acquisition for moderator technology in collaborative projects.**

I. Introduction

A BUSINESS can differentiate itself from its competitors and compete efficiently and effectively, through well-targeted exploitation of its knowledge and expertise. Knowledge exists in all business functions, including purchasing, marketing, design, production, maintenance and distribution, but knowledge can be notoriously difficult to identify, capture, manage and reuse [1]. Hence, manufacturing enterprises are faced with a two-fold knowledge challenge; (1) To thoroughly understand and efficiently capture their valuable knowledge and expertise, and (2) to ensure that such knowledge is effectively reused and exploited to gain the best possible competitive advantage. This paper presents a knowledge discovery and data mining (KD/DM) framework for learning, sharing, and reuse of experience and expertise within multiple discipline project teams using the concept of a manufacturing system engineering moderator (MSEM).

A manufacturing system may be engineered or re-engineered for a variety of different reasons. Projects may range from partial or comprehensive overhaul of existing resources to a complete design of new manufacturing facilities and systems, and are generally executed by multidisciplinary project teams. The manufacturing system must satisfy many different requirements so compromises generally have to be made to achieve a balanced design for the new or re-engineered facility. Project team members must be aware when decisions that they make may affect other team members. The complexity of this problem increases when manufacturing projects are large and members are located in multiple sites, as the greater "awareness" that comes with "face to face" meetings can be very difficult to achieve. Therefore, an intelligent support system such as moderator technology is necessary [2,3]. Moderators have been utilized for several types of collaborative team projects, including both product and manufacturing system design. Recently they have also been applied in the context of globally extended manufacturing teams, where the challenges of knowledge sharing and awareness of available expertise were addressed through the adoption of a manufacturing system engineering ontology to enable semantic interoperability across the extended project teams [4,5,6].

To date, all moderators have been designed and implemented as specialist software systems, consisting of a moderation module, multiple expert modules and a knowledge acquisition module. Until now, all knowledge acquisition for the prototype moderators has been done manually, based on human expertise and experience [3,4]. There is however substantial potential for moderators to "learn" and update themselves from knowledge discovered in the existing operational databases of manufacturing companies. Our data mining case studies have already shown that relevant knowledge for future designs and redesigns can be identified by exploring operational data collected during product manufacture [7,8]. This paper therefore introduces a new KD/DM framework and proposes that the moderator's knowledge acquisition module should incorporate "learning", updating and reuse elements which exploit knowledge discovery techniques.

II. Moderator Concepts

A. *Multidiscipline Design*

The engineering moderator concept was first proposed during the MOSES[a] research project as a support tool for product design project teams [2]. It was coordinating software for Concurrent Engineering design, to raise awareness among the inter-working cross disciplines that

[a] *MOSES- Model Oriented Simultaneous Engineering Systems- EPSRC Project Number (GR/H24273)*

0-7803-9701-0/06/$20.00 ©2006 IEEE

1

gift journ@l
Global Journal of e-Business & Knowledge Management
2007, Vol. 3, No 1, pp 1-8

Engineering Moderator to Universal Knowledge Moderator for Moderating Collaborative Projects

A. K. Choudhary
Wolfson School of Mechanical and Manufacturing Engineering
Loughborough University, Loughborough, Leicestershire, UK, LE113TU

J. A. Harding
Wolfson School of Mechanical and Manufacturing Engineering
Loughborough University, Loughborough, Leicestershire, UK, LE113TU

H. K. Lin
Department of Industrial Engineering and Management
I-Shou University, Kaohsiung, Taiwan ROC.

Abstract

One of the major issues in multi-discipline collaborative projects is how to best share and simultaneously exploit different types of expertise, without duplicating efforts or inadvertently causing conflicts or loss of efficiency through misunderstanding of individual or shared goals. This research discusses the development of moderators from the initial concepts of an engineering moderator to ongoing research into universal knowledge moderator on semantic web. Moderators are knowledge based specialist intelligent software systems which support each individual of a team to perform his particular role from a position of strength, using his preferred method of working whilst understanding the need of other team members and the whole team working on a collaborative project. Current research proposes an improved framework based on semantic web and knowledge discovery for moderators to enhance globally collaborative e-manufacturing work through better interoperability and efficiency. A universal knowledge moderator prototype system consisting of an enterprise data integration module, knowledge discovery module and moderation module is developed to significantly improve the moderation activities and performance of collaborative e-manufacturing.

Keywords: Collaborative Project, Moderator, Semantic web, Knowledge discovery, E-manufacturing

Introduction

Current market trends indicate that demand is on the increase for highly customized products with ever shortening life cycle times and this trend is expected to accelerate. Manufacturing enterprise will achieve and sustain competitive advantage by improving productivity, responsiveness and flexibility. Manufacturers are striving to meet these demands by focussing on core competencies and migrating towards knowledge-based manufacturing [1].

A manufacturing enterprise can differentiate itself from its competitors and compete efficiently and effectively, through well targeted exploitation of its knowledge and expertise. Knowledge exists in all business functions, including purchasing, marketing and design, production, maintenance and distribution, but, knowledge can be notoriously difficult to identify, capture, manage and reuse [2]. Mapping out where "knowledge" resides and identifying the conditions that foster its generation and re-use has become a necessity. The knowledge assets of an enterprise reside in many different places, e.g., knowledge bases, filing cabinets, peoples' minds and expertise, and are distributed across the enterprise. Therefore, enterprises need to know their knowledge assets and how to manage and make use of these assets to get maximum return [3]. Hence, manufacturing enterprises are faced with the following twofold knowledge challenge;

- To thoroughly understand and efficiently capture their valuable knowledge and expertise, and
- To ensure that such knowledge is effectively reused and exploited to gain the best possible competitive advantage.

In this regard, knowledge discovery, knowledge management and knowledge engineering are currently topics of importance to manufacturing researchers and managers intent on exploiting current assets.

Need for Engineering Moderators in Collaborative projects

A manufacturing system may be engineered and re-engineered for a variety of different reasons. Projects may range from partial or comprehensive overhaul of existing resources to a complete design of new manufacturing facilities and systems. Such projects are generally performed by multidisciplinary project teams. Project team members must be aware when decisions that they make may affect other team members. The strength and success of a team depends on how well each individual can contribute the

The Universal Knowledge Moderator for Globally Distributed and Collaborative e-Manufacturing

H.K. Lin
Industrial Engineering and Management,
I-Shou University, Kaohsiung,
Kaohsiung,
Taiwan ROC
hklin@isu.edu.tw

J.A. Harding, A K Choudhary
Mechanical and Manufacturing Engineering
Loughborough University
Leicestershire LE11 3TU, UK
J.A.Harding@lboro.ac.uk
A.K.Choudhary@lboro.ac.uk

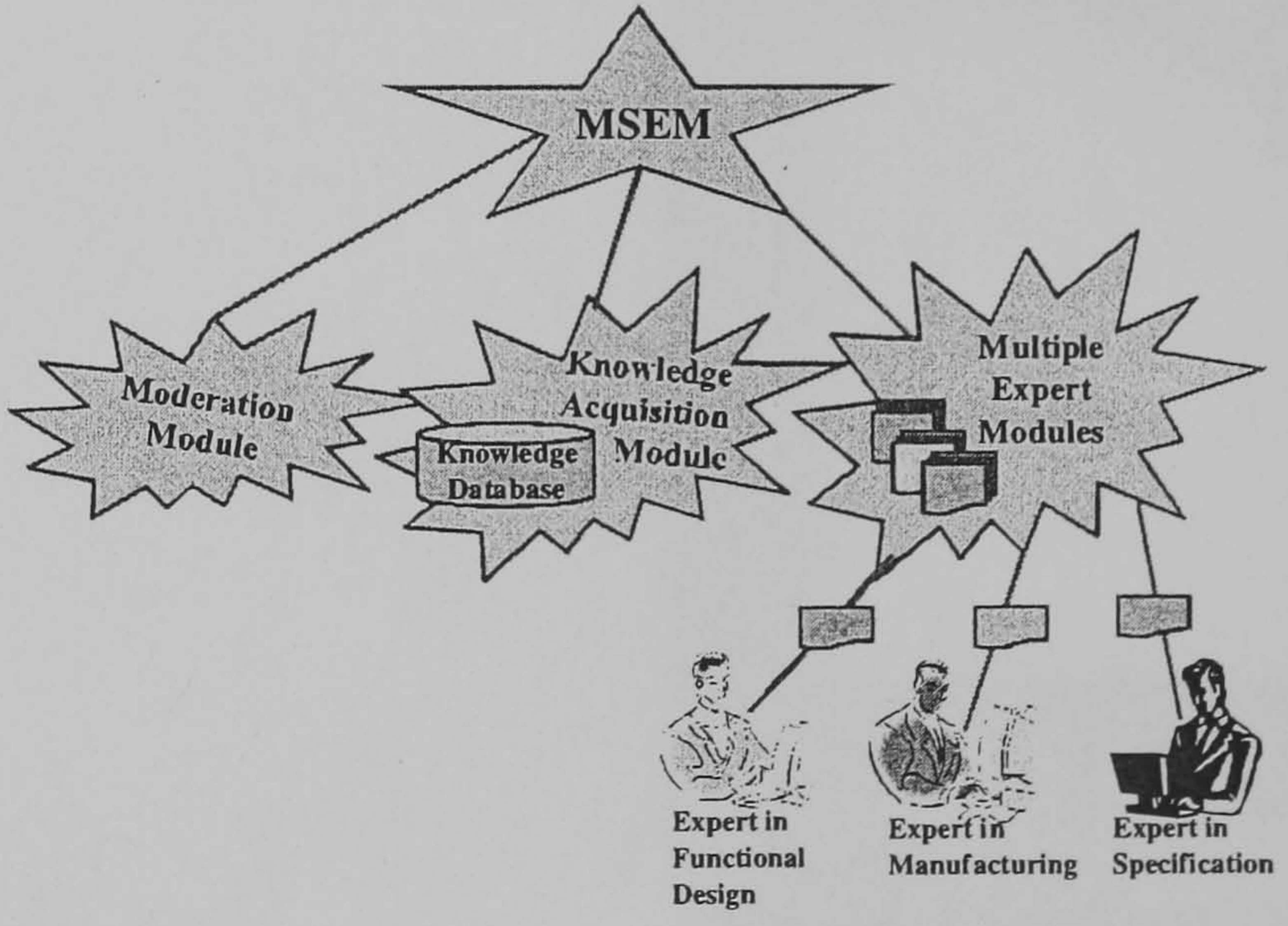
Abstract- **The new challenge of enterprise interoperability is to share knowledge inter-business via networks to enable all partners to benefit and WIN, WIN, WIN in their markets. The Universal Knowledge Moderator (UKM) project envisages the delivery of collaborative knowledge services which learn through Knowledge Discovery Module; exploit interoperability services through a Semantic Web based mediated schema within Universal Inter-Experts Module; and increase inter-team awareness whilst improving coordination through Moderation Module, to achieve the global visions of enterprise interoperability. The overall aim of UKM is to significantly enhance the moderation activities for universal knowledge cooperation, using inter-connected semantic web systems.**

I. INTRODUCTION

Manufacturing industry needed to cut its costs and overheads, and there were growing demands for flexible, fast, well-planned manufacturing system design. Manufacturing systems therefore started to move from distributed manufacturing and global manufacturing towards cross-organization e-manufacturing, using inter-connected web systems. The major goal of most manufacturing organizations is the development and adoption of inter-operational approaches that require companies to coordinate their activities effectively and efficiently across different enterprise and discipline boundaries.

One of the major issues in multidiscipline distributed and collaborative projects is how best to share and simultaneously exploit different types of expertise, without duplicating efforts or inadvertently causing conflicts or loss of efficiency through misunderstandings of individual or shared goals. Therefore, it is necessary to raise the awareness among the project teams' members working on a collaborative project to reduce or avoid the conflicts and misunderstandings that inevitably occurs in collaborative projects [1].

Moderator technology, which encompasses intelligent support systems, provides a possible approach to resolving these issues. The concepts of a Moderator and Manufacturing System Engineering Moderator (MSEM) to support a MSE team have been suggested and previously reported in MOSES [2] and MISSION [3] research projects. The main function of a moderator is to support a design group or team by raising individual members' awareness of the needs and experiences of other team members and this concept has been successfully demonstrated in both the product and manufacturing system design domains. Moderator Technology in the form of knowledge based software support systems, consisting of Multiple Expert Modules, a Knowledge Acquisition Module (KAM) and a Moderation Module (shown in Figure 1). The complexity of moderator technology increases when manufacturing projects are large and members are globally distributed in the context of Extended Enterprises (EE)/ Virtual Enterprise (VE). In such circumstances, the core functionality of the moderator remains the same, but its knowledge must be extended to provide greater understanding of the extended environment in which it is operating.



H. K. Lin, J. A. Harding and A. K. Choudhary, "The Universal Knowledge Moderator for globally distributed and collaborative e-manufacturing," *Industrial Informatics, 2008. INDIN 2008. 6th IEEE International Conference on*, pp. 1227-1231, 2008