A CASE-BASED REASONING APPROACH TO IMPROVE RISK IDENTIFICATION IN CONSTRUCTION PROJECTS

By
Yan Tan

Submitted in accordance with the requirements for the

Degree of Doctor of Philosophy.

The University of Leeds

School of Civil Engineering

Construction Project Management Group

March 2006

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

To My Family

ACKNOWLEDGEMENT

I would like to express my sincere gratitude and appreciation to my supervisors, Professor Nigel J. Smith and Dr. Denise A. Bower, who provided professional guidance, technical commentary, and encouragement throughout all aspects of this research work. Without their help, this research would not have been carried out as smoothly.

A special thanks to the WUN committee and Professor Jeffery S. Russell that offered me the opportunity to visit the University of Wisconsin-Madison, USA. There I found useful resources that broadened the scope of my research. I was influenced by Professor Russell's persistent pursuit of excellence which motivated my thirst for knowledge. I have greatly benefited from suggestions through discussion with Professor Andy Hanna.

My thanks also go to MindBox, who provided at discounted rate, their excellent product ART*Enterprise which is as the fundamental shell for the model development of this research.

This research could not have taken place without the financial support of the ORS Award Scheme, the University of Leeds, and the School of Civil Engineering.

I also would like to acknowledge Dorothy Carr and Betty Vaughan, who provided support and help during my hard time. Without this support, I might not have been able to continue my study.

Finally I must thank my family for being always with me. Their unwavering support and undaunted faith provided me with the confidence to pursue and complete this endeavour. They were a tremendous source of support and encouragement along the way.

To my friends and colleagues, it was a great journey. Thank you all.

ABSTRACT

Risk management is an important process to enhance the understanding of the project so as to support decision making. Despite well established existing methods, the application of risk management in practice is frequently poor. The reasons for this are investigated as accuracy, complexity, time and cost involved and lack of knowledge sharing. Appropriate risk identification is fundamental for successful risk management. Well known risk identification methods require expert knowledge, hence risk identification depends on the involvement and the sophistication of experts. Subjective judgment and intuition usually from part of experts' decision, and sharing and transferring this knowledge is restricted by the availability of experts. Further, psychological research has showed that people have limitations in coping with complex reasoning. In order to reduce subjectivity and enhance knowledge sharing, artificial intelligence techniques can be utilised. An intelligent system accumulates retrievable knowledge and reasoning in an impartial way so that a commonly acceptable solution can be achieved.

Case-based reasoning enables learning from experience, which matches the manner that human experts catch and process information and knowledge in relation to project risks. A case-based risk identification model is developed to facilitate human experts making final decisions. This approach exploits the advantage of knowledge sharing, increasing confidence and efficiency in investment decisions, and enhancing communication among the project participants.

TABLE OF CONTENT

ABBREVIATIONS

AE	Art*Enterprise
AHP	Analysis Hierarchy Process
AI	Artificial Indigence
AMP	Association for Project Management
ANN	Artificial Neural Network
APL	Adaptable Program Loans
ASCE	American Society of Civil Engineering
BNFL	British Nuclear Fuels
BP	Backpropagation
BR	Board Report
BSI	British Standard Institute
CALIBRE	Candidate LIBrary Retrieval
CATRAP	Cost and Time Risk Analysis Program
CBR	Case-based Reasoning
CORBA	Common Object Request Broker Architecture
CPAS	Case-based Procurement Advisory System for Construction
DBMS	Database Management System
DDE	Dynamic Data Exchange
DI	Data Integrator
DLL	Dynamic Link Library
DSS	Decision Support System
EA	Environment Assessment
EIRR	Economic Internal Rate of Return
FIRR	Financial Internal Rate of Return
IRR	Internal Rate of Return
NPV	Net Present Value
WB	World Bank
EAP	Environmental Action Plan
EIA	Environmental Impact Assessment
ENPV	Economic Net Present Value
ENR	Engineering News Records

ERL	Emergency Recovery Loans
FH2	Second Fujian Highway project
FIL	Financial Intermediary Loans
FNPV	Financial Net Present Value
FPECD	Fujian Provincial Expressway Construction Directorate
FPTD	Fujian Provincial Transport Department
FY04	The 2004 report of status of projects in execution
GA	Generic Algorithms
GOC	Government of China
GTK	Graphic Toolkit
GUI	Graphic User Interface
IBRD	International Bank for Reconstruction and Development
IDA	International Development Association
IP	Indigenous Peoples Plan
ISDS	Integrated Safeguards Data Sheet
IT	Information Technology
ITC	International Tendering Company
KBS	Knowledge-based System
KM	Knowledge Management
LIL	Learning and Innovation Loan
LM	Machine Learning
MBR	Model-based Reasoning
MIS	Management Information System
MOC	Ministry of Communication
MOP	Memory Organization Packets
MS	Microsoft
MUKCA	Multi-User Knowledge Capture Administrator
NTHS	National Trunk Highway System
ODBC	Open Database Connectivity
PAD	Project Appraisal Document
PFI	Private Finance initiative
PID	Project Information Document
PMI	Project Management Institute
PRAM	Risk Analysis and Management for Project

RAPs	Resettlement Action Plans
RBS	Rule-based System
RPL	Resettlement Plan
SIL	Specific investment loans
SIM	Sector Investment and Maintenance loans
SP	Screen Paint
TAL	Technical Asistance Loan
TCP/IP	Transmission Control Protocol / Internet Protocol
V&V	Verification and Validation
WACC	Weighted Average Cost of Capital
WB	World Bank
WBS	Work Breakdown Structure
ZMECD	Zhangzhou Municipal Expressway Construction Directorate
ZZE	Zhangzhou-Zhao'an Express

LIST OF FIGURE

LIST OF TABLE

CHAPTER 1

INTRODUCTION

1.0 INTRODUCTION

This thesis examines research conducted over a three-year period in managing knowledge in relation to project risks by utilizing case-based reasoning. The first section of this chapter gives a brief and initial overview of the general subject area for the research. This is followed by the author's justification of the purpose of the study. Next, the primary aim of the research is stated and the objectives that have been set for achieving this aim are highlighted, and for balance, an examination of the limitations and scope seen to apply to this study is made. The chapter concludes with a detailed outline of the structure of the report.

1.1 OVERVIEW

Construction projects are becoming increasingly complex because the components that are used are increasing in complexity. Projects are undertaken in complex technical, economic, political, and social environments and management must deal with a broad range of issues, requirements and problems in directing projects to successful conclusions. The increasing complexity of both projects and project environments cause many projects to fail to meet their expected budgets, schedules, functionality, and quality goals. Dalcher (1993) states that because of the increase in the complexity and scope of project, the ability to bring them to a successful completion dramatically decreases. To cope with these problems, it is necessary to carry out a risk analysis of the project at an early stage. Risk management is an important way to enhance the understanding of the scope and potential problems of the project concerned.

Research into risk management of construction projects started in the 1960's and thrived in the mid-1980, the period in which quantification analysis blooming up and achieved considerable success. Since the early 1990's risk analysis in project management has attracted more and more attention. A number of innovative techniques have been tested in systematic risk models for project management. Sensitivity analysis, Monte Carlo simulation, system simulation, fuzzy logic, AHP (Analytical Hierarchy Process) ANN (Artificial Neural Network), and Expert Systems are good examples.

Despite the number of methods being developed for project risk management, in practice the application of risk management is still limited. One reason for this is the accuracy and complexity of current analysis tools. Accurate analysis relies upon accurate risk factors involved in the analysis. Thus qualitative risk analysis should be given more attention as it determines whether or not the quantitative stage should be carried out and what the quantification variables are. Obviously, the main purpose of risk identification is to gain knowledge about uncertain situations and potential problems. For such work, present approaches largely depend upon the involvement of experts and their sophisticated experience and knowledge. However, the impact of human behaviour in terms of subjective judgement and psychological influence usually comes along with an expert's decision making. In order to manage knowledge in an appropriate manner in relation to project risks, reduce the subjectivity, and provide an impartial solution, an intelligent method that mimics human reasoning processes is considered as an appropriate approach in this research.

1.2 PURPOSE

The purpose of this thesis is to demonstrate an approach to identify risks in construction projects based on previous experience in similar projects. This approach simulates the human reasoning processes and provides a common knowledge platform to allow knowledge sharing among project participants in order to support decision making in project risk identification.

It is the author's expectation that this study will increase the exposure of risk management thus encouraging better adaptation of risk management techniques in construction projects, increasing the accuracy of estimates, increasing confidence and efficiency in investment decisions, and finally enhancing communication among project participants. As a result, it should make suggestions easier to accept and thus improve the ultimate performance of the project. This should bring about a change in focus of traditional risk management and arouse more attention on the basic issues of risk management, such as risk identification in the project appraisal stage. These may be particularly beneficial in coordinating the relationship between the project participants, simplifying the risk identification process, and promoting the application of risk management to the construction industry.

1.3 AIM AND OBJECTIVES

The aim of this research is to develop an IT based approach to facilitate risk identification in construction projects by using previous information and knowledge appropriately.

To overcome present application deficiencies and to achieve the above aim, the following measurable objectives have been set up as a guide to the study:

- to review project risk management, trace its development, and investigate current problems in application;
- to understand knowledge/knowledge management and its development, and evaluate the features of managing information and knowledge for project risk identification;
- to choose the most appropriate method and tool to cope with current problems;
- to develop an approach for project risk identification, collecting, processing, and sharing knowledge, and reasoning, in an impartial manner;
- to validate and verify the result of the research and any guidelines arrived at through this research; and
- to demonstrate developed model by applying it to a case study.

1.4 LIMITATIONS AND SCOPE

The undertaken research focuses on projects sponsored by the World Bank. After exploring a various available data, it was decided to focus on the appraisal stage because usually after a project has been sanctioned, 80% of the cost is frozen and the opportunity to reduce costs during the subsequent implementation phase is relatively small (Smith, 1999). Therefore, excellent appraisal can usually lead to successful project and risk identification is most useful in this stage. Transportation projects were selected because among all of the various World Bank supported projects, this sector was the closest to the construction industry and there were sufficient cases available for this study. The East Asian & Pacific Area and European Areas were chosen because the author is familiar with the background culture and policies in these areas. In addition, cases in these two areas would build a satisfactory case base with reasonable similarity and case coverage and the number of cases is sufficient for a model testing purposes.

The projects considered in this research were funded by the World Bank. Because of the differences in funding and policies, other transportation projects are not within the scope of this thesis. However, some of the fundamental outputs can be referenced, and relevant principles can be borrowed to build another model for applications outside the domain of this thesis.

In this research all reasoning is based on the World Bank project documents; other potential risks do not fall within the scope of this thesis. The identified risks of each project are individual to project. Therefore it is not possible provide a genetic risk list in each category, such as political, design, financial, environmental etc along with a probability percentage, as other risk identification methods do. The model is designed to provide identified risks and information for similar projects in order to support human experts' decision making.

1.5 OUTLINE OF THESIS

In order to present contextual thoughts in a comprehensible format, as well as new findings from this research, the thesis includes eleven chapters, which are illustrated in Figure 1.1.

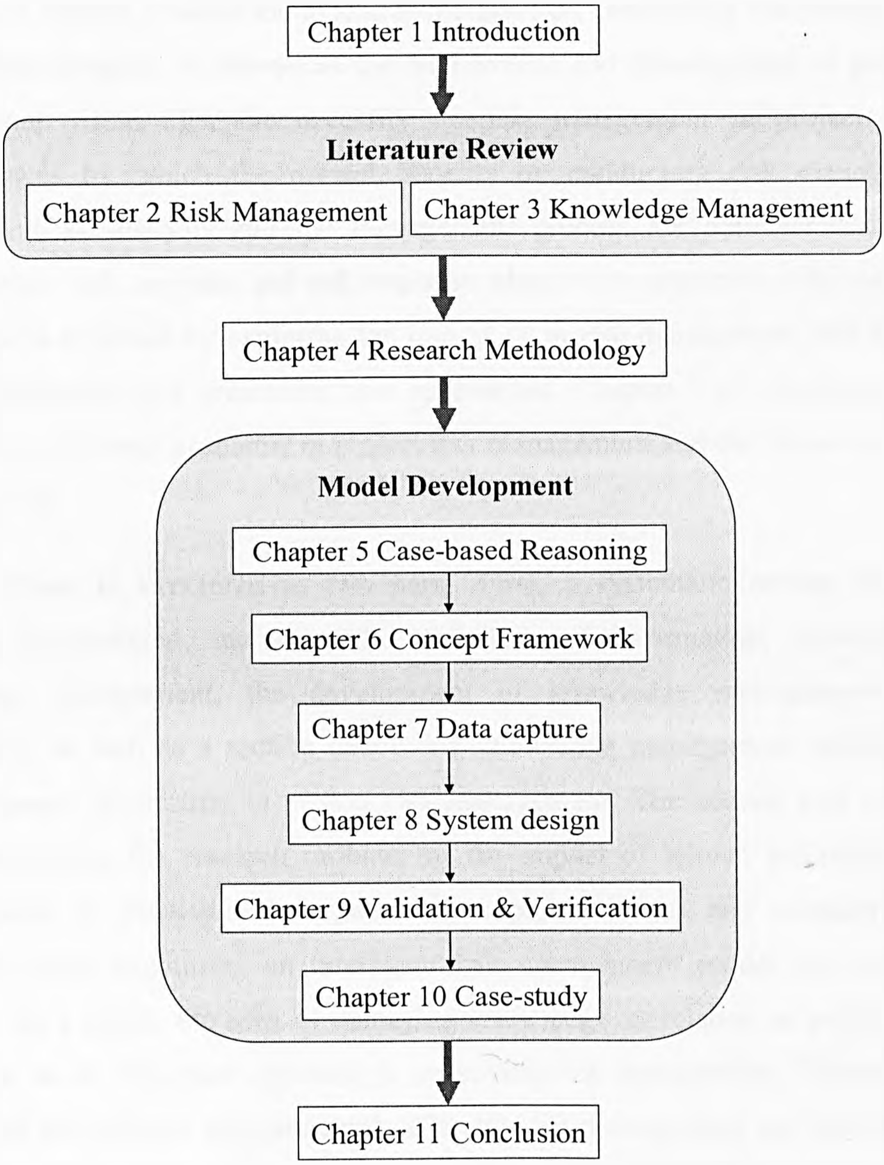

Figure 1.1 The structure of the thesis

These chapters have been divided into two sections. The first section, comprising Chapters One to Four, addresses the first two objectives, whilst the second section, consisting of Chapters Five to Ten, detail the practical side of this research which serves to realise other objectives and to further develop the first two objectives. Chapter One introduces the study, describing the aims and objectives and outlining the structure of the rest of the study, whilst Chapter Eleven concludes by bringing together the salient points of the review of project risk identification and the results of the analysis conducted in the later stages of this research.

The second chapter reviews the available literature on conducting risk management in construction projects. It introduces the background and development of project risk management. From this, the necessity of risk management in projects and the circumstances by which the project benefits by conducting risk management is explained. Risk analysis and the management process are demonstrated by: risk identification, risk analysis, and risk response, along with respective relevant methods used. This is followed by exploring the role of IT in risk management and evaluating current limitations and presenting new approaches. Chapter Two concludes with an elucidation of current problems in project risk management and the importance of risk identification.

Chapter Three is structured in two parts. First, a systematic review of relevant literature is provided, including the definition of information, knowledge and knowledge management, the development of knowledge management and its application, as well as a section examining knowledge management techniques and their proposed application in project risk management. The second part of Chapter Three formulates the research problem of the impact of human behaviour on risk management. It discusses the system development issues and considerations of relevance when structuring an intelligent risk management model for construction projects. As a result, the idea of managing knowledge in relation to project risks is concluded as an important approach to improving risk management. There is a deep analysis of the features and processes of knowledge management and their reflection on processes of case based reasoning and a discussion on the appropriate method to solve current problems.

Chapter Four outlines the research methodology applied in order to achieve the aim and objectives of this research, as set out in section 1.3. It introduces the definition of research, and demonstrates a typical seven step process used to carry out research projects. The proposed research approach is then presented, from the formulation of a research problem and managing the research data, to selecting the research methods. The latter section of this chapter introduces the research methods used to carry out this research.

Chapter Five introduces the mechanisms and approaches for the intelligent reasoning model. It compares other relevant methods and examines why the intended method is an appropriate choice. The reasoning method is described, in terms of theory, algorithm, features and relevant applications. Issues for consideration are highlighted and comparison of a range of packages is drawn. The latter sections discuss the approach and procedure for evaluating and selecting an appropriate modelling tool.

Chapter Six discusses the initial modelling framework and procedures. The requirements of the prototype model are specified. The structure of the intelligent model developed for the present report is then explained. The model is divided into five modules: criteria formulation, graphic user interface, knowledge base, CBR application, and data mapping. These modules are designed for various requirements to cope with different stages of the simulation. The following sections of Chapter Six explain these modules in detail.

In Chapter Seven the process of data capture, as adopted for this research, is presented. The types of data required are first identified and then classified. The framework of capturing research data is then introduced. The data collection process is discussed and some of the difficulties encountered are highlighted. The chapter goes on to outline the initial manipulation and interpretation of the data, and its use to derive a theoretical data structure, which is used for the rest of the study. Chapter Seven concludes by outlining the features of data resource, and identifying some of the key indicators used for assessing the intelligent model.

Chapter Eight provides a detailed description of how the intelligent model is designed. According to the system framework described in Chapter Six, and the data captured in Chapter Seven, it proposes five steps: data integration, case index, case representation case retrieval, and case adaptation. The detail of internal elements and structure is provided with both textual and visualised information, along with a description of how methods are chosen for the system development. Chapter Eight concludes with a summary of the intelligent model and its potential functionalities.

The development of an intelligent risk identification model to manage knowledge in relation to project risks as derived in Chapter Six is outlined in Chapter Eight. Chapter Nine reports on the verification and validation of the intelligent risk management model to establish confidence in the model's outputs. The role of verification and validation in research is initially introduced and the approach adopted to verify and validate the prototype model is outlined. The chapter goes on to discuss the results of the verification and validation techniques as applied to the model. This Chapter concludes with an explanation of the validation made from the analysis within the Summary section.

Chapter Ten provides a case study in terms of the content of the knowledge base, the operation process of the intelligent model and relevant guidelines for using the model. The stages of development are introduced and the processes that users have to go through within each stage are outlined. It then explores the outcomes of the simulations. The analysis is conducted and reported together with the simulations. Chapter Ten is concluded by outlining the overall simulation result, critiques and potential problems of the system.

This thesis is concluded in Chapter Eleven. It explores the results reported in the previous chapters, and conducts an in-depth analysis of the recorded outcomes of the simulations. A conclusion is then drawn from the research undertaken. Here the elements of the thesis, resulting from realisation of the objectives set out earlier in this chapter, are highlighted. A summary of the outcomes is made for each objective and the implications outlined. The chapter demonstrates how the aim of this research is achieved by realising its objectives. Chapter Eleven concludes with details of areas identified during this research that may provide potential avenues for further research.

CHAPTER 2

RISK MANAGEMENT

2.0 INTRODUCTION

This chapter provides an overview of risk management in construction projects. Firstly, the definitions of risk and risk management are explored. Then the need for risk management in construction projects is explained by exploring current problems in the construction industry. Four well known risk management frameworks are then introduced, followed by a description of relevant tools for the three main processes: risk identification, risk analysis and risk response. Subsequently, the role of information technology in risk management is introduced. The latter section of this chapter discusses the problem of applying risk management in practice.

2.1 RISK DEFINITIONS

The word risk derives from the Latin risicare, "to dare", which came to the English language in the mid 17th century. In the second quarter of the 18th century the anglicised spelling began to appear in insurance transactions (Flanagan and Norman, 1993). There is no commonly accepted definition for the term "risk". All risk concepts have one element in common, the distinction between reality and anticipation, e.g.

"Risk is the potential for the occurrence of undesired, negative consequences of an event" (Rowe, 1988).

"Risk is the probability that an adverse event occurs during a stated period of time" (Royal Society, 1992).

These two definitions focus on the negative effect of risk. However, in some circumstances risk elements may change from hazard to opportunity. When managed professionally, risk taking can provide real opportunities to maximise potential benefits for all concerned, and yield higher profit and/or benefit returns than low risk events. The following are definitions of risk for construction projects that consider both sides of risk.

Association of Project Management (2002): "a combination or frequency of occurrence of a defined threat or opportunity and the magnitude of that occurrence".

Project Management Institute (2000): "is an uncertain event or condition that, if it occurs, has a positive or a negative effect on a project objective".

Broadly speaking, risk can be classified into four components: risk, opportunity, uncertainty and complexity. The definitions from APM and PMI talk about risk as the negative factor and opportunity as the positive factor. The risk associated with complexity can be quite substantial. Because of the broad scope of complexity, more and more actors are involved, either directly or indirectly, and complexity goes hand in glove with fundamental uncertainty.

The definition of risk in construction projects also concerns uncertainty, such as:

HM Treasury (2000): "the uncertainty of outcome, within a range of potential exposures, arising from a combination of the impact and probability of potential events".

BS 6079-3 (British Standards Institution, 2000): "is the uncertainty inherent in plans and the possibility of something happening that can affect the prospects of achieving business or project goals".

Hertz and Thomas (1983) determined the relationship between risk and uncertainty, which is expressed as: *Risk= Uncertainty + Damage.* Chapman and Ward (2002) defined uncertainty as "a lack of certainty; involving variability and ambiguity".

Taken altogether, risk contains a broad range of understanding. It not only represents hazards and threats, but also embodies opportunity. It contains uncertainty from various sources, which give rise to and shape risk. The complexity of risk associated with uncertainty exploits a broader connotation. As a result, to understand the risk in construction project management, it is necessary to keep a dynamic and flexible vision, and to bear mind its negative, as well as its positive, variable and complicated features.

The definition of project risk management is given by several authors. BS 6079-3 (British Standard Institute, 2000) defines risk management as the "systematic application of management of policies, procedures, and practices to the tasks of analysis, evaluating and controlling risk". Smith (2003) states "risk management is about communication to make better decisions on a real project under conditions of uncertainty". Risk management refers to the process of reducing the risks to a level deemed tolerable by the project manager and to assure control, monitoring, and communication (Morgan, 1990).

In general, risk management in a construction project embraces the inclusion and the evaluation of potential risks with analytical and quantitative approach combined with a real understanding of probability and uncertainty; the pro-active control of significant risks and threats to the achievement of project objectives; and harnessing the experience and knowledge of the entire management population in a project to anticipate and overcome risks. It is the best way to manage uncertain components, control the negative effects, discover and create the potential opportunity, and save the project from overrun, delay and unsatisfactory quality.

2.2 *THE NEED FOR RISK MANAGEMENT*

2.2.1 Risk in Construction

A construction project is complex. Depending on the scale of a project, the project team usually comprises a client, an architect, a structural engineer, a building service engineer and miscellaneous members such as health and safety regulators and other specialist subcontractors (Anumba, *et al.*, 2003). Each team member is in charge of a variety of tasks and resources. The relationship between members of the project team is often complex and can lead to a difficult working environment.

The construction industry and its clients are widely associated with a high degree of risk due to the nature of construction business activities, processes, environment and organization (Akintoye and MacLeod, 1997), as its components are continually faced with a variety of situations involving many unknown, unexpected, frequently undesirable and often unpredictable factors (Fong, 1987). The process of taking a project, from initial investment appraisal to completion and into use, is complex and entails time. It requires a multitude of people with different skills and interests and the co-ordination of a wide range of disparate, yet interrelated, activities (Flanagan and Norman, 1993). Such complexity, moreover, is compounded by many external, uncontrollable factors, which may cause delays to the schedule, overrun of costs, and/or influence the quality of the project.

A survey carried out by Baldwin et al (1999) found that in UK 52% of all construction projects end with some type of claim. In 1994, Hartman (1997) introduced a study, of more than 8,000 projects conducted by the Standish Group, which found that only 16 were able to satisfy the famous triple constraints of project management: to get the job done on time, within budget, and according to specifications. According to the World Bank's 1999 annual review, 63% out of 1778 projects funded by the World Bank between 1974 and 1998 had experienced significant cost overrun.

Consequently, the construction industry has a poor reputation for coping with risk, with many projects failing to achieve their cost and schedule goals. It is generally agreed that risk management can help control the time and cost-overrun of construction projects.

All projects involve risks of various kinds (ICE & FIA, 1998). A study on perceptions and tendencies in risk, undertaken in the United States, compared a survey of ENR's (Engineering News and Record) top 100 US contractors of 1992 with an ASCE (American Society of Civil Engineering) survey concluded in 1997. The results indicate that certain risks are consistently associated with either contractor or owner. Moreover, the engineer and other relevant parties, such as the insurance company, are all involved in the project risk.

In recent times there have been several examples of major projects which have cost up to twice the budgeted amount to construct, opened twelve months late, performed well below the specified level of reliability, and generated less than 50% of the forecast annual revenue. Consequently, governments, funders, and lenders have become extremely reluctant to accept the risks inherent in such investments. When considering investment in terms of their overall lifetime performance, measurements are made against the original objectives. If these objectives are not achieved, the project as a whole will usually have failed (ICE &FIA, 2000).

Akintoye and MacLeod (1997) conducted a survey based on general contractors and project managers, the construction industry's perception of risk associated with its activities, and the extent to which the industry uses risk analysis and management techniques. They concluded that risk management is essential to construction activities in minimizing losses and enhancing profitability.

Risks and uncertainties exist in the whole of the project life cycle, and are related to all construction participants. The implication of several of the risks likely to be encountered in construction projects are illustrated in Figure 2.1. The single-line investment curve represents the 'most likely' outcome of the investment. An idea of the spectrum of uncertainty arising from estimates and predictions is shown in the shaded area. All uncertainties, particularly those which cause delay, will affect investment in the project.

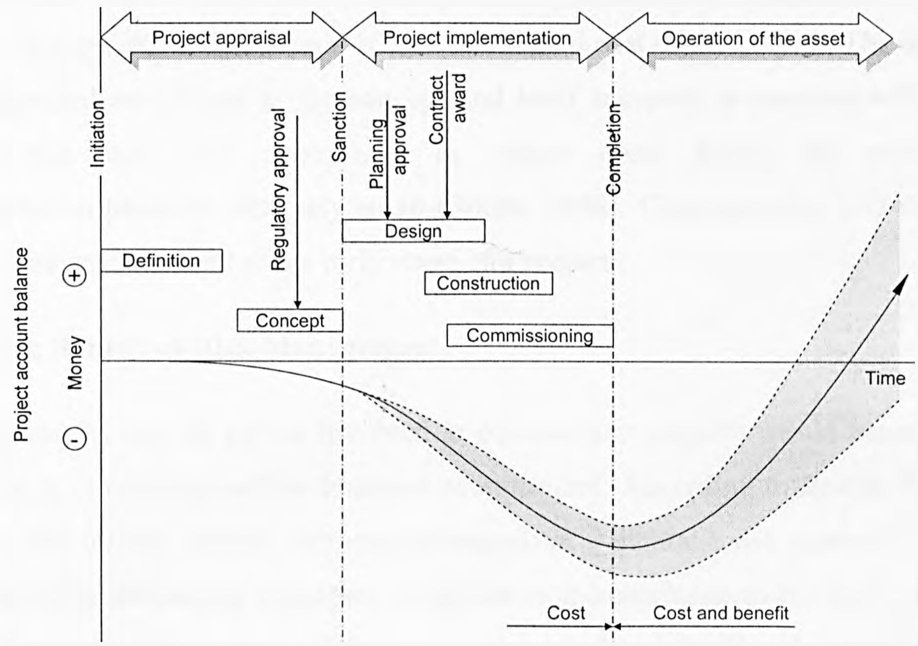

Figure 2.1 Project life cycle: cash flow (Smith, 1998)

Project appraisal or feasibility study is an important stage in the evaluation of a project. It is necessary to consider alternatives and to identify and assess risks at a time when data is uncertain or unavailable (Smith, 1998). The successful project is dependent on the effort expended during the appraisal phase preceding sanction. Figure 2.2 plots the percentage cost against time, showing how the important decisions for any project are made at the start of that project.

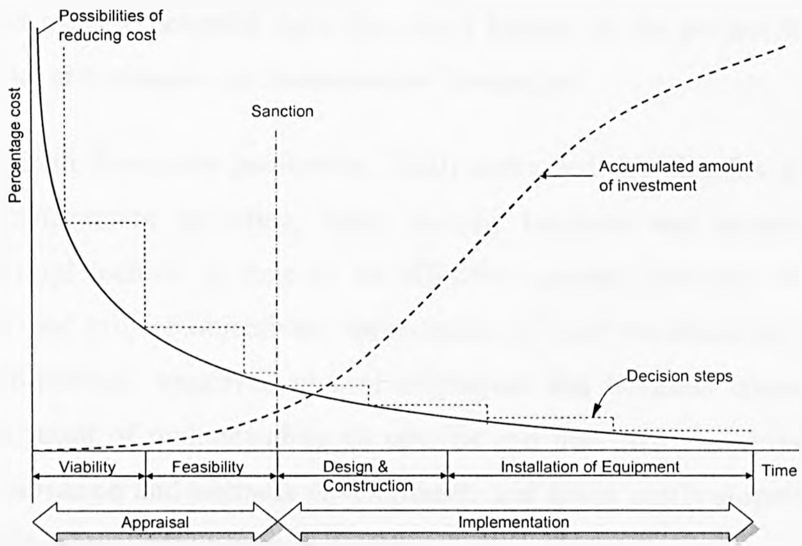

Figure2.2 Percentage cost plotted against time (Smith, 2002)

As shown in Figure 2.2, the percentage of expenditure to appraisal of major engineering projects rarely exceeds 10% of the capital cost of the project. The outcome of the appraisal as defined in the concept and brief accepted at sanction will freeze 80% of the cost. The opportunity to reduce costs during the subsequent implementation phase is relatively small (Smith, 1998). Consequently, it is worth to carry out risk management at the early stage of a project.

2.2.2 The Benefit of Risk Management

It is well known that all parties involved in construction projects would benefit from reductions in uncertainty before financial commitment. According to British Standard Institutes (BS 6079-3, 2000), "without the benefit of systematic risk analysis, it is not always possible for project managers to exploit their knowledge to the full". Simister (1994) investigated the usage and benefit of project risk analysis and management in 1992. According to Williams (1994), risk management provides two benefits. First, it provides a mechanism for analysis and management of risk. Second, it provides a common language for a means of communication across a project with complex structure. Construction projects usually involve long timescales, many uncertainties, and complex relationships among the participants. To enhance the assessment of the project, potential risks that could happen in the project life cycle should be identified and analysed as early as possible.

To identify and analyse potential risks that could happen in the project life cycle as early as possible, can enhance the assessment of the project.

BS 6079-3 (British Standards Institution, 2000) addressed the benefits gained from project risk management including: more realistic business and project planning; actions being implemented in time to be effective; greater certainty of achieving business goals and project objectives; appreciation of, and readiness to, exploit all beneficial opportunities; improved control of project and business costs; increased flexibility as a result of understanding all options and their associated risks; greater control over innovation and business development; and fewer costly surprises through effective and transparent contingency planning.

Further potential benefits of using risk management techniques as discussed in AMP (2004) and RAMP (ICE & FIA, 2000) are:

- an increased understanding of the project, which in turn leads to the formulation of more realistic plans, in terms of both cost estimates and timescales and the role of engineers and actuaries in risk analysis and management issues generally;
- an increased understanding of the risks in a project and their possible impact, which can lead to the minimisation of risks for a party and/ or the allocation of risks to the party best able to handle them;
- an understanding of how risks associated with a project can lead to the use of a more suitable type of contract;
- an independent view of the project risks which can help to justify decisions and enable more efficient and effective management of the risks;
- a knowledge of the risks in a project which allows assessment of the risks and which also tends to discourage the acceptance of financially unsound projects, so as to lead to better decisions and give greater confidence to investors and other stakeholders;

- a contribution to the build-up of statistical information of historical risks that will assist in better modelling of future projects and reducing risk exposures, and to the wider community of professionals and managers engaged in projects; and
- facilitation of greater, but more rational, risk taking, thus increasing the benefits that can be gained from risk taking.

2.3 RISK MANAGEMENT APPROACHES

Risk management is a systematic approach of management policies, procedures and practices applied to the tasks of establishing the context, identifying, analysing, evaluating, assessing, treating, monitoring and communicating risks in a way that will enable organizations to minimize loss and maximize opportunity in a cost-effective way (British Standard Institute BS IEC 62198, 2001). A risk management system should establish an appropriate context, set goals and objectives, identify and analyse risks, influence risk decision-making, and monitor and review risk responses (Edwards and Bowen, 1998).

2.3.1 Risk Management Frameworks

In order to carry out a sufficient risk analysis and make the risk management maximize benefit for the construction project, a variety of risk management frameworks have been developed. The most successful examples are: RAMP (Risk Analysis and Management for Project) by the Institution of Civil Engineering and the Faculty and Institute of Actuaries; APM's (the Association for Project Management) "Project Risk Analysis and Management"; Risk Management from British Standard 6079-3; and PMI's (Project Management Institute) guide of project risk management. In addition, the CRM Manual (1987) provides a procedural, task-based guide to construction risk management, as did Flanagan and Norman (1993) and Raftery (1994).

AMP's guide splits the overall process into two constituents: risk analysis and risk management. Risk analysis includes two sub-stages: qualitative analysis, which allows the main risk sources or factors to be identified and associated with some form of assessment; and quantitative analysis, which uses more sophisticated techniques to measure the uncertainty in cost and time estimates and probabilities combined in individual uncertainties. In this framework, risk management is more about risk response resulting from risk analysis.

PMI's systematic project risk management process is illustrated in Figure 2.3. The risk management content is related to the following processes: risk management planning, risk identification, qualitative and quantitative risk analysis, risk response planning, and risk monitoring and control. The details of each component are listed in the relevant box.

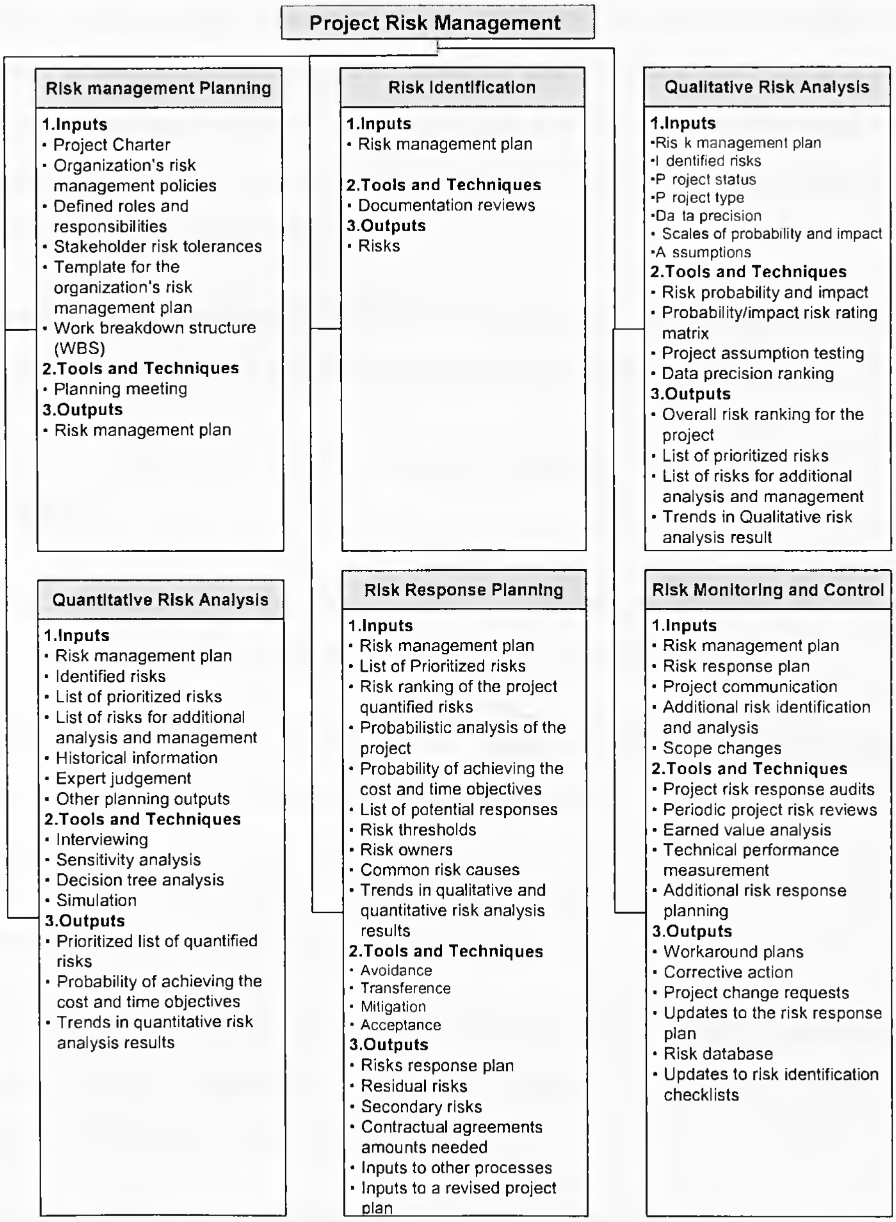

Figure 2.3 Project risk management overview (PMI, 2000)

RAMP is a comprehensive and systematic process for identifying, evaluating and managing risks in a capital investment project. It covers the entire life of a project from inception to close-down (ICE &FIA, 2000). The process consists of four activities: process launch; risk review; risk management; and process close-down. In process launch, the baseline objectives, scope and plans for the project are defined, as well as the underlying assumptions on which these are based. Risk review repeats key stages or decision points by systematically identifying risks and entering them in a risk register. It evaluates these risks to determine their likelihood and impact and potential relationships, and identifies mitigation measures to avoid, reduce or transfer risks. Risk management implements the mitigation strategy and risk response plan in risk review. It identifies and manages new or changing risks and designates individual charges to risks within their areas of responsibility. Finally, PRAM is closing down, when the initial objectives have been achieved.

The risk management process of BS 6079-3 (British Standard Institute, 2000) is a goal driven framework. It aims to answer five questions at different levels:

- What is at risk and why? - Context: business and project objectives and boundaries.
- What (and where) are the risks? - Risk identification: sources of risk, what are the risks, how do they arise, and grouping and associations.
- What is known about them? - Risk analysis: characteristics, classification, estimates of likelihood, and potential consequences.
- How important are they? - Risk evaluation: set criteria, decision ranking, and selection of priorities.
- What should be done about them? - Risk treatment: identify options, evaluate options, plan treatment measures, assess secondary risks, allocate responsibilities, and implement treatment.

Although these process models may differ in detail and the way they are drawn, they all tend to show a series of key activities. They are:

- identification of risk issues;

- analysis and assessment of the risks for their potential impact on the project;
- developing responses.

In a general risk management procedure, risk identification classifies and records each risk and qualifies the risk by reaching a unique description of the risk element; risk analysis estimates risk factor's likelihood of occurrence and potential impact on the project in terms of timescales, cost or quality; and risk response identifies the persona, team or company that will be responsible for the risk, plans risk mitigation strategies and executes a risk control plan. The following sections will explain these processes and their associated methods.

2.3.2 Risk Identification

Risk identification determines which risks might affect the project and documents their characteristics (PMI, 2000). Al-Bahar and Crandall (1990) defined risk identification as "the process of systematically and continuously identifying, categorizing, and assessing the initial significance of risk associated with a construction project". During risk identification, the collection and primary processing of a large volume of initial data is carried out. This results in a set of risk events, which are further classified and assessed. Therefore, risk identification is a fundamental step before risks can be analysed and an appropriate response can be determined. The objectives of risk identification are to compile a list of the main risk resources and a description of their likely consequences, perhaps including an initial approximation of the potential effect on estimates of cost and time (Perry, 1986).

There are a number of methods of risk identification, such as brainstorming, Delphi, checklist, structured interviews, questionnaires, review historical data, flow chart, fault tree, physical inspections, organizational charts, outside specialists or consultants, intuition, and hazard indices. Here the author only explains the major methods in detail.

Checklist

Checklist is a hazard identification technique which provides a listing of typical hazardous substances and/or potential accident sources which need to be considered

(British Standard Institute, BS 8444, 1996). Checklists are derived from the risks encountered in previous projects. Risk issues can be identified through an examination of what occurred on previous projects plus an overall understanding of the issues that are likely to be problematic on future projects. These issues can be formalized into lists and structured in a way that suits the particular type of project (Webb, 2003). New projects can then be examined against the list and an opinion formed on each point raised. There are three basic types of checklist: simple checklist/ ticklist, yes/no questionnaire, and descriptive checklist (Edwards, 1995). A checklist provides a convenient means for the project manager to quickly identify possible risks. It takes the form of either a series of questions or a list of topics to be considered. Checklists are an important vehicle through which an organization can record and capitalise on its experience (Merna, 2002).

Brainstorming

Brainstorming is one of the approaches that takes place among a group of knowledgeable staff. In an atmosphere of free speculation, people are invited to produce as many ideas as possible of the risks that might arise. This can lead to a very large number of ideas, some of which will be wild speculation. All risk ideas put forward have to be analysed and categorized into the those that are real and need to be either dealt with or monitored, and those that are largely imagined or are extremely unlikely and can be ignored (Webb, 2003). Originating in the 1950s as a business management tool, brainstorming involves the collective generation of ideas by a group comprising key project personnel and others in an environment free of criticism. The underlying principles are that group thinking is more productive than individual thinking, and the avoidance of criticism aids the production of ideas. Group members are encouraged to "free-wheel" in their thought association and build on each other's ideas. The most promising ideas generated are then selected, developed and verified (Merna, 2002). It is the most unstructured form of risk identification but can be invaluable in many circumstances (Edwards, 1995).

Delphi

Delphi is perhaps the best-known method of using group judgements in forecasting. The Delphi technique is a method for the systematic collection and collation of judgements from isolated anonymous respondents on a particular topic, through a set of carefully designed sequential questionnaires interspersed with summarised information and feedback of opinions, derived from earlier responses (Chapman, 1998). Smith (1998) addressed the Delphi procedure in risk identification: "it starts with the formation of a team of experts that represent all aspects of the project; the experts meet and formulate an exact definition of the risk that is being considered; they then discuss the risk, paying particular attention to its causes and the interdependencies it has within the project; subsequently, they give their opinions as to the probability of occurrence of the risk and the impact of the risk on the project, should it occur; the experts can also give a cost assessment of the risk based on the probability of occurrence and possible impact". Some of the benefits of the Delphi method are that participants may be remote from the risk management process such as conformance, personality characteristics, compatibility, and peer pressure (Merna, 2002).

Review of Historical Data

Review of historical data is a hazard identification technique that can be used to identify potential problem areas and also provide an input into frequency analysis based on accident and reliability data (BS 8444, 1996). It collects the previous project data, records, experience and failure, plus the project manager's knowledge, experience and judgement, to identify risks in the current project. It enables learning from previous experiences, both good and bad. Meanwhile historical data can provide an initial list of possible risks for brainstorming discussion.

2.3.3 Risk Analysis

Risk analysis is the systematic assessment of decision variables which are subject to risk and uncertainty. Risk analysis systematically uses the information available to characterize the risks, determine how often the specified events could occur, and then judge the magnitude of their likely consequences. The risk analysis process comprises: the establishment of probabilities of occurrence of adverse events; the setting of assumption of occurrence of adverse events, the setting of assumptive bounds to associated uncertainties, and the measurement of the potential impact of risk event outcomes (Edwards and Bowen, 1998).

There are a range of successful risk analysis methods suitable for different project characteristics and analysis purposes. The main techniques currently used for risk analysis are:

Probability Analysis

Probability analysis is the first risk analysis method which can be traced back to the 1650's. Pascal and Fermat laid out the founding principles of the theory of probability, which is the mathematical concept at the heart of the modern concept of risk. By 1730 de Moivre had proposed the structure of a normal distribution, and he was the first to define risk as a chance of loss (Smith, 2003). Published in 1763, two years after the death of the Reverend Thomas Bayes, Bayes' theorem is the basis for all inference problems using probability theory as logic. Stated as a simple equation, Bayes' theorem shows that

$$\text{Expected outcome of an event} = \sum_{i=1}^{n} P_i \times V_i$$

where $P_1 + ... + Pn = 1$, P_i and V_i are the outcome of probability and value of the Nth event respectively.

"Bayes' theorem can be applied by hand in decision tree calculations but, increasingly, all types of probabilistic analysis require the use of risk management software packages (Smith, 2003).

Probability analysis is used to provide information, such as estimates of the likelihood of achieving certain project targets and the likely range of outcome of the project, in terms of its duration and economic parameters. The situations of risk are defined as those in which the potential outcomes can be described in reasonably well known probability distributions (Haimes, 1998).

Sensitive Analysis

Sensitivity analysis and deterministic analysis checks what change in the value of a dependent variable occurs of the value if one or more variables that determine the dependent variable, changes. The sensitivity analysis should be done for all risks that may have a considerable impact on the project, in order to quantify the impact of those risks on the dependent variables. The sensitivity analysis can be used to identify the variables that need to be considered for the performance of a probability analysis. The main limitation of the sensitivity analysis is that in indication of the likely probability of occurrence of changes in key variables is given (Merna, 2002).

Monte Carlo Simulation

The most common form of risk analysis uses "sampling techniques", usually referred to as "Monte Carlo Simulation" (PRAM, 2000). Monte Carlo simulation is a process for developing data through the use of a random number generator. It should be used for problems involving random variables with known or assumed probability distributions. This technique requires the selection of different values from a probability distribution, the values corresponding to their probability of occurrence as defined by the probability distribution. In the analysis phase, the identified risk is modelled by estimating a pessimistic, a normal, and an optimistic value, known as a triangular distribution, although others can be used (Smith, 1998).

Influence Diagrams

The influence diagramming technique involves mapping out the project, identifying the sources of risk and possible responses to these risks. This information is then represented diagrammatically (Smith, 1998). Influence Diagrams provide a powerful means of constructing models of the issues in a project which are subject to risk. As a result they are now used as the user interface to a computer based risk-modelling tool thus allowing the development of very complex risk models that can be used to analyse the cost, time and economic parameters of projects (PRAM, 2000).

Decision Tree

Decision trees are diagrams that depict a sequence of decision and chance events, as they are understood by the decision-maker. The decision tree is made up of two types of nodes: decision nodes and chance event nodes. A decision node represents a decision that has to be made and a chance event node represents an event that has a chance of occurring, possibly a risk. A decision tree starts at a decision point node on the left-hand side and the information is conveyed going across the page from left to right. At the time represented by a specific node all prior decisions, or decisions at the time represented by a specific node, all prior decisions, or the decisions to the left of the node, have been made and uncertainties related to prior chance event nodes have been removed. Each decision node should have at least one branch, or arrow, coming from it and these branches represent the decision alternatives. The procedure for constructing a decision tree begins with the identification of action available at each of the decision points. Once this has been completed, it is necessary to identify the chance events points, or uncertainties, in the project and establish the possible alternative outcomes of each chance event. Finally, the decision tree should be evaluated to obtain the expected values for following each alternative course of action (Smith, 1998).

Bayesian Network

In contrast to the traditional definition of probability using relative frequencies, the Bayesian probability theory interprets probability as a degree of belief. It takes the view that probabilities are subjective i.e. they represent the strength of belief of an observer about whether certain events will take place. It is assumed that this observer has some prior beliefs that will change as a result of seeing the outcome of the experiment. Bayesian theory provides formalism for this transformation from prior to posterior belief (Anonymous, 1993). Bayesian inference can be used to enhance the integration of available numeric and subjective information (Englehardt, 1997), as prior belief can be quantified and incorporated with experimental evidence into a final probability assessment.

2.3.4 Risk Response

Once a clear understanding of the threats and opportunities facing a project has been established, the final phase of risk management, the risk response process, commences (Merna, 2002). Risk response is the action required to reduce, eliminate, or avert the potential impact of risks on a project. In many cases it is presented as a plan of action or strategy. A response strategy is normally identified for risks with a high or medium likelihood and impact assessment, or for those classed as unacceptable risks or critical risks. Figure 2.4 illustrates a risk response process.

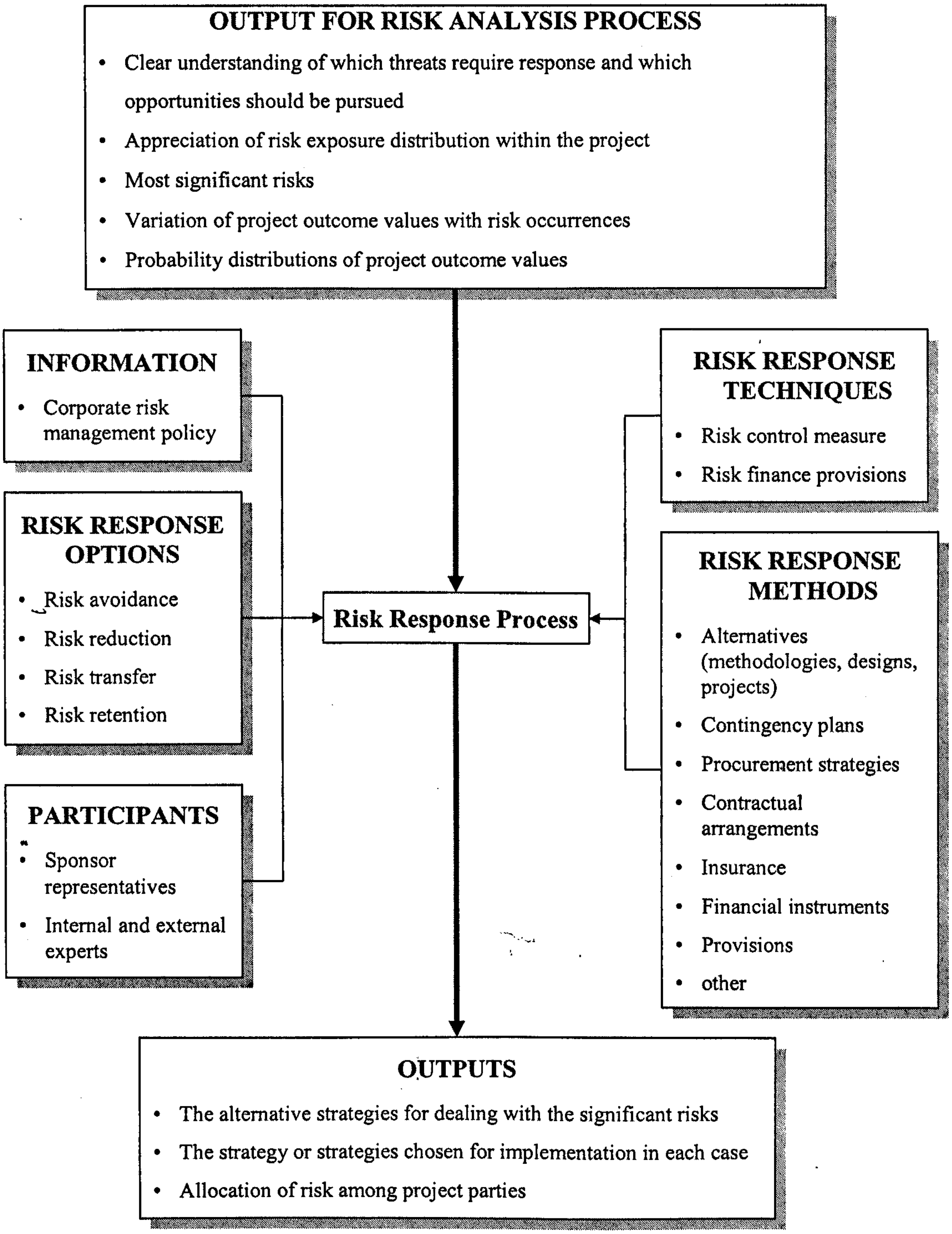

Figure 2.4 The risk response process (Merna, 2002)

The usual risk response approaches include risk avoidance, risk transference, risk reducing, risk management, and contingency fund.

Avoidance

Risk avoidance is mainly directed at removing the cause of the risk by analysing the situation to determine the elements that are creating the risk. When this cannot be achieved, the avoidance has to be approached with care as it can involve making decisions. This includes: deciding not to no-bid a project in which the risk exposure to the company is perceived to be too high; leaving the risk with the contractor when the contractor is believed to be in the best position to mitigate the risk, or that a more acceptable price may be achieved; the acceptance of an alternative lower risk technology path, which may lead to reduced quality performance, or to a less acceptable solution; insisting that escape clauses are included in the contract, which releases the supplier if the risk occurrence happens (Cater *et al.*, 1996).

Transfer

Generally, risk transfer means transferring the activity on to subcontractors who understand the subject and are more capable of identifying and mitigating the risks involved. The reason for doing this may be twofold: first, the organization is likely to encounter them in terms of specialist technology involved or, second, the organization is better able to sustain the effects of the risk if it should materialize, possibly because it is more secure financially or can rapidly bring resources to bear on the problem (Webb, 2003). Thompson and Perry (1992) stated that risk transfer can take two basic forms. The property or activity responsible for the risk may be transferred, for instance, hiring a subcontractor to work on a hazardous process; or the property or activity may be retained, but the financial risk transferred, for example by using services methods such as insurance. The most common form of risk transference is taking out insurance as here the financial effects of risk are transferred to an insurance company in exchange for a fee.

Reduction

The purpose of risk reduction is to take an action, generally a single action, that will reduce the risk to an acceptable level. Since the significance of a risk is related to both its probability of occurrence and its effect on the project outcome if it does occur, risk reduction may involve either lowering its probability or lessening its impact (Merna, 2002). An approach to risk reduction is to anticipate those risks that are amenable to some form of work that will alter the situation favourably and then include this work in the project plan. According to Smith (1998), risk can be reduced by obtaining additional information, performing additional tests/simulations, allocating additional resources, and improving communication and managing organisational interfaces. Within the reduction also lays the ability to procure information or the services of specialist consultants to support the project team in their activities. This path also addresses those risks that involve a choice of technology or programme activity where an extra cost may be involved, but the impact of the risk reduction will not affect other factors within the programme such as timescales or performance (Cater *et al.*, 1996).

Retention

Risk retention is the method of handling risks by the company who controls them. The risks, foreseen or unforeseen, are controlled and financed by the company or contractor that is fulfilling the terms of the contract (Baker *et al.*, 1999). There are two methods (Carter and Doherty, 1974), active and passive. Active retention is a deliberate management strategy after a conscious evaluation of the possible losses and costs of alternative ways of handling risks. However, passive retention occurs through neglect, ignorance or absence of decision. For example, a risk has not been identified and handling the consequences of that risk must be borne by the contractor performing the work. According to Flanagan and Norman (1993), the risks suitable for retention are those which occur frequently but result in small losses.

Contingency

The contingency provision is an additional a sum to the project estimate to cover aspects that are not known, such as those risks that are assessed to be of a low likelihood and impact and for risks that have not been revealed during the identification process, for instance, the residual risks. Contingency is often formalized in the project estimating process, and a structure defining who has control of each level of contingency may be set up (Webb, 2003). These types of risks show themselves during the implementation of the activity. The more thoroughly risk management is implemented, the lower the number of residual risks that require to be catered for within this category and the higher the likelihood that they will fall into a lower impact rating. Residual risks and concerns may be mitigated either by the inclusion of a sum of money within the risk contingency or, subject to senior management's confidence in their risk management methodology, by the acceptance of the risk exposure (Cater *et al.*, 1996).

2.4 THE ROLE OF IT IN RISK MANAGEMENT

Technological developments in the last few decades have elevated the role that information plays in the survival of companies, and is causing a revolution in the way organisations in general treat information, information systems and its associated technologies (Edum-Fotwe, et al, 2001). Information technology plays an important role in project risk management, from information processing, information analysis to decision support.

2.4.1 Risk Management Systems

Along with the development of both the hardware and software of computer systems, a number of software has been developed for a variety of applications in the construction industry, as well as for project risk management. The application of information systems in project risk management starts by focusing on quantitative analysis. At this stage, modelling is carried out by using certain modelling methods, spreadsheet add-in packages, network-based packages or hybrid packages. These packages are mostly based on probability analysis and operational research approaches. Since the early 1990's, researchers realized the importance of qualitative risk analysis, hence, some qualitative packages then came onto the market. These systems usually use questionnaires and checklists. At the same time, whether or not to provide a risk mitigation strategy in the software became important criteria by which to evaluate risk management systems. Some packages provide both qualitative and quantitative risk analysis.

Quantitative Models

Quantitative models are the most popular tools applied in risk management practice. These packages are largely based on Monte Carlo simulation or CMP/PERT network along with a rational database. Table 2.1 lists the well-known risk quantitative software on the market. These packages are classified into four categories: spreadsheet add-in, directly modelling risks, planning packages add-in and decision support systems.

Table 2.1 Quantitative risk analysis software tools

Type	Package Name	Vendor	Theory based
Spreadsheet	Crystal Ball	Decisioneering	Monte Carlo
	@Risk	Palisade Corporation	Monte Carlo, sensitivity and scenario analysis
	Risk Analyzer	MacroSystems	Monte Carlo
	CONTROL-IT	Jerry Fitzgerald & Associates	Delphi
	RI$K	Tecolote Research, Inc	Monte Carlo
Modelling	DynRisk	University of Oslo	Influence diagram
	DPL	Syncopation Software	Influence diagram, decision tree
	Definitive Scenario	Definitive Software	Influence Diagram, Monte Carlo
	Buddy	Countermeasures, Inc.	CMP Network, relational database
	PRA	Katmar software	Monte Carlo
Planning	Monte Carlo	Primavera	Monte Carlo, network
	Open Plan	Welcome	Network
	Risk+	C/S Solutions, Inc.	Microsoft Project
Decision support	Temper	Finland	Probability
	Futura	Proha Group	Danish Lichtenberg method
	Assessing Risk	Internal Audit	AHP
	BDSS	OPA, Inc.	Bayesian network

- Integrated with spreadsheet

Packages integrated with spreadsheet mostly use Monte Carlo simulation to calculate the cost of projects undertaken or in making decisions that are subject to uncertainty. These packages are principally of use where the decision alternatives are a number of well-defined, discrete courses of action.

Crystal ball is probably the most well known spreadsheet add-in package. It is a suite of Excel-based applications for risk analysis, Monte Carlo simulation, optimization, time-series forecasting, and real options analysis. Crystal Ball is already successfully applied in a broad area such as consulting, cost estimation, engineering, marketing, portfolio management, project management, real options, research/ public policy and six sigma (Decisioneering, 2004).

- Modelling

Risk modelling packages are mostly based on influence diagrams. Occasionally they also employ Monte Carlo, CMP/PERT network, and decision tree for system development.

Dynrisk is the one which has achieved the most success and its basic principle is employed to develop Definitive Scenario. Dynrisk is used to build models which are simulated and analyzed to determine the optimal outcome. The strengths of Dynrisk are its technical robustness, flexibility and ability to handle complex models.

- Add-in planning packages

Risk management systems add-in planning packages, such as Microsoft Project and Primavera Project Planner (P3), allow the use of a number of time and cost performance parametres by employing network and relational databases. Network-based packages permit the inclusion of interdependencies between activities, so that these packages are well suited to the modelling of projects rather than just decision processes (Smith, 1998). Relational databases usually accompany the application of these packages to cope with large volumes of project data.

Monte Carlo for Primavera is one of the success stories. Integrated with Primavera P3, Monte Carlo supports quantitative analysis of risks for the whole project as well as individual segments and quanfities risks built into the schedule, thus providing mid-project corrections to accumulate acutal performance data of the project (Primavera, 2004).

- Decision support systems

The decision support systems are based on various theories, such as probability analysis, Bayesian Theory, AHP and the Danish Lichtenberg method. The Danish Lichtenberg method is a structured approach to modelling the future which enables decisions and plans to be made with a greatly enhanced understanding of the key issues, and the assumptions and risks that are involved (Proha Group, 2004). This method was successfully used in the development of Futura in project risk decision making.

Among these methods, Analysis Hierarchy Process (AHP) is a well applied risk decision support system. Mustafa (1991) demonstrates a risk assessment method by using the Analysis Hierarchy Process. Later, it was introduced to assess the risks of constructing the Jamuna Multipurpose bridge project in Bangladesh. Recently, Nazari (1999), developed a decision support system for the assessment of the viability of Super Projects pre-appraisal phrases decision making, using the AHP method for risk analysis.

Qualitative Models

While quantitative models had achieved considerable success, researchers have also realized the importance of qualitative analysis in project risk management processes. In practice, qualitative analysis is carried out first. If qualitative analysis cannot give sufficient detail, quantitative analysis will be carried out to provide numerical evidence.

Table 2.2 Qualitative risk analysis software tools

Package Name	Vender	Theory based
RANK-IT	Jerry Fitzgerald & Associates	Delphi
RiskPAC	CSCI	Questionnaire, knowledge base
JANBER	Eagan	Questionnaire and checklist
Risk Alert	J. E. Boritz Consultants Limited	Sensitivity analysis, benchmarking, decision tree

The familiar qualitative packages are JANBER, Risk Alert, RANK-IT and RiskPAC. The quantitative models are mostly based on a questionaire and checklist. Among these, RiskPAC integrates both qualitative and quantitative capability. The benchmarking facility of Risk Alert also enables comparison of both the quantitative and qualitative features.

Intelligent Systems

Recently, the application of artificial intelligent techniques to risk management has attracted more attention, both in quantitative and qualitative terms. In quantitative analysis, fuzzy logic and artificial neural networks have already achieved considerable success. Meanwhile, rule-based reasoning, case-based reasoning, knowledge-based systems and machine learning were gradually applied to the risk qualitative analysis.

- Artificial Neural Networks

Artificial Neural Network (ANN) is a system loosely modelled on the human brain. The multiple layers of simple processing elements are called neurons. Each neuron is linked to certain of its neighbours with varying coefficients of connectivity that represent the strengths of these connections. Learning is accomplished by adjusting these strengths to cause the overall network to output appropriate results (Klerfors,1998). Like the human brain, ANNs learn from experience, generalise from previous examples to new ones and abstract essential characteristics from inputs containing irrelevant data (Chen and Hartman, 1998). They deal with problems where there are complex relationships between inputs and outputs and where the input data is distorted by high noise levels. ANN also allows self-learning, self-organisation, and parallel processing, and is well suited for problems involving matching input patterns to a set of output patterns where deep reasoning is not required (Boussabaine and Kaka, 1998).

Hashemi and Stafford (1993) developed a risk assessment system by ANN. A three layer Backpropagation (BP) network was trained to perform risk assessment on a set of toxicological data and gave decisions like those given by experts. Al-Tabtabai and Alex (2000) presented modelling the cost of political risk in international construction projects with the BP network. Chen and Hartman (1998) combined different neural networks in their risk assessment and contingency allocation system. The model was based on BP; General Regression Neural Networks was used for the proposed cost model; and BP and Probabilistic Neural Networks was used for the time model.

- Machine Learning

Machine learning is a computer program system to learn from experience, E, with respect to some class of tasks, T and performance measure, P; only if its performance at tasks T, as measured by P, improves with experience E (Mitchell, 1997). The field of machine learning studies the design of computer programs which are able to induce patterns, regularities, or rules from past experiences. Learner, a computer program, processes data D, representing past experiences, and tries to either develop an appropriate response to future data, or describe in some meaningful way the data seen.

Stone *et al.* (1990) addressed a system for managing risk in civil engineering by Machine Learning from previous failures for future projects, and to carry this out in a manner that facilitates the expression of the inherent uncertainties. The system was built through the representation of case histories of previous projects in the form of event sequence diagram, the development of a hierarchically structured knowledge base of concepts at differing degrees of generality, and the use of the discrimination and connectivity machine learning algorithms to extract patterns of commonly occurring features.

- Fuzzy Logic

Fuzzy Logic is an extension of the classical Boolean or binary logic. It was formulated by Zadeh to model vagueness intrinsic to human cognitive processes (Baloi and Price, 2003). As a branch of logic, it uses degrees of membership in sets rather than strict true/false membership. Using fuzzy logic, sets may be defined on vague, linguistic terms. These terms, which cannot be defined meaningfully with a precise single term, may be formally defined in mathematical logic by using fuzzy logic (Carr and Tah, 2001).

Kangari (1988) presents an integrated knowledge-based system for construction risk management using fuzzy sets. Tah *et al.* (1993) try a linguistic approach to risk management during the tender stage for contingency allocation, using fuzzy logic. Carr and Tah (2001) identify the relationships between risk sources and the consequences on a project by using fuzzy association and composition. Recently, Baloi and Price (2003) developed a fuzzy decision framework for contractors to handle global risk factors affecting construction cost performance at project level and evaluated different decision-making technologies.

- Rule-based System

The rule-based system is the most well-known reasoning system in Expert Systems. A rule-based system consists of a bunch of IF-THEN rules and facts. Facts are things, which are true about the domain, and rules describe relations or phenomena in the same problem domain. Rules are represented as if-then statements, often defining logical relations between concepts of the problem domain (Bratko, 1986). It is presented as if a particular set of conditions exists, then responds with a particular set of behaviours.

CORA is one of the successful risk analysis packages built with rule-based reasoning. It combines risk expert rules files generated by risk expert and local information files populated by field personnel to estimate the cost of risk and to optimize the selection of risk mitigation, transfer and recovery (International Security Technology, 2004).

Leung *et al.* (1998) developed a knowledge-based system for identifying potential project risks which was tested in a transmission line construction project. The domain risk knowledge was collected by interviewing project managers, project engineers, and senior personnel who were involved in similar projects. Risk causing knowledge was represented in the form of IF/ THEN type rules. Ninty seven rules, which cover 14 external and 24 internal risk factors, had been developed by the interviews.

- Analytic Hierarchy Process

AHP is a method of breaking down a complex unstructured situation into its component parts; arranging these parts, or variables, into a hierarchic order; assigning numerical values to subjective judgements on the relative importance of each variable; and synthesising the judgements to determine which variables have the highest priority and should be acted upon to influence the outcome of the situation. It is a method for handling cognitive complexity and for providing a basis for eliciting, discussing, recording, and evaluating the elements of a decision (Al-Harbi, 2001). The well known package to support the creation of AHP modelling is Expert Choice.

2.4.2 Pros and Cons of IT

Advantages

There are many advantages to using computer systems to carry out project risk management. The following list highlights some of the more significant advantages (Smith, 1998).

- **Speed and accuracy.** As construction project risk analysis usually involves a number of variables compared with manual calculations, computer systems have terrific advantage of carrying out quantification analysis both in speed and accuracy.

- **Knowledge sharing.** A computer network is the most efficient way for knowledge sharing. By using intranet the project information and experience of individual experts can be shared within a certain institution/ project group. And by using internet, broad knowledge is shared all around the world. It also makes information exchange more efficient by using electronic project sources. The additional techniques, such as net-meeting and e-commerce, greatly improve communication and enhance the procurement process.

- **Flexibility.** Computers are very flexible in the way in which they can accept information, enabling most projects to be modelled using a computer. The programs used to model projects can be either off-the-shelf packages or tailored to the needs of the user (Smith, 1998).

- **Reduces subjective judgment.** Few people have a reliable intuitive understanding of risks because of their different backgrounds and experiences. However, as construction projects usually last a long time, the experience held by individual experts is limited. Moreover, when using a method that is based on group work, such as brainstorming, peer pressure may influence decision making. Intelligent computer systems can reason, based on a common principle, and provide unbiased solutions to support decision making.

- **Storage and memory.** Computer systems can store a great range of information such as: text documents; graphics; pictures; sounds; movies; and dynamic scenario simulations. This can save space and cost for keeping the project documents and other relevant information.

Limitations

Despite remarkable advantages, using information systems for project risk management is also subject to the limitations listed below.

- **Poor data leading to inaccurate models.** A model of a project is only as good as the data that is input, so if the input data are inaccurate then the model will not accurately reflect the project (Smith, 1998). In traditional quantification systems especially the information required might not be available in full, so that some information input under certain assumptions may lead to inaccurate results.

- **Security.** Because information systems use electronic information, they are occasionally subject to disaster by virus and the loss of documents. A virus may create damage at different levels, some can destroy the hard disk with all the files stored therein, and some just destroy certain files. The counter-claim group of the 'China Yellow River multipurpose project' lot three was subject to the CIH virus in April 1999. As a result, even with the backup, there was still a lot of work to be redone, which dramatically influenced progress of the project and caused additional costs.

- **Not able to fully reflect real life complications.** The model produced is usually only a part representation of real life and, therefore, does not necessarily accurately reflect the reaction of the actual project to real life complications. It is impossible to be sure that the model will react in exactly the same way as the real project because the project is yet to occur and everything is based on what is expected to happen; the only exception is when the project being modeled is identical to a previous project (Smith, 1998).

- **Reliance on computer output.** Too much reliance is placed on the output from computers and often there is insufficient checking of the model or the program used to create the model. It is difficult to tell whether or not a project model is an accurate representation of reality. If the model is very inaccurate this can be easily detected, but if it is nearly accurate then this is much more difficult to detect (Smith, 1998).

2.4.3 Selection of Information Systems

The main features to be considered in evaluating individual risk management packages are functionality, flexibility, usability and cost.

Functionality

Different systems use different methods to simulate project risk analysis to a certain level. The chosen package should satisfy the requirement of risk analysis that it is supposed to achieve. It is important to explore the available functions that the packages support and the outputs the packages can provide. For instance, clarify which kind of risk analysis is required: quantitative, qualitative or both; and what kind of output is expected, the overall risk level, the precise figure of the probability, the potential loss which may occur, the mitigation/response strategy and any certain cost/schedule simulation to be carried out.

Flexibility

Apart from the functionality, it is also necessary to consider the flexibility of the system. Whether it is well integrated with the additional software with which it is expected to integrate such as Excel, Project and P3, and whether the computer hardware components match the package required running environment.

Usability

How long does it take to learn the package and to build the risk management model? Are the technique support and online help sufficient? Is the user interface friendly enough? As packages have various capacities to deal with project information and some of them are particularly suitable for large projects, it also important to choose an appropriate package to cope with the relevant project.

Cost

Taking into account the functionality of the package, the costs of using a risk management package in practice should also be considered, such as the cost of the package, cost of training personnel and relevant maintenance.

2.5 PROBLEMS IN CURRENT RISK MANAGEMENT

2.5.1 Risk Management in Practice

Even though there are a range of benefits by integrating risk management with project management, the limited use of risk management in construction is surprising considering the presence of risk and uncertainty in every phase of the project development cycle (Uher and Toakley, 1999). The main problem of present studies is the difficulty in promoting research achievements effectively into real application. A survey, carried out by Leung *et al.* in 1998, on the practice of risk management in building services, electricity supply, and public transportation industries in Hong Kong, shows risk management approaches are not widely accepted in project management.

Ward *et al.* (1991) gave reasons why risk management, particularly risk analysis, has not been used more effectively in construction. They identified 'culture issues' such as lack of knowledge, negative attitudes and mistrust of risk analysis as being the main reasons preventing its greater use. Akintoye and MacLeod (1997) carried out a survey of why contractors and project managers did not use risk analysis and management. Lack of familiarity featured prominently amongst the reasons provided by the respondents for non-use of formal risk management techniques. This is followed by the claim that the amount of calculation involved using the techniques are unwarranted in order to meet the project's objectives of time, cost and quality. Further reasons covered lack of information and knowledge, requirement of sound data to ensure confidence, and subjectivity. Simon and Richard (1992) investigated the difficulties of applying risk analysis. They found them to be: size of project, complexity of the problem, new market, and new product.

To summarize the above, the major limitations identified in applying risk management in construction project management are drawn from four fields: accuracy, complexity, time and cost, and communication.

Accuracy

Risk management begins with risk identification, which is typically carried out by human experts. The methods undertaken are usually based on assumptions, which might be influenced by intuition, biases and subjectivity. As a result, the data input to risk analysis tools might inevitably inaccurate, incomplete, or over-simplified (Carter, *et al.*, 1996). This can directly influence the output results of risk analysis. Moreover, different risk management approaches focus on different analysis objectives. No method is perfect. Consequently, errors may always exist in the analysis, and to choose the appropriate method is the key process. Furthermore, because of difficulties in coping with complicated uncertainty throughout the analysis process, the analysis accuracy can be influenced by the changing situation.

Complexity

Because of the complex nature of many construction projects, risk management relies on a combination of several methods together rather than a single method. Popular risk analysis methods, such as Monte Carlo, influence diagram, sensitive analysis, and decision tree, all need profound understanding. These techniques relate to the knowledge of mathematics, statistics and operational research, which may not be easy for all project participants to understand and use correctly. Furthermore, when using risk analysis software, the problem of accumulating many interacting valuables in a statistical way is complex (Carter, *et al.*, 1996).

Time and cost

The time and cost involved in using risk management tools is another issue. The time required by individual analyst to fully learn the various risk management theories can be considerable, particularly for the more complex analysis methods. Moreover, training the relevant staff, learning and choosing the available risk management approaches, obtaining input estimates and assessing the probabilities, and understanding and interpreting outcomes, all need time. Meanwhile, to carry out these activities relevant investment in term of advanced technology, material, and labour input is unavoidable. Clearly this will result in the employer incurring costs during this learning process.

Communication

The lack of knowledge and information which is decision making based can also influence the risk management quality. A typical project involves a range of disparate professionals. In many cases, the participants are geographically distributed, making the need for effective information and communication technologies acute (Anumba, *et al.,* 2002). Smith (1999) points out that one of the main obstacles when introducing risk management to an organization is the lack of openness and communication within the organisation. For effective collaborative working between the parties in a construction project team, it is essential that enabling information and communications technologies are available (Anumba, *et al.,* 2001). However, as discussed in Chapter Two, these information and communication technologies are not always sufficiently available.

In addition, the communication gap developed between disciplines, particularly between contractor and client, may result in experienced contractors hiding contingency allowances in their estimate to reduce the cost of a bid in order to submit the most competitive price and win the contract.

2.5.2 Importance of Risk Identification

Exploring the development of project risk management, the research of quantitative risk analysis achieved significant success in the 1980s. In the 1990s, a variety of computer software was developed to improve the application of risk quantitative analysis. As quantification methods have been well developed, it is now more important to identify the potential risks, evaluate the possible influences and make responsive decisions. Therefore attention should be point to qualitative analysis because qualitative analysis determines whether quantitative analysis is necessary, which quantitative analysis method should be used, the scope of quantitative analysis, and the element used in quantitative analysis.

As described in the previous section, accuracy is the main issue of the limited application of risk management in practice. Risks can only be measured, controlled, transferred or otherwise managed once they have been identified and so the process by which risks are identified assumes an importance that is quite distinct from any particular risk (Toakley and Ling, 1991 and Dickson, 1991). Risk identification provides fundamental information for subsequent forecasting stages: assessment, ranking, classification, and judging the probability and impact of potential risks. The accuracy of quantitative analysis is based on the input elements extracted from the risk identification stage. As a result the identification process plays an important role in risk management. In fact, it is believed that the main benefits of risk management originate from the identification rather than the analysis stage (Uher, 1993).

However, the risk identification stage has not been adequately addressed in risk management literature (Raftery, 1994). To ultimately improve the accuracy of risk analysis and increase the application of risk management, it is essential to attach importance to the risk identification stage.

2.5.3 Limitation of Risk Identification

The traditional methods used for risk identification include checklists, brainstorming, examining historic data, the risk register and the Delphi method.

Checklists are a simple and inexpensive way of generating information. However, the possibility of ambiguities and subjective alter the accuracy of results (Edwards, 1995). Furthermore, the accuracy of a checklist depends on a thorough understanding of what could go wrong. Without extensive past experience and documented fault and hazard histories, a checklist is not soundly based. Moreover, its adequacy also depends on the circumstances of its use being the same as those in which it was created; if they differ, the checklist could be out of date, or inapplicable, and dangerously misleading. Checklists, even when appropriate, need to be reviewed periodically (Redmill, 2002).

In systems which are not so well understood, perhaps because they are only now being planned or designed , techniques which employ the creativity of human investigation are required. Brainstorming is sometimes used. Brainstorming enables the project's personnel to hear what the other members of the project team see as risks and then to use these ideas to give themselves inspiration in identifying additional project risks. It is one of the most popular methods in carrying out risk identification because of its simplicity and speed. But although creative, there is usually little formality in this process. Information for risk identification may also be obtained from audits and formal or informal interviews with staff, all of which depend on human abilities, attitudes and thoroughness (Redmill, 2002). Merna and Njiru (2002) addressed limitations of this technique as its dependence on the group composition, conformance, personality characteristics, compatibility, and peer pressure.

Examination of historic data from previous projects ensures that corporate knowledge is utilized. However, an organization may not deal with the same project twice and the data from a previous similar project may not have been recorded well. Therefore, this technique can only be used successfully in limited cases.

The Delphi method provides a communication process allowing a group of individuals as a whole to deal with complex problems. Nevertheless, the drawbacks of the Delphi method are that it is its unreliability, over sensitivity of results, ambiguous questions, and time consumed (Chan et al., 2001) and it is expensive in terms of the resources used, the cost of resources and the time undertaken. The success of this method largely relies upon the selection of the panel of experts.

2.6 SUMMARY

Construction projects have a poor reputation caused by overdue schedules, overrun budgets and unsatisfactory quality. Risk management, especially in project appraisal, can help project participants to understand the potential problems of a project in order to achieve the project's goals. Although a variety of process models have been developed to guide risk management, they all tend to share a series of key activities consisting of risk identification, risk analysis and risk response.

Information technology can be a way of liberating the human brain from complicated calculations, helping with storage of information and dealing with basic logical reasoning, and providing contributable benefits for processing risk management in construction projects. A number of packages have been developed to facilitate the risk management process, including qualitative models, quantitative models and artificial intelligent systems. Existing successful risk analysis packages include decision support, modeling, spreadsheet add-in, and planning add-in systems. To carry out advanced numeric analysis and cope with knowledge in relation to project risks, serial Artificial Intelligent techniques were introduced to project risk management, for instance, ANN, rule based systems, machine learning and fuzzy logic. To employ appropriate IT in project risk management practice, it is also necessary to consider the time and cost involved as well as to compare the functionality, stability and flexibility.

However, despite a variety of risk management and identification methods being developed, these approaches have not resulted in expected benefits to the construction industry. Risk management was not well used in the real project management. The major limitations identified in applying risk management in construction project management are dawn from four aspects: accuracy, complexity, time and cost, and communication. Current project risk management is more concerned with quantitative analysis but qualitative analysis, such as risk identification should be paid more attention.

Risk identification is the fundamental stage for successful risk management. Risk identification is fundamental in project risk management as it underlies the analysis approach to be used and the elements to be considered later on. However, the existing risk identification methods have various limitations. In a checklist the possibility of

ambiguities and subjective alter the accuracy of results. Brainstorming depends on the composition of the group, conformance, personality characteristics, compatibility and peer pressure. Examination of historic data is restricted by the project information available. The Delphi method is insufficiently reliable, over sensitive to the results of ambiguous questions and is time consuming. As a result, a new approach is required to overcome the limitations of existing risk identification methods.

CHAPTER 3

KNOWLEDGE MANAGEMENT

3.0 INTRODUCTION

Having found out that inadequate risk identification is the main impediment to improving the application of risk management, it is the way to choose an appropriate approach to improve risk identification. The risk identification process requires a great deal of information and knowledge. Whether or not the information and knowledge gained from previous projects can be shared and transferred will significantly influence the result of risk identification. Increasingly, knowledge management has attracted significant attention in the field of project management, especially in the view of the organization's strategy and human resources. What is knowledge and knowledge management? How is project decision making undertaken? Is knowledge management is useful for project risk management? And how can it be used to identify project risks? All these questions are answered in this chapter. It then describes the limitations of current risk identification methods and explores the influence of human behaviour in decision making. Finally this chapter demonstrates an appropriate approach for risk identification.

3.1 KNOWLEDGE AND KNOWLEDGE MANAGEMENT

Before attempting to address issues of knowledge management in relation to project risks, it is appropriate to develop some perspective regarding what knowledge and knowledge management are.

3.1.1 Definitions

Fleming (1996) addressed the issue that information, knowledge and wisdom are more than simply collections; rather, the whole represents more than the sum of its parts and has a synergy of its own. Stenmark (2002) collected a variety of definitions regarding data, information and knowledge, listed in Table 3.1.

Table 3.1 Definitions of data, information and knowledge (Stenmark, 2002)

Author(s)	Data	Information	Knowledge
Wiig (1993)	-	Facts organized to describe a situation or condition	Truths and beliefs, perspectives and concepts, judgements and expectations, methodologies and know how
Nonaka and Takeuchi (1995)	-	A flow of meaningful messages	Commitments and beliefs created from these messages
Spek and Spijkervet (1997)	Symbols not yet interpreted	Data with meaning	The ability to ask the meaning
Devenport (1997)	Simple observations	Data with relevance and purpose	Valuable information from the human mind
Davenport and Prusak (1998)	A set of discrete facts	A message meant to change the receiver's perception	Experiences, values, insights, and contextual information
Quigley and Debons (1999)	Text that does not answer questions to a particular problem	Text that answers the questions 'who', 'when', 'what', or 'where'	Text that answers the questions 'why' and 'how'
Choo et al. (2000)	Facts and messages	Data vested with meaning	Justified, true beliefs

Data

Data is just a meaningless point in space and time. It is like an event out of context, a letter out of context, a word out of context. Since it is out of context, it is without a meaningful relation to anything else. As Fleming (1996) indicated, a collection of data for which there is no relation between the pieces of data is not information.

Information

Information is quite simply an understanding of the relationships between pieces of data, or between pieces of data and other information. It generally does not provide a foundation for why the data is, what it is, nor an indication as to how the data is likely to change over time. Information has a tendency to be relatively static in time and linear in nature. Information is a relationship between data and, quite simply, is what it means, with great dependence on context for its meaning and with little implication for the future.

Knowledge

Beyond relation there is pattern (Bateson, 1988) where pattern is more than simply a relation of relations. Pattern embodies both a consistency and completeness of relations which, to an extent, creates its own context. Pattern also serves as an archetype with both an implied repeatability and predictability.

When a pattern relation exists amidst data and information, the pattern has the *potential* to represent knowledge. It only becomes knowledge when one is able to realize and understand the pattern and its implications. The patterns representing knowledge have a tendency to be more self-contextualizing. That is, the pattern tends, to a great extent, to create its own context rather than being context dependent to the same extent that information is. A pattern which represents knowledge also provides, when the pattern is understood, a high level of reliability or predictability as to how the pattern will evolve over time, for patterns are seldom static. Patterns which represent knowledge have completeness to them that information simply does not contain.

Wisdom

Wisdom arises when one understands the foundamental principles responsible for the patterns representing knowledge being what they are. And wisdom, even more so than knowledge, tends to create its own context. These foundational principles are universal and completely context independent.

So, in summary, the following associations can reasonably be made:

- **Information** relates to description, definition, or perspective (what, who, when, where).

- **Knowledge** comprises strategy, practice, method, or approach (how).
- **Wisdom** embodies principle, insight, moral, or archetype (why).

Figure 3.1 shows the relationships among data, information, knowledge and wisdom in terms of different understanding levels and context independence. However, this figure does not exactly explain the complete relationship among them as data, information, knowledge, and wisdom are interwoven and interrelated in more complicated ways than this simple model suggests. To look at information purely in terms of the degree to which it has been processed, i.e., the data, information, knowledge hierarchy (Davenport, 1997, Checkland and Howell, 1998), oversimplifies the complex relationship between the three intangibles.

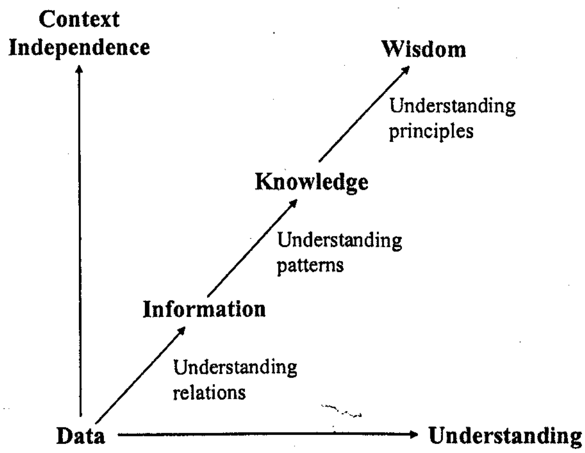

Figure 3.1 The relationship of data, information, knowledge and wisdom

Firstly, the relationship among them is not linear and the distances between two entities next to each other are not equal, as the figure shows. Furthermore, the direction of format transferring is not fixed, since there are many examples that use knowledge to derive information and to create data out of information. Finally, it is not illustrated that the entities in a higher position are more valuable than the lower ones, i.e. knowledge is more valuable than information, which in turn is superior to data is not clear.

As in practice, the terms 'information' and 'knowledge' are often used interchangeably. Davenport and Prusak (1998) defined knowledge as a "fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information". Knowledge is the whole body of cognitions and skills which individuals use to solve problems (Probst *et al.*, 1999). It is a complex concept which consists of information and skills acquired through experience; truth and belief, perspective and judgments, expectations and methodologies (Egbu *et al.*, 2003).

A common approach to considering knowledge often highlights its relationship to information in terms of difference. Nonaka and Takeuchi (1995) stated: "Information provides a new point of view for interpreting events or objects, which makes visible previously invisible meanings or sheds light on unexpected connections. Thus information is a necessary medium or material for eliciting and constructing knowledge". Although the relationship between information and knowledge may be seen as "closely associated," it should be more appropriately seen in terms of a "Dynamic and interactive relationship" (Watson, 2001). Information facilitates the development of knowledge, which creates more information that deepens knowledge. Choo (1998) has viewed this dynamic interactive relationship as part of the process of knowing which facilitates the capacity to act in context. Stewart (1997) notes: "The idea that knowledge can be slotted into a data wisdom hierarchy is bogus, for the simple reason that one man's knowledge is another man's data". Furthermore, Boisot (1998) interprets this dynamic relationship among data, information and knowledge, as shown in Figure 3.2: data can be characterized as a property of things, knowledge is a property of agents predisposing them to act in particular circumstances; information is that subset of the data residing in things that activates an agent it is filtered from the data by the agent's perceptual or conceptual apparatus.

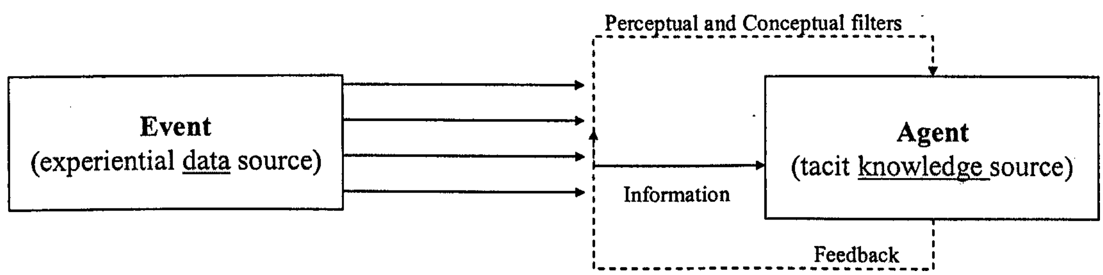

Figure 3.2 Data, information and knowledge (Boisot, 1998)

3.1.2 Knowledge Management

The concept of knowledge management first emerged in the 1960s (Drucker, 1969) with the recognition that the post-industrial society was characterized more by intellectual, as opposed to manual work. Since then the concept has become increasingly popular in academic and practitioner circles where a variety of perspectives have emerged (Inkpen, 1996). Knowledge management is about "creating, acquiring, capturing, sharing and using knowledge" (Currie, 2003). It involves the acquisition, storage, retrieval, application, generation, and review of the knowledge assets of an organization in a controlled way (Watson, 2003).

Quintas *et al.* (1997) defined knowledge management as "the process of continually managing knowledge of all kinds to meet existing and emerging needs, to identify and exploit existing and acquired knowledge assets and to develop new opportunity." Webb (1998) defined knowledge management as "the identification, optimization and active management of intellectual assets to create value, increase productivity and gain and sustain competitive advantage." It is important to recognize that knowledge management involves the sharing of knowledge, as well as other processes. Several authors have identified these different processes. For example, Rugles (1997) categorized theses processes as generate, codify, and transfer. Siemieniuch and Sinclair (1999) identified five processes: generate prorogate, transfer, locate and access, and maintain and modify. Tiwana (2002) identified five categories as find, create new, package and assemble, apply, and reuse and revalidate knowledge. Laudon and Laudon (2000) recognized that these processes can be cyclical and iterative and that they all have different process requirements.

Knowledge may be divided into two forms: explicit and tacit (Nonaka, 1991). Tacit knowledge is personal knowledge which in practice is difficult to communicate fully to others, it has a technical as well as cognitive dimension. Explicit knowledge is the knowledge that has been articulated, coded and recorded (Kermally, 2002).

Explicit Knowledge

Nanaka and Takeuchi (1995) define explicit knowledge or codified knowledge as knowledge that can be articulated in formal language including grammatical statements, mathematical expressions, specifications, and manuals. Such explicit knowledge, they conclude, can be transmitted easily and formally between individuals. Explicit knowledge is formal and systematic. Choo (1998) suggests that explicit knowledge is knowledge that is made manifest through language, symbols, objects, and artifacts. Explicit knowledge in construction projects can further be object based, i.e. software codes, databases, technical drawings and blueprints, mathematical formulas, project plans, and statistical reports, or rule based routines and procedures. It is therefore easily communicated and shared throughout the firm. For example, explicit knowledge is embodied in a computer program or set of procedures for selecting procurement.

Tacit Knowledge

Tacit knowledge is defined as "highly personal" and not amenable to formalization and standardization (Currie, 2003). In addition, tacit knowledge is not easily communicated to others. Nonaka (Currie, 1991) claims that, "tacit knowledge is also deeply rooted in action and in an individual's commitment to a specific context: a craft or profession, a particular technology or product market, or the activities of a work group or team. Tacit knowledge consists partly of technical skills: the kind of informal, hard to pin down skills captured in the term "know-how". A master craftsman after years of experience develops a wealth of expertise "at his finger-tips". But he is often unable to articulate the scientific or technical principles behind what he knows". Nonaka argues that the knowledge creating company should attempt to make tacit knowledge explicit. This is likely to be achieved in the ubiquitous team approach adopted by Japanese companies where individuals work together in an attempt to create and share knowledge for product development and technological innovation (Davenport and Prusak, 1998).

3.1.3 Knowledge Management in Construction

The discourse of knowledge management is increasingly evident within the project management literature. Knowledge management enables the creation, communication, and application of knowledge of all kinds to achieve business goals (Tiwana, 2000). Turner (1999) observes that project teams consist of "knowledge workers". This clearly resonates with widespread notions of a knowledge economy. The issue of how better to share knowledge across teams and between knowledge workers therefore becomes of central concern to project managers. Drew (1999) described four types of knowledge: 1) what we know, we know; 2) what we know, we don't know; 3) what we don't know, we know; and 4) what we don't know, we don't know. He emphasized that most knowledge management programs were concerned with processes for sharing and distributing existing knowledge.

The issue of knowledge sharing also becomes increasingly important to many project based organizations as they turn themselves into service companies that are increasingly divorced from the physical work of construction. The epistemological orientation of the project management discipline tends towards a functionalist, managerial framework of knowledge that readily accepts the link between knowledge and competitive advantage perceived elsewhere (Lanzarra and Patriotta, 2000). From this perspective, the challenge is how knowledge can be captured, acquired or appropriated (Kamara et al., 2002, Kululanga and McCaffer, 2001, and Egbu, 1999). Lanzarra and Patriotta (2000) criticize this orientation for its lack of scrutiny on knowledge by itself and its tendency to conceptualize knowledge as an objective, transferable commodity.

Much effort has also been expended on the codification of project management into "bodies of knowledge" (Dixon, 2000 and PMI, 2000). The underlying assumption is that such bodies of knowledge retain any meaning once divorced from context. Most project managers would readily concede that there is little substitute for experience, thereby implying that knowledge derived from experience cannot easily be codified. Here lies the issue for project managers: how can a project's team members collaborate better through knowledge sharing? If this challenge is to be addressed, the nature of knowledge must be addressed.

"The construction industry delivers large, expensive, built facilities at the end of a construction process. It is a strong, knowledge-based industry that relies heavily on the knowledge input by different participants in a project team. The management of construction project knowledge should clearly address the spectrum from tacit to explicit knowledge. The construction industry relies heavily on its experts in specialized areas throughout the project life, from design to use (Carrillo *et al.,* 2004).

Knowledge management is particularly relevant to the UK construction industry with its current focus on collaborative working, knowledge exchange and the creation of new networks to increase competitiveness and profitability (Moodley *et al.*, 2001). The publications of the UK report Rethinking Construction (DETR, 1998), highlighting the low levels of company profitability, forced a number of construction organizations to reconsider the way in which they manage their business and the role of learning and knowledge in achieving performance targets.

Carrillo *et al.* (2004) carried out a survey to study the impact of knowledge management practices on business performance in major UK construction organizations. The results show that 77.4% of the responses were aware of the benefit of knowledge management. They concluded that the main benefits of implementing knowledge management within an organization are seen as the need to share tacit knowledge of individual projects and to disseminate best practice.

3.2 DECISION MAKING

The term "decision making" in investment refers to a decision made about the acceptability of the investment that should be very carefully examined (Hammond and Haylen, 1994). The goal of all decision-making techniques is to map out the probabilities, consequences, and financial options, with the intention of constructing some kind of balance sheet that can provide guidance to decision makers. Decision-making is a game of imperfect information involving the future, change and human action and reaction (Flanagan and Normen, 1993).

3.2.1 Decision Making in Risk Management

To take a risk is to take a gamble. Therefore, before actually taking it, it is necessary to understand what the risk contains and decide what to do. The more the project managers discover the exact nature and level of risks they face, the better they are able to prepare for it. To understand risk, there are a number of tasks to be undertaken, such as calculate the odds, weigh up all the facts and apply personal experience, knowledge and intuition. As explained in section 2.1, risk embodies hazard, opportunity, complexity and uncertainty. The risk analyzer needs to estimate probability, choice and decision. Information and knowledge in relation to risk includes both empirical data and insight obtained by interpretation. It is necessary to distinguish between what is known and what is unknown. Many decisions are opaque in the sense that they are hard to understand, solve or explain. They often involve numerous performance objectives which are generally conflicting, each making demands on scare resources such as time, money or technological capability (Flanagan and Normen, 1993). Project managers rarely consider how a solution was arrived at, or how to justify the use of the chosen approach for decision-making.

Risk management is about communication to make better decisions on a real project under conditions of uncertainty (Smith, 2003). As already explained in Chapter Two, the decision made in the appraisal stage can significantly affect the whole project. Consequently to carry out risk management in the appraisal stage can more likely lead to a successful project. At appraisal stage decisions are made mainly based on opinions. Intuition, expert skill and judgment will always influence decision-making, but a set of tools is now needed which will enable risk management techniques to be put into practice in the construction industry (Flanagan and Normen, 1993).

3.2.2 Intuition

Decision-makers rely upon both intuition and formal models to assess the worthiness of an alternative. Many decision-makers place great emphasis on intuition reasoning, following their feelings rather than their thoughts. As Isenberg (1985) pointed out, intuitive thought is not the opposite of rational thought. It is based upon both the accumulation of experience, which allows the decision maker to perform well learned operations rapidly, and on mental leaps which enable him to synthesize seemingly isolated information to produce results which represent more than the simple sum of its parts. Intuition is the acknowledgement of some 'gut feeling' of a situation and the best course of action to take. Whilst this is probably rooted in experience, it is much more tenuous. Decisions based on experience can be justified, whilst those based on intuition cannot (Flanagan and Normen, 1993).

Experience, intuition and judgment have their rightful place in decision making; often they present the only legitimate and available recourse. However, relying solely on experience, professional feel and intuitional hunches does not guarantee that the best course of action is chosen. The project managers may solve the problem and implement the solution skillfully, but the solution itself may be inferior or wrong (Flanagan and Normen, 1993).

3.2.3 Bias

Good decisions are based upon proper analysis and intuition. Facts help to formulate the basis for the decision, and intuition guides us to the decision. On occasions there has to be a balance between analysis and intuition. The judgmental ability of humans is flawed by numerous biases which distort the perception of reality. These biases affect the way of interpreting the past, predicting the future, and making choices in the present (Flanagan and Normen, 1993). Table 3.2 lists the main biases which have been identified by psychologists.

Table 3.2 Bias and its effects (Flanagan and Normen, 1993)

Bias	Effects
Availability	Judgment of probability of easily recalled events is distorted
Selective perception	Expectations may bias observations of variables relevant to a strategy
Illusory correlation	Encourages the belief that unrelated variables are correlated
Conservatism	Failure to sufficiently revise forecasts based on new information
Law of small numbers	An over estimation of the degree to which small samples are representative of population
Wishful thinking	The probability of desired outcomes judged to be inappropriately high
Illusion of control	Over estimation of the personal control over outcomes
Logical construction	Logical construction of events which cannot be accurately recalled
Hindsight bias	Over estimation of the predictability of past events

An individual lacks the multiple viewpoints required in risk identification, is subject to the inside view and an overconfidence bias, and is unlikely to carry out a thorough investigation (Redmill, 2002). Two common biases worthy of note are those of 'availability' and the 'illusion of control'. The availability bias is the tendency of a decision-maker to judge a future event as being likely if he can easily recall past occurrences of the event. This may indeed be a good measure of probability, since frequently occurring events are more readily recalled; but recall tends to be biased in favour of recent events, and those which appear as dramatic occurrence (Flanagan and Normen, 1993).

The illusion of control describes the tendency of decision-makers to overestimate their skill and the impact it will have on the outcome. This results in a tendency to express an expectation of success which exceeds the objective probability. This form of 'wishful' thinking can have dramatic repercussions in construction, when a commitment to a cause such as technological innovation or energy saving can blind decision maker to the inherent risks (Flanagan and Normen, 1993).

3.2.4 Experts and Experience

Experts are an individuals who have some degree of training, experience and /or knowledge greater than that of the average man. In general, they are substantive experts who, in a given domain, assess events in their field of expertise. Experience is built up over time through working in, and developing an understanding of, a particular aspect of their work. The experience can reside in an individual or in the corporate experience of the company which has been shared by the individuals within it. Nevertheless, reliance upon gut feelings frequently results in poor decision making (Flanagan and Normen, 1993).

Experience plays a valuable role in the way an expert works. Experience serves as a data base that can be used to fill gaps in the details learned about unfamiliar circumstances. The mind unconsciously searches its data base of experience automatically. Over the years our minds develop categories, methods and filing systems for all their experiences. It is this fund of previous experience that helps the brain separates the relevant from the irrelevant, all without interrupting conscious thought processes (Flanagan and Normen, 1993).

Experience can lead to bias in decision-making. When something happens to a person, he/she takes that event as being representative, when often it is not. It is tempting to solve present or future problems on the basis of extrapolating the past into the future. It takes a lot of wisdom, skill and nerve to use information that disagrees with past experience. Experience is the strongest resource available to the decision maker, it is also the most likely cause of a blinkered approach. People feel comfortable with information that validates their previous experience; they are reluctant to use data that is hostile or discomforting to their view (Flanagan and Normen, 1993).

Intuition, bias and experience can all influence decision making in relation to project risks. Obviously, the main purpose of risk identification is to gain knowledge about uncertain situations and potential problems that might arise in a given project. Well known risk identification methods, such as Delphi, brainstorming, a checklist/questionnaire, are mostly based on human decisions. Subjective judgment and psychological influence usually accompany with the experts' decision. This could be expressed as the impact of human behaviour.

3.3 INFORMATION AND KNOWLEDGE IN RISK MANAGEMENT

Nonaka (1991) states that knowledge and information are similar in some aspects, but different in some: while information is more factual, knowledge is about beliefs and commitment. The benefits of organizing and managing information and knowledge in relation to project risks have been recognized by many researchers. Al-Bahar and Crandall (1990), Perry and Hayes (1985), and Tah and Carr (2001) have proposed a variety of risk categorization schemes, with the goal being to assist users in risk identification by focusing attention on similar risks. Prompt lists either in a generic form that refer to all types of projects, and lists that refers to specific types of projects, have also been suggested as a way of organizing and managing risk knowledge. Examples of the latter include prompt lists for coastal projects (Simm and Cruickshank, 1998) and for river and estuarine construction (Morris and Simm, 2000).

3.3.1 Typical Knowledge Sharing in Risk Management

Risk management is an important tool to aid decision making in projects regarding more accurate cost estimates and plans. There is a gap between the theory and the techniques proposed to manage risk and what people do in practice (Flanagan and Normen, 1993). As explained in Chapter Two, to improve risk management it is now more important to improve risk qualitative analysis, especially risk identification.

In order to identify risks in an effective way, previous experience is important. Williams (1994) advocated a 'risk register' component in the management system of a project to generate an accessible database of risk experiences. Edwards and Bowen (1998) concluded: "risk management techniques were only useful as the willingness of the project participants to become knowledgeable and skilled in them". Chapman (1997) noted "for both parties to manage their risks effectively, it may be important to move towards a cooperative shared information approach to management." Thus to manage information and knowledge relating to project risks is the key to improving risk identification.

A successful knowledge management initiative will install a learning and knowledge sharing culture and environment and provide vision and effective leadership to overcome learning barriers (Maqsood *et al.*, 2003). Sharing knowledge in project risk management can reduce time and effort, speed up the decision-making process, provide an effective way of inducting new staff, encourage the use of knowledge and promote collaboration, capture knowledge for organizational use and encourage the transfer of best practice (Kermally, 2002).

Nonaka and Takeuchi (1995) presented the four modes of knowledge conversion:

- Socialisation: acquiring knowledge from the design engineer, site manager, etc.;
- Externalisation: converting tacit knowledge into explicit knowledge;
- Combination: transforming explicit knowledge into future explicit knowledge by integrating different bodies of explicit knowledge; and
- Internalisation: transferring explicit knowledge into tacit knowledge.

Table 3.3 Methods to carry out the knowledge converting modes

Socialisation	Externalisation	Combination	Internalisation
Brainstorming, informal meeting, discussions, dialogues, observation, training.	Meeting, workshops, building hypotheses, models.	Virtual library, reports, publications, conferences.	Facilitation skills, knowledge zone, client/contractor feedback review, development counselling.

The methods used in different knowledge converting modes can be used formally or informally in the project risk management as shown in Table 3.3. Most of these methods rely upon non-numerical information. Therefore, qualitative methods are more suitable in coping with knowledge in relation to project risks.

A range of knowledge management tools, both IT and non-IT, are used to facilitate knowledge sharing. Qualitative methods in risk analysis have been used for some time to deliver a great deal of relevant knowledge. These include examining historical documents, brainstorming and Delphi. The following are three methods that project participants use to acquire and share knowledge.

Reading and writing-through paper work

Examining historical documents is the most essential and basic approach to acquire knowledge relating to project risks, as shown in Figure 3.3 knowledge transfer occurs along with the transfer of project documents and it is shared when other project participants read it. Historical data is valuable for obtaining a perception of risks in previous projects (Artto, 1997). The previous project data, records, common causes of failure, plus the project managers' knowledge, experience and judgement can provide information useful for consideration when carrying out risk identification. Therefore, it is necessary to learn from post experience, in terms of both success and failure. Furthermore, historical data provides an initial list of possible risks for a brainstorming discussion.

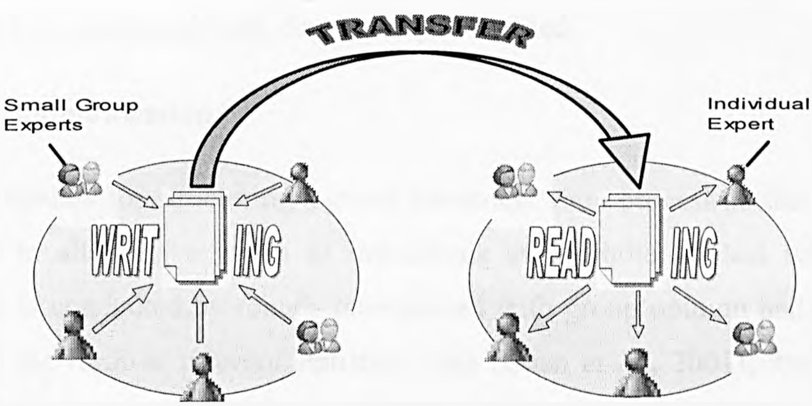

Figure 3.3 Transferring and sharing knowledge by reading and writing

Talking through face to face communication

Brainstorming is a typical method of people sharing and transferring knowledge through face to face communication. Figure 3.4 illustrates the format of this type of method. Small group of experts/ individual experts sit around a table to share their knowledge.

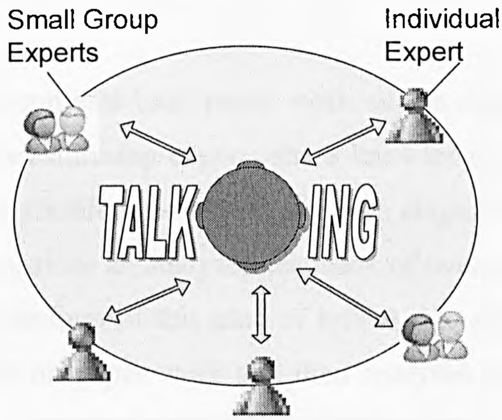

Figure 3.4 Sharing and transferring knowledge by talking

Formal brainstorming is sometimes used in this context. Outlandish suggestions are encouraged (Raftery, 1994). Brainstorming involves the collective generation of ideas by a group comprising key project personnel and others in an environment free of criticism (Merna and Njiru, 2002). The underlying principle is that group thinking is more productive than individual thinking. Brainstorming encourages wide ideas and avoidance of criticism of 'silly' ideas. The most promising ideas generated are then selected, built on, and combined, developed and verified.

Integrated communication

Delphi is a method for structuring a group communication process so that the process is effective in allowing a group of individuals as a whole to deal with complex problems. It is conducted by rounds interspersed with group opinion and information feedback in the form of relevant statistical data (Chan *et al.*, 2001). Panel members remain unknown to one another and respond to a series of questionnaires. The iterative nature of the procedure generates new information for panelists in each round, enabling them to modify their assessments beyond their own subjective opinions.

The first round of Delphi identifies the selection risk factors according to the experts' responses. The result of this analysis is given statistically. In the second round of Delphi, the result of first round is given to the experts to ask if they want to change their opinions in view of the feedback, so as to refine the selection risk factors. The third round provides a classified list of refined risk factors and asks the experts to evaluate the impact level of the risk factor and opportunity of occurrence. Similarly to round two, round four refines the level of risk factor.

Examining historical documents uses paper work as the medium to indirectly share knowledge, while in brainstorming experts share knowledge face to face which does not need any medium to transfer knowledge. As each stage, Delphi uses a coordinator to collect opinions and provide an analyzed feedback of overall knowledge for experts. The coordinator is the medium of this kind of knowledge sharing. He communicates individual experts based on paper work and then analyses and reports back to them after a highly intelligent appraisal of the raw knowledge. This integration improves the quality of knowledge sharing.

3.3.2 Risk Register

Having realized the importance of the role of information and knowledge in project risk management, nowadays a risk register is gradually used as a means of accumulating information and knowledge in relation to the project risks in an organization or firm. It is an easy to understand method of documenting perceived risks and their importance, and recording actions taken to manage them. It is used as an effective means of documenting perceived risks and their importance, and recording actions taken to manage them. It is used by some construction firms, such as Amec Civil Engineering, Bovis Lend Lease Ltd, Ove Arup and Partners.

A risk register can be a very simple document, and is a powerful means of communication when information is to be shared between the various organizations, or indeed between parts of the same organization working on a particular building or construction project (Clayton, 2002). The risk register itself is an extremely effective tool to enable everyone involved in the project to consciously evaluate and manage the risks as part of the decision making process. It also provides a platform on which the mitigating actions and decisions can be made in the future through ensuring a greater understanding and acceptance of the visible risks.

A risk register consists of three entities, the register of the risks itself, which is the main focus of the system, and two supporting documents, to include information on the risk owner and the risk reduction and/or mitigation plan (Patterson and Meailey, 2002). Williams (1994) states that the risk register has two main roles: the first is that of a repository of a corpus of knowledge; the second role is to initiate the analysis and plans that flow from it. As such a risk register should be used to keep track of the risks of the project (Williams, 1993). Chapman and Ward (1997) state that, to enable the documentation of the sources of the risk and their responses, as well as their classification, "the risk register identify phase involves compiling a list, log or register". Within this they identify that the documentation produced through the utilization of project risk management can be regarded as a "by-product rather than a central concern" (Chapman and Ward, 1997). Ward (1999) takes this further when he states that the purpose of the summary risk register is to help the project team review project risks on a regular basis throughout the project. Barry (1995), on the other hand, uses the risk register as "a comprehensive risk-assessment system". Within this, the risk register is used as a formal method of identifying, quantifying and categorizing the risks, as well as providing the means of developing a cost-effective method of controlling them (Barry, 1995).

A risk register or risk log, that assists a project's lifecycle is a mechanism for the management of risk information on a specific project (Department of Premier and Cabinet, Tasmania, 2002 & Construction Industry Research and Information Association, 1996). Contents that have been suggested for registers include information about the timing of risks, details of risk interdependencies, and reporting on the nature of risk impacts and risk ownership.

Williams (1994) and Carter *et al.* (1995) state that the project risk register should be held within a database system. Several risk register systems were developed, mostly in the form of Microsoft Access databases. Risk Radar (Integrated Computer Engineering Inc., 2002) allows the input of information on identified risks, qualitative information about their impacts and probabilities and details of mitigation and contingency steps. SiteRisk (Anderson and Madsen, 2001) contains details of construction activities, risks, and events in different tables which are linked through relationships in the database. The Risk Register Database System developed for the automotive industry (Patterson and Neailey, 2002) contains information such as risk ownership, assessments of probability and impact of the risk and phase/time by which the risk should be evaluated. The system allows reports which act as status updates on the risks of the project to be generated. The 'Risk Register' software developed by Incom Pty Ltd. (2004) provides functions for recording information on the category of risks, its association with assets, and information on the likelihood and consequences of risk. The software also has the capability of providing graphical outputs that show the distribution of risk over time. A spreadsheet-based software termed RiskCom allows the user to record information during the different stages of risk management (Hall, Cruickshank and Godfrey, 2001). Pre-Programmed help functions that provide instructions and information on different stages of the risk management process guide the user throughout its different stages.

3.3.3 Artificial Intelligent Model

As most data storage and complex calculation demands have been met, the research focus of computer science has now moved to knowledge management, helping humans coping with intricate problems with non-numerical information. This is similar to the development of the project risk management approach. In the 1980s, the quantitative theory was successfully established; then in the 1990s, the relevant theory was integrated into various risk analysis software. More recently the qualitative stage of risk management, such as risk identification, has become more important, as it decides whether or not the quantitative stage carries on and what the elements of the quantification process are. Obviously, the main purpose of risk identification is to gain knowledge of uncertain situations and potential problems that might arise in a given project.

The well-known approaches that make use of advances in artificial intelligence, such as rule-base reasoning, case-based reasoning, machine learning, and fuzzy logic, have been advanced as tools for the management of knowledge and information related to risks. Among these is a knowledge-based risk identification system (Leung et. al., 1998) that uses rules to identify risks related to extra high voltage transmission line construction projects. Kangari (1989) presents an integrated knowledge-based system for construction risk management using fuzzy logic to calculate overall risk of a project by combining values for different risks that are provided by users. Stone et al. (1990) addressed a system for managing risks in Civil Engineering by Machine Learning from previous failures to future projects, and to carry out this in a manner that facilitates the expression of the inherent uncertainties. Hashemi and Stafford (1993) developed a risk assessment system by Artificial Neural Network (ANN). A three layer Backpropagation (BP) network was set up to perform risk assessments on a set of toxicological data and gave decisions similar to those given by experts. Al-Tabtabai and Alex (2000) modeled the cost of political risk in international construction projects within the BP network. Chen (1998) combined different neural networks in her risk assessment and contingency allocation system. Knowledge about project risks is captured in templates that set out typical activities of specific project types in the Technical Risk Identification and Mitigation System (Best Manufacturing Practices Centre of Excellence, 2001). Actions that act as risk mitigators are attached to each element of the template. Reasoning about the level of risk is based on the level of compliance of each action.

3.4 LIMITATIONS OF RISK KNOWLEDGE

3.4.1 Knowledge in Risk Identification Methods

As explained Chapter Two, to improve risk management it is now more important to improve risk qualitative analysis, especially risk identification. Williams (1994) advocated a 'risk register' component in the management system of a project to generate an accessible database of risk experiences. Edwards and Bowen (1998) concluded: "risk management techniques were only useful as the willingness of the project participants to become knowledgeable and skilled in them". Chapman (1997) noted "for both parties to manage their risks effectively it may be important to move towards a cooperative shared information approach to management." Thus to manage information and knowledge relating to project risks is the key to improving risk identification.

Ugwu, *et al.* (2004) addressed previous interaction with some firms in the UK construction industry, revealing that the primary source of construction knowledge and expert experience remains with the project team members who have had experience of installing designed facilities. Kermally (2002) stated that without effective communication, tacit knowledge remains tacit and organizations lose out. The knowledge sharing and transferring methods discussed above have various limitations:

- Project documents may not be well recorded, so might not able to provide as much information as project participants expect. For instance, only 438 of 2791 World Bank projects in transportation, energy and mining, and public sector development have any risk assessment in the project appraisal document.

- Knowledge relating to project risks is difficult to acquire because construction projects have a long life cycle. As a result the number of projects in which an expert can be involved is limited. Moreover, because in the difference of background and experience, knowledge transference between experts is another issue. As humans are influenced by psychological factors and their environment, it is difficult to transfer knowledge from one to another in exactly the same way. Any loss of knowledge during the transfer may lead to inaccurate judgment and decision making.

- Delphi somehow overcomes peer pressure in brainstorming but the availability of experts can not be guaranteed due to illness, moving to different jobs, tiredness, making irrational decision and preoccupation with other matters. Sometimes risk experts are scarce, hence become very expensive and beyond the means of the project team. Furthermore, documenting human expertise is extremely difficult and time consuming.

- To solve the above problems, artificial intelligent systems are also employed in a variety of research to improve risk identification. Leung et al. (1998) introduced a rule-based system to identify project risks. However the rule-based systems are too restrictive to handle tacit knowledge (Watson, 2003). In addition, these rules may not be fully understood or accepted by the other project participants. The knowledge management activities are identified as generation, propagation, location, capture, access, maintainence and use of knowledge.

- Carter *et al.* (1995) provides an extremely descriptive version of a "risk register", and they state that "the methodology calls for a simple database of the risks in a project to be constructed". However there are various types of logs, forms and registers stated in the methodology and although examples are given, unless they were all kept in electronic format, their maintenance could be repetitive.

3.4.2 Influence of Human Behaviour

Human factors refers to "individual, project team and organizational factors which influence the behaviour of people and the climate at work, in a way which can increase or decrease the productivity of a construction project" (Thevendran, 2003). From this definition human factors are categorized into three main groups, which are further divided into 3 sub-categories. These are as follows:

- Individual factors: capability, knowledge and skill, stress, motivation, emotional and cultural;
- Project team factors: management, communication and coordination, task and supervision; and
- Organizational factors: system and procedures policies and standards.

Human factors are unequivocally the single most important element that can affect a project's success (Thevendran, 2003). More recently firms have incorporated knowledge management methods and practices into construction management. Knowledge management in relation to project risks must not neglect human issues and people's management practices.

Risk identification techniques cannot accurately forecast the effects of human factors because of their complex, unpredictable and qualitative nature (Redmill, 2002). These techniques that have been developed are useful in dealing with overt or conspicuous risks factors but they cannot adequately address the risks of human factors, which are widely recognized as making a substantial contribution to construction project failure (Oldfield and Ocock, 1997). The processes in current risk identification involve subjectivity. Kumaraswamy and Thorpe (1996) state "qualitative judgments can be compared, rated and ranked to reduce the subjective elements, but a residual subjectivity can only be minimized rather than eliminated."

There is always uncertainty, the need for judgment, considerable scope for human bias, and inaccuracy. The results obtained by one risk identification are unlikely to be obtained by others starting with the same circumstances (Redmill, 2002). Lowrance (1980) pointed out the estimate of risk, whether made by scientists or lay people, cannot escape containing elements of subjectivity.

Wharton (1992) advises that failure to cope with uncertainty in the management of technological risk abound. Their causes include overconfidence in scientific knowledge, the underestimation of the probability or consequences of failure, not allowing for the possibility of human error and plain irresponsibility concerning the potential risks to others. Thus, we should recognize and allow for the subjectivity inherent in our analysis and decision-making. Formal models clearly have a role to play in revealing the blind spots in intuitive reasoning, particularly when the complexity of the decision makes it opaque. We are less often able to base decisions on past experience because of the uniqueness of many modern construction problems which demand a more analytical approach (Flanagan and Normen, 1993).

In summary, risk identifications depend on the subjective choice of techniques, and each technique not only carries its own propensity for error, but also is based on human judgment. If the adverse effects of subjectivity are to be reduced, it should be determined at the definition-of –scope stage which techniques are most appropriate, given the nature of the system to be studied (Redmill, 2002).

The main purpose of risk identification is to cope with uncertain situations and find potential problems, while uncertainty sources from lack of knowledge. Hence, knowledge accumulating and sharing is the key to breaking the bottleneck of risk identification. Group techniques, such as brainstorming and Delphi, might broaden the perspective, "but the limitations of human processing information still will often preclude optimal decision-making" (Pender, 2001). Psychological research finds that on average humans make future planning decisions based on the three most recent decisions made by the same manager. According to Pender (2001), it seems that about nine decisions attributes that a person can effectively encompass each time, which illustrates that human have a limited information processing capability. Consequently, they cannot directly deal with complex problems even though the information may be available in some form. Furthermore, psychologists also found that by modelling a judge's decision-making process, the results provided by using that model was more accurate than the judge's own decision (Pender, 2001). This reflects the unsteady nature of human decision.

Moreover, the result of risk identification is the input of elements of risk analysis. It includes a volume of textual information, which numerical techniques find hard to handle. There are many techniques for risk identification, and all depend on human observation, judgment, and creativity. As well as being the key attributes of an effective study, they introduce subjectivity and therefore the potential for bias. Currently risk identification is largely based on the subjective judgment of humans. Decision-making is usually based upon incomplete, contradictory information. For such work, present procedures largely depend upon the involvement and the empirical knowledge of the experts. However, the restriction of humans processing complex information, potential subjective influences, and experts' availability, constrains the appropriate application of risk management. Experience associated on the individual experts. However, when carrying out risk identification for a particular project, suitable experts might not always be available due to a conflict with other duties or illness. Also the experience and knowledge of an expert might be lost if the expert died or retired. In addition, subjective judgment and psychological influence usually accompanies the experts' decision, which is almost unavoidable.

3.5 APPROPRIATE APPROACH

3.5.1 Potential Concept Model

Knowledge management literature is relevant to this study since it delineates between different forms of knowledge generation, notably explicit and tacit. According to Ugwu, *et al.* (2005), how effective the techniques are in extracting explicit and tacit knowledge embedded within an organization's cooperated memory is an issue which should be considered when developing a knowledge based model. In the context of developing a risk identification model, firms may acquire knowledge by the capturing, codifying, storing and distribution of knowledge, and through practice guide the decision-makers.

A method is useful only if the user can accept it. To improve risk identification in construction projects, we need an easily understandable, easily accessible, easily communicatable, commonly acceptable and reliable approach. Risk identification is a process of managing knowledge and information. But knowledge can be generated within one project and then buried in unread reports and arcane filling systems, or lost because people move on (Carrillo and Anumba, 2002).

In terms of choosing an appropriate intelligent technique, the difficulty of acquisition of knowledge and necessary data, explanation capacity, difficulty of development and appropriate domain are the main criteria under consideration. Different techniques have their pros and cons. ANN is particularly good at coping with numerical information; rule-based reasoning is suited to highly structured, well understood domains which change little over time; case-based reasoning goes well with poorly understood or highly dynamic domains where past decisions may form the basis for sound future decision making; and fuzzy logic has the advantage of dealing with explicit knowledge that can be explained and understood.

Managing knowledge relevant to project risks will help capture, accumulate, access, exchange and implement this knowledge in coping with risks in new projects. To improve current risk identification, ideally we can find an expert who has sufficient experience to facilitate the risk analysis and decision making. The knowledge accumulated by this expert contributes to the knowledge on which identification, analysis and decisions are based. However, in reality such an expert is difficult to find; instead we try to employ intelligent computer systems to take on this role.

To solve the above problems, artificial intelligent systems are also employed in a wide variety of research to projects in order to improve risk identification. Leunga et al. (1998) introduced a rule-based system to identify project risks. However the rule-based systems are too restrictive to handle tacit knowledge (Watson, 2003). In addition, these rules may not be fully understood or accepted by the other project participants. The knowledge management activities are identified as generation, propagation, location, capture, accessibility, maintainance and use of knowledge.

3.5.2 CBR is a Method for KM

The main knowledge management activities were identified as the acquisition, analysis, preservation and use of knowledge. CBR in particular matches the cycle of knowledge management. Figure 3.5 explains how the CBR can meet each of these requirements by superimposing steps of the CBR cycles upon Boisot's initial model as discussed in Figure 3.2.

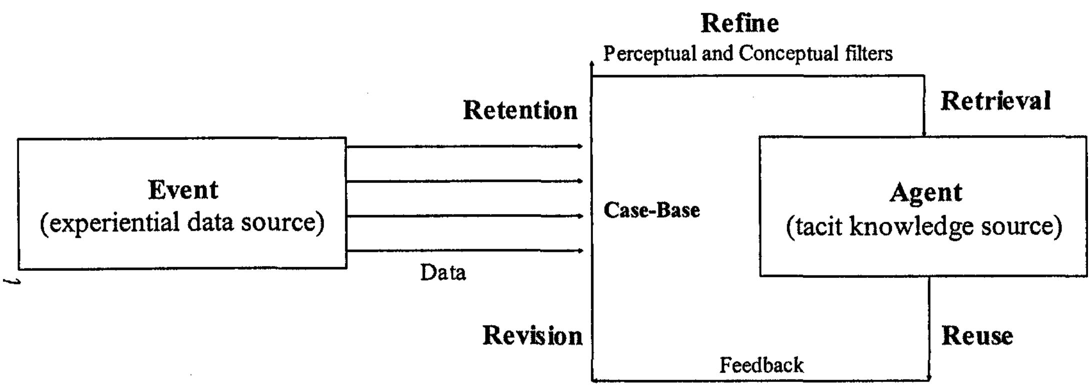

Figure 3.5 The CBR-cycle supporting knowledge management

Now it can easily be seen that the event is the episodic data that comprises a case. The case is retained in a case-base but typically undergoes some pre-processing and filtering. The agent or the reasoner retrieves the case to solve some revision or adaptation of the case's solution, resulting in new episodic data and hence a new case being created. The new case is reviewed by comparing it against cases already retained in the case-base and if it is judged useful it is retained. In addition, the use of the case may result in the case-base's indexing scheme, or the feature weights being refined. This completes the knowledge management cycle indicated on Boisot's original diagram.

In construction there are many situations where detailed information to evaluate uncertainty is not available. The intelligent systems, like humans, solve complex problems despite limited and uncertain knowledge and their performance improves with experience. Jung et al. (1999) explained that the case-based reasoning technique is good for the risk analysis process because it is useful for tasks that are experience-intensive, that lead to inconsistent outcomes, that have incomplete rules applied to them, and for which it is hard to acquire domain experience. In risk identification, knowledge is gained from documents or participants of previous projects. Case-based reasoning can be used to identify risk from non-standard situations, and can contribute further evidence to support decision making.

The intended risk identification model to cope with knowledge in relation to project risks was developed, based on the algorithm of case-based reasoning. This resembles just like human reasoning, solving the problem by referencing previous experience and saving the new solution as knowledge for future use. CBR is particularly applicable to highly dynamic or poorly-understood domains, or where expert knowledge is difficult to divine or encode. It may learn from preview experience and can continue the learning process as it is used and developed. The power of this model can be harnessed to help human experts solve more quickly a complex problem of risk identification. This will reduce the work load of human experts in accumulating complex and large scale construction information and making complicated reasoning, and provide sufficient support for decision making of risk management.

3.6 SUMMARY

Risk identification is a qualitative process which is based on a great deal of information and knowledge. To improve qualitative analysis, information and knowledge management is a key issue. There is a dynamic relationship among data, information, knowledge, and wisdom. Risk identification processes a range of raw data and information on the project and transfers some of it to knowledge. Whether or not it is advisable to share and transfer the knowledge is important for risk identification.

Current risk identification methods are largely based on previous experience. Well known risk identification methods, such as a questionnaire/checklist, brainstorming, examining historical documents and Delphi are all based on previous knowledge or experts' experience. However, experts' bias and availability may directly influence the result• of risk identification as well as the risk analysis results. Furthermore, psychological researches found that people make decisions on the futher based on the three most recent decisions made by the same manager, and there are about nine decision aspects that a person can effectively encompass each time. This illustrates that humans have limited information processing capability. This limited information processing capability means that they cannot directly deal with complex problems, even if the information is available in some form. Psychologists also found that modelling a judge's decision-making process resulted in a more accurate outcome than the judge's own decisions. As a result, apart from collecting knowledge, overcoming the limitations of human reasoning is another core issue if risk identification is to be improved. A knowledge based system that mimics the human reasoning process to solve a current problem by referencing previous similar situations becomes a suggested approach to solve the current problem of project risk identification.

CHAPTER 4

RESEARCH METHODOLOGY

4.0 INTRODUCTION

This chapter introduces the research methodologies that have been applied in order to achieve the aim and objectives of this research. The definition of research is discussed and the core points of the research are found to be experience and reasoning. Then a seven-step process to complete the research is explained. Following this guidance, the proposed research approach to achieve this study is described, from the formulation of a research problem, managing the research data, to selecting the research method. The later section of this chapter introduces the research methods that are used to carry out this study.

4.1 RESEARCH METHODOLOGY

4.1.1 Definition

The word 'research' is composed of two syllables, 're' and 'search'. The dictionary defines the former as a prefix meaning again, anew or over again and the latter as a verb meaning to examine closely and carefully, to test and try, or to probe. Together they form a noun describing a careful, systematic, patient study and investigation in some field of knowledge, undertaken to establish facts or principles (Grinnell, 1997).

Grinnell (1997) further stated "research is a structured inquiry that utilises acceptable scientific, methodology to solve problems and creates new knowledge that is generally applicable".

Merriam-Webster English Dictionary defines research as studious inquiry or examination, especially investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws.

According to Kerlinger (1986), "scientific research is a systematic, controlled empirical and critical investigation of propositions about the presumed relationship about various phenomena".

Leedy (1989) defined research from a more utilitarian point of view starting that research is a procedure by which we attempt to find systematically, and with the support of demonstrable fact, the answer to a question or the resolution of a problem.

From these definitions it is clear that research is a process for collecting, analysising and interpreting information in order to answer questions. Kumar (1996) generalized certain characteristics of a quality research process. "It must, as far as possible, be controlled, rigorous, systematic, valid and verifiable, empirical and critical". Walliman (2001) stated that "what is certain is that there are many different opinions about and approaches to research". However, as a means of achieving a greater comprehension of our world, research was distinguished from two other basic and ancient means, those of experience and reasoning.

Experience comes from in knowledge and understanding gained either individually or as a group or society, or shared by experts or leaders, through day-to-day living. Reflective awareness of the world around us, present to a degree, provides invaluable knowledge. The immediate form of experience is personal experience, the body of knowledge gained individually through encountering situations and events in life. When solutions to problems are not to be found within the personal experience of an individual, people may turn to those who have wider or more specialist experience for advice.

Knowledge gained from experience forms an essential aid to our understanding and activities in everyday life. However, it does have several limitations as a means of methodically and reliably extending knowledge and understanding the world. This is because learning from experience tends to be rather haphazard and uncontrolled.

Conclusions are often quickly drawn and not exhaustively tested, 'common sense' is invoked as self-evident, and the advice of experts is frequently misplaced or seen as irrelevant. Despite these shortcomings, experience can be a valuable starting point for systematic research, and may provide a wealth of questions to be investigated and ideas to be tested.

Reasoning is a process that draws conclusions by chaining together generalized logical rules, starting from scratch. There are three basic types of reasoning: deductive, inductive and a combination of both called inductive/deductive. An argument based on deduction begins with general statements and, through logical argument, comes to a specific conclusion. A syllogism is the simplest form of this kind of argument and consists of a major general premise, followed by a minor, more specific premise, and a conclusion, which follows logically.

Inductive reasoning starts from specific observations and derives general conclusions there from. The value of inductive reasoning was revealed by Bacon in the 1600s. By careful and systematic observation of the events in the world around us, many theories have been evolved to explain the rules of nature.

However, deductive reasoning was found to be limiting because it could only handle certain types of statement, and could become increasingly divorced from observation and experience. Purely inductive reasoning proved to be unwieldy and haphazard, and in practice was rarely applied to the letter.

When inductive and deductive reasoning were combined to form inductive/deductive reasoning, the to-and-fro process of developing hypotheses inductively from observations, charting their implications by deduction, and testing them to refine or reject them in the light of the results, formed a powerful basis for the progress of knowledge, especially of scientific knowledge. It is the combination of experience with deductive and inductive reasoning, which is the foundation of modern scientific research. Three characteristics of research can be seen to distinguish it from gaining knowledge purely by experience or reasoning (Walliman, 2001).

Gaining experience is an uncontrolled and haphazard activity, while research is systematic and controlled. Reasoning can operate in an abstract world, divorced from reality, while research is empirical and turns to experience and the world around us for validation.

Unlike experience and reasoning, research aims to be self-correcting. The process of research involves rigorously testing the results obtained, and the methods and results are open to public scrutiny and criticism.

4.1.2 Research process

Kumar (1996) introduced an eight-step model to carry out the research process. To consider the characteristics of this research, the author modified the model to a seven-step process as shown in Figure 4.1. The tasks identified in arrows are the operational steps to be followed in order to conduct a study. Topics identified in rectangles are the required theoretical knowledge needed to carry out these steps. The tasks identified in circles are the intermediary steps that must be completed to proceed from one step to another.

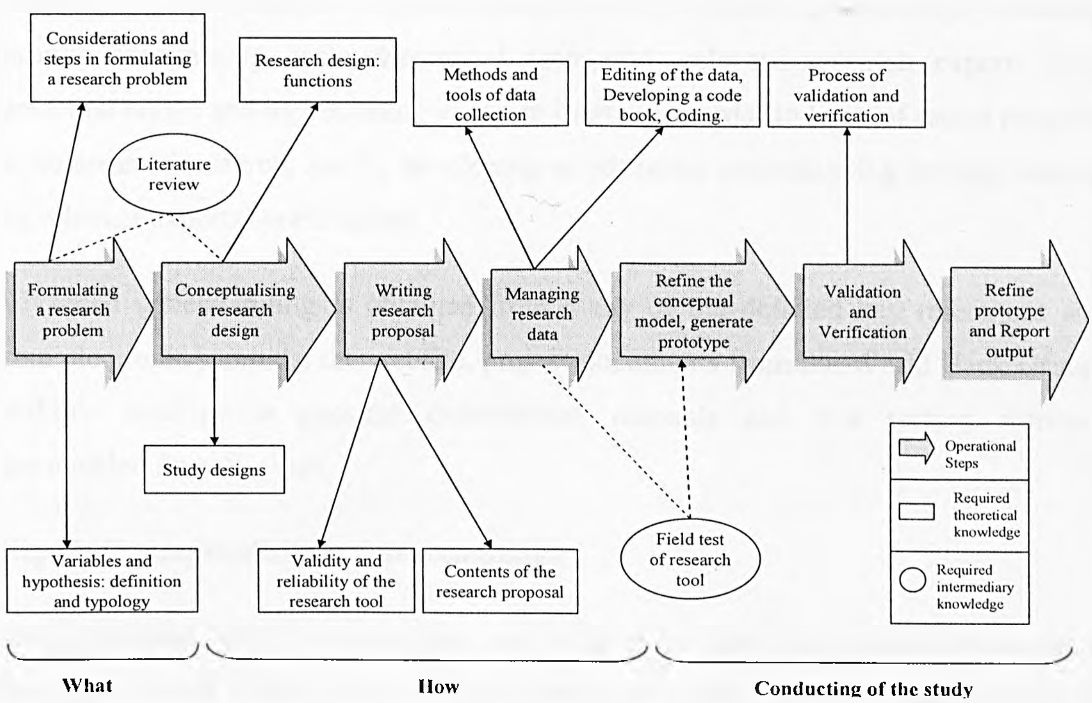

Figure 4. 1The Framework of Research Process (Source: Kumar, 1996)

Step I: Formulating a research problem

To formulate the research problem, a great deal of reading and analysis of relevant literature has been undertaken with a view to developing a basic modelling system for risk identification by hybrid intelligent reasoning. Literature was chosen as broadly as possible with the hope of obtaining an overview of the whole industry. Weak assets that can be improved by drawing lessons from business, project management and first hand practical experience and aspects which require innovative techniques to explore a new development space could be determined from this overview. This procedure should highlight the gap between the demands and limitations of the project.

The reading materials included academic papers, books, website information, information from relevant news groups and mailing lists. The content is divided into two broad categories: the first category includes knowledge of project management such as estimating, change, forecasting, planning, risk management, contract conditions and standards; the second category includes information technology, covering Software Engineering, Windows Programming, ANN, Expert system, Case-based Reasoning, Rule-based Reasoning, AHP, and Machine Learning.

The chain of knowledge is systematically built by collecting references resourced through recognizing basic theoretical facts and, relevant research experts from elemental books and by searching in recent journals in order to learn of recent progress in the areas of research, and by developing an advanced understanding through reading the relevant experts' publications.

Advanced understanding is obtained from many of the detailed key references and from electronic journals, the website, project documents from the World Bank Group, artificial intelligence package descriptions, manuals and free testing software downloaded from the web.

Step II: Conceptualising a research design

Having decided what to study, the next stage is to plan the course of the study. Kerlinger (1986) states that a research design is a plan, structure and strategy of investigation so conceived as to obtain answers to research questions or problems. The plan is the complete scheme or program of the research; it includes an outline of what the author will do from writing the hypotheses and the operational implications to the final analysis of data.

Among the various AI algorithms, consider the features of project risk identification, an intelligent reasoning technique which matches the reasoning process of human beings is chosen as the basic modelling paradigm. In this study, the risk identification modelling framework is built to understand the following work, including the kind of programming and tools that are needed, which data should be collected, and how the information is connected and integrated.

Several packages were considered when developing the system, such as Dreamweaver, Visual C++, Delphi, Oracle Database and Visual Basic. Finally the packages used in system design include ART*Enterprise, Microsoft Office (Access, Excel and Word) and Borland C++.

The modelling system consists of three components: a graphic user interface, an application model, and a knowledge base. The system development begins with familiarization of the software's environment by writing some small examples of programmes in different sections of the package; then a simple system is developed to integrate the separate sections. The whole procedure follows the processes of planning, analysis requirement, concept design, detailed design, programming, testing, and maintenance.

Step III: Writing a research proposal

A research proposal details the operational plan for obtaining answers to research questions. A transfer report works as a research proposal. It illustrates the problem identified in the previous stage through literature review, describes the method for carrying out the research and clarifies the proposed framework to develop the system. The potential limitations of this modelling are also explained. In the end, the proposed schedule to guide subsequent research is provided.

Step IV: Managing research data

Once the concept of research has been identified, the kind of data that is needed to carry out the study should be decided and the ways to collect and analyse data were chosen.

The construction of a data collection tool is an important aspect of research as it determines the nature and quality of the information. This organises input to the study and output is entirely dependent on it. The availability of data is the main problem in many research projects. Furthermore, the choice of a particular method for collecting data depends upon the purpose for which information is collected and the skills of the researcher. Due to some unforeseen issues in this research, such as a change of manager in the first chosen company, contact with alternative companies before finally using the World Bank project data, some problems were encountered.

Step V: Refine the conceptual model and generate a prototype

When the real data is put into the intended frame of the conceptual model, potential problems and difficulties were explored. To identify these issues, a more feature focused, reliable, and flexible prototype was generated.

Step VI: Validation and verification

When the development of the system had been completed, validation and verification of the system was carried out by following a systematic evaluation scheme.

Verification and validation, also referred to as evaluation, are formal methods for testing a computer program (Ng, 1996). Gupta (1991) has classified verification as 'white box' testing designed to determine if the system completely and accurately implants the user specifications, while validation was classified as 'black box' testing designed to determine whether the system meets the user's need. In essence, verification determines whether the system was built correctly and validation determines if the right system was built.

System verification includes checking the consistency and completeness of the system (Botten et al, 1989). According to Morell (1988), a system is inconsistent if it asserts something that does not reflect the modelled domain. The aspects influencing the consistency of a CBR system include the consistency of cases (Curet, 1995) and the features used to represent the cases (Kriegsman and Barletta, 1993). An incomplete system is one that cannot respond to all possible situations that may arise within the domain (Cragun and Steudel, 1987).

In terms of validation, the sensitivity, accuracy, usability, and efficiency of the system must be checked by the developer (Marcot, 1987). Sensitivity is the effect on system performance due to the changes of system parameters. The term accuracy means how well a system reflects reality. Usability is about the human interaction and user friendliness of the system. The efficiency can be examined by checking the running time of the system (Rajamoney and Lee, 1991).

Step VII: Refine prototype and report output

When the system was successfully validated and verified, an analysis was provided. Meanwhile, as a result of its evaluation the initial prototype was refined. A report of the system's output and further analysis were then provided.

4.2 RESEARCH DESIGN

Research is often presented as a fixed, linear series of stages, with a clear start and end. A research design has four core problems: what questions need to be asked, what data is relevant, what data needs to be collected, and how the results are to be analysed.

This thesis is written in four sections: Introduction, Literature Review, Model Development and Testing, and Conclusion. The first part consists of introducing the background to the research, setting up the aim and objectives, exploring the scope and limitations of the research. The second part makes an extensive literature review of work related to the subject of the research and demonstrates the appropriate approach. The third part comprises the processes of developing a model in order to achieve the aim and objectives and also describes the research methods used in each process. The fourth part highlight the key findings from this research, identifies the contribution to knowledge, and suggests future work that be carried out in order to improve the model and/or the method. These four steps are described in Figure 1.1.

The methodology is integrated with three components, as shown in Figure 4.2. It starts from identifying the problem and writing a research proposal, collecting and analysing research data, and finally detailing the research methods chosen to carry out the research work.

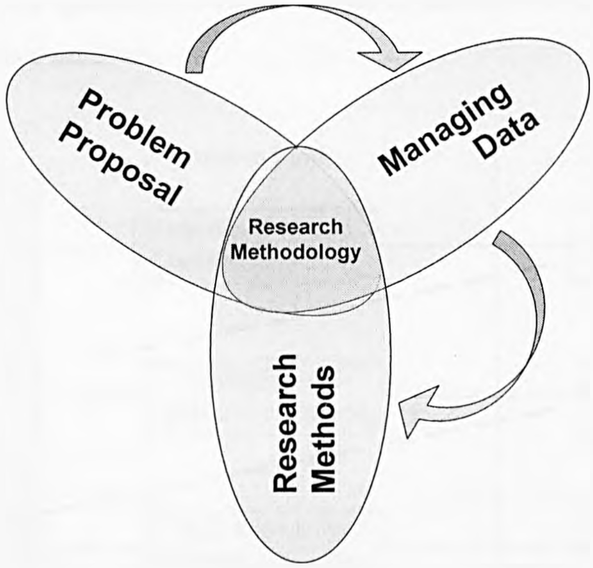

Figure 4.2 Components of research methodology

To explain each of the above components in detail, the following section outlines the overall research methodology applied to the research.

4.2.1 Identify the Problem and Prepare a Proposal

The first part of the research methodology is to identify the research problem by reviewing relevant literature, which is explained in detail in chapters two to four. Figure 4.3 describes how the knowledge chain is built up from identification of the problem to formulation of the research proposal.

The literature review begins with existing problems in construction projects. Construction projects have a poor reputation in overrun, over due and unsatisfactory quality. Analysis of potential risks at the early stage of a project can help the understanding of the scope and potential problems related to projects and support decision making. A range of risk management methods were well developed during the 1960's to 1980's. Since the 1990s, the explosive growth in computers and computer software has made possible the everyday use of a sophisticated risk analysis model. However, most of the models are based on quantitative analysis, such as Monte Carlo simulation, which is a popular fundamental theory from which most packages are derived. However, qualitative risk analysis is also important, such as risk identification. Risk identification is the fundamental stage of risk management as it decides whether or not quantitative risk analysis is carried out and what elements are used in the quantitative analysis.

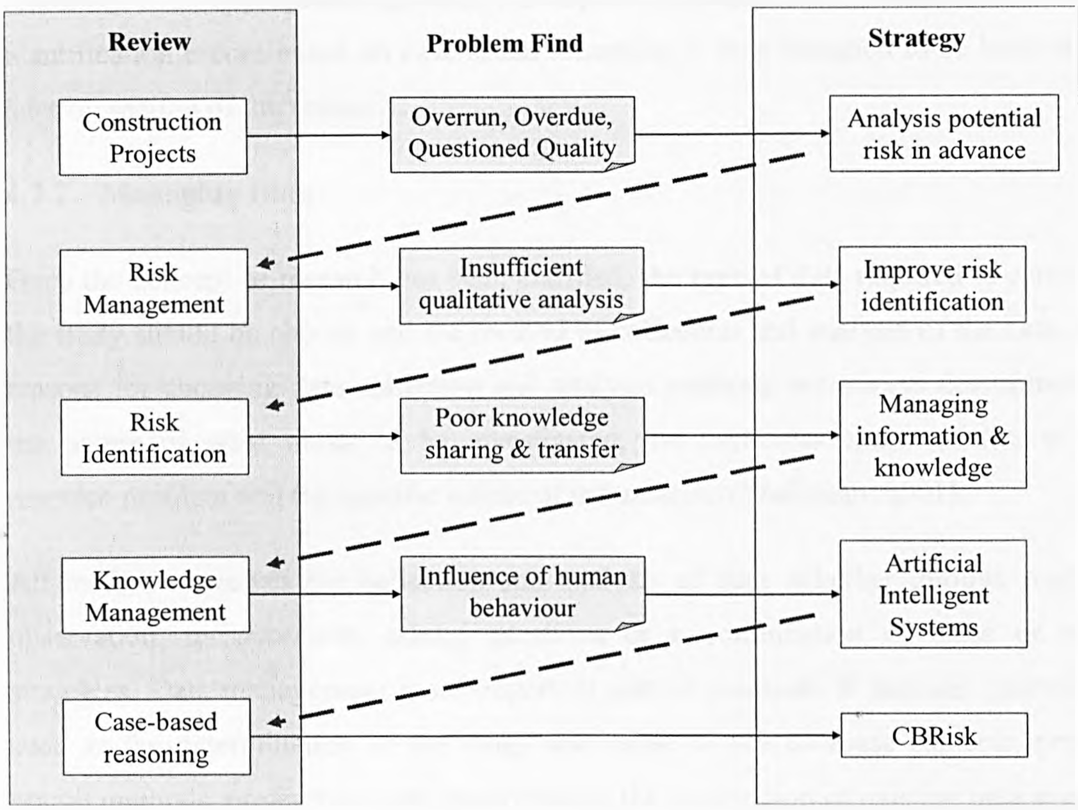

Figure 4.3 Procedure of problem setting and preparing proposal

At present well known risk identification methods largely rely on the experience of project experts. Sharing and transfer knowledge was restricted by experts' availability and job changing. As a result, subjective judgement and intuition usually accompany an expert's opinion. In addition, knowledge management is worth making a pilot for risk identification. A risk register is already used to collect information and knowledge of previous projects in several construction firms. Nevertheless, psychological research has proved that a person can effectively cope with about nice decisions attributes at a time, which illustrates that humans have limited information processing capability. Consequently, they cannot deal directly with complex problems even though the information may be available in some form. Therefore, the artificial intelligence system is considered suitable for project risk identification. Psychological research also finds that generally human make future planning decisions based on the three decisions made by the same manager, which is a perfect match for the algorithm of case-based reasoning. Meanwhile risk identification deals with a great deal of non-numerical information which is also suitable for case-based reasoning. Consequently, case-based reasoning is an appropriate approach for improving project risk identification. A risk identification model based on case-based reasoning is then intended to be built in the later on section of the research.

4.2.2 Managing Data

Once the concept of research has been clarified, the type of data required to carry out the study should be chosen and the method of collection and analysis of the data. The reasons for choosing data collection and analysis methods are always determined by the nature of what needs to be investigated, the particular characteristics of the research problem and the specific source of information (Walliman, 2001).

All research involves the collection and analysis of data, whether through reading, observation, measurement, asking questions or a combination of these or other strategies. Data management is an important part of research. It includes procedures such as the determination of the range and scope of the database subjects, present search methods, preferences and requirements; the exploration of existing data sources which are being, and can be, used to locate the research; a survey of suitable database computer programs, including design features, formats, search methods, print options and updating characteristics; and reviews of possible methods of management, maintenance, quality control and periodical updating of the database.

Data Collection

The construction of a data collection instrument is an important aspect of research as it determines the nature and quality of the information stored. During data collection it should be decided what data might be required, what sources will be used, and the availability and possible methods for collecting the data. This is the input of the study and the output is entirely dependent on it. When considering what data might be required, the source, the availability and the possible methods of collecting data are considered. The availability of data is the main problem in many research projects. Furthermore, the choice of a particular method of collecting data depends upon the purpose of collecting information and the researcher's skills. The author experienced some unforeseen issues during collection of data for this research. A managerial change in the first chosen company, time lost in contacting alternative companies, and the World Bank project documents which were used to collect the primary data.

Documentary information can take many forms and should be the object of an explicit data collection plan. For instance, the following variety of documents can be considered: project information document, project appraisal document, environment assessment document, integrated safeguards data sheet, resettlement plan, board report, and implementation completion report.

The collected data is verified with the intention of reviewing its quality. Verification consists of a review by key informants of the representation, consistency and completeness of the information, which is outlined in Chapter Eight.

Data Analysis

Data analysis should be carried out in relation to the research problem. To help judge the type of and amount of data required a decision should be made on the methods which will be used to analyse the data. In return, a decision on the appropriateness of analytical methods must be made in relation to the nature of the research problem and the specific aims of the research project (Walliman, 2001). Hakim (2000) comments that secondary analysis is any further analysis of an existing dataset which presents interpretation, conclusion, or knowledge additional to, or different from, that presented in the first report or the inquiry as a whole and its main results.

When analysing the data, the tools, techniques and resources required are taken into consideration. The author examined raw data using many interpretations in order to find linkages between the research object and the outcomes with reference to the original research questions. Throughout the evaluation and analysis process the author remained receptive to new opportunities and insights.

The tactics used in analysis aim to improve the accuracy and reliability of the findings. Specific techniques include placing information into categories, creating flow charts or other displays and tabulating frequency of events. Both quantitative data and qualitative data contribute to the development of the model. Quantitative data is most useful in understanding the rationale or theory underlying relationships. The statistic analysis method is employed in investigating both types of dataset at a profound level. Another technique is to use multiple investigators to examine the data and the patterns. When multiple observations converge, confidence in the findings increases.

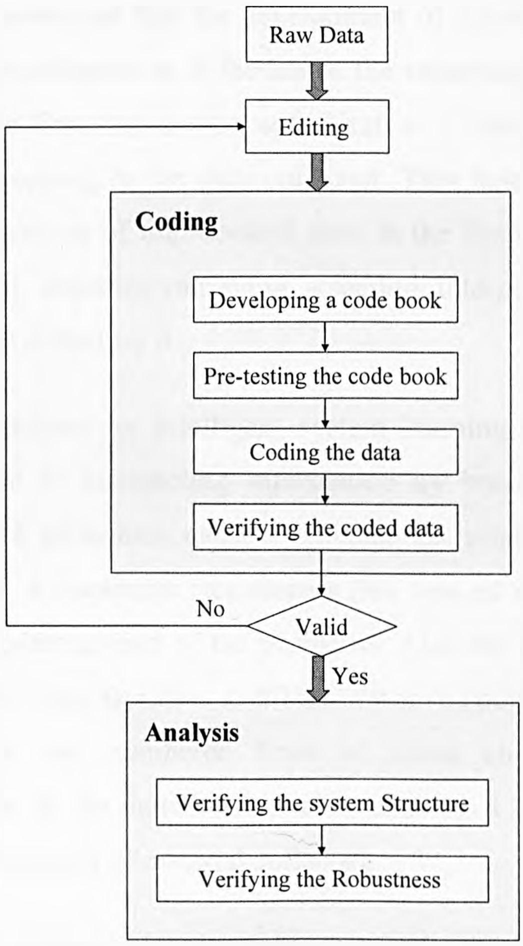

Figure 4.4 Data capture processes

Figure 4.4 shows the data processing procedure. The raw data is first edited in the appropriate format. The standard format data is then coded, which includes four processes: development of a code book, pre-testing of the code book, coding the data and verifying the coded data. This process is accompanies by developing the model by programming it. If the coded data is valid, then it will reach the analysis stage, otherwise, it will return to the editing stage and the coding will be redone until it is valid. Data analysis includes two processes: verifying the systems structure and verifying its robustness, which is explained in detail in Chapter 9.

As the data accumulates, a valuable step is to organize the shapeless mass of data by building typologies and taxonomies, that is, by identifying differences in the data and thereby forming subgroups within the general category. Using these new typologies helps to clarify the relationships among concepts.

Walliman (2001) points out that the development of a coding system is an important aspect of forming typologies as it facilitates the organization of copious amounts of data and provides a first step in conceptualization. Codes are labels or tags used to allocate units of meaning to the data collected. This helps to prevent data overload resulting from mountains of unprocessed data in the form of ambiguous words. The process is analytical, requiring retrieving, selecting, interpreting, and summarizing the information without distorting it.

As this research focuses on intelligent system learning from previous knowledge, coding is important in interpreting information by breaking the data into smaller analytical units, such as themes, causes/ explanations, relationships among people and emerging concepts. A computer can process this type of situation and can develop a better integrated understanding of the problems. Also the computer program used for analysing qualitative data facilities to fill and retrieve coded information allows codes to be attached to the numbered lines of notes and for the source of the information/opinion to be noted. This also enables a rapid retrieval of selected information from the mass of material collected.

Data analysis causes from the tools, techniques and resources used. The data analysis method can be categorized into quantitative and qualitative. Quantitative analysis is based on statistical analysis in order to verify and validate the model. On the other hand, as risk identification is concerned with human experts' experience and knowledge, as subjective human feelings and emotions were difficult to quantify, qualitative analysis methods were evolved (Cohen and Manion, 1994). As a result, and as is apparent from the statistical analysis, a case study is used to demonstrate the system's performance and an explanation of the validation of the model is also provided. A mixture of quantitative and qualitative research is usually appropriate for all types of research strategies.

4.2.3 Selection of Research Method

The selection of research methods throws up several questions: what to research, what tool to use, what data is relevant, how to process the data, and how to validate the result.

Walliman (2001) stated that models are a method of selectively mimicking reality in a form which can be manipulated. In order to make a model it is necessary to understand the system which lies behind the phenomena in reality, to understand which variables are important and how they interact. Qualitative models emphasize the relationships between entities without trying to quantify them, while quantitative models not only describe the relationships but also accurately measure their magnitude.

This research combines both qualitative and quantitative methods. Quantitative analysis uses the syntax of mathematical operations to investigate the properties of data. Qualitative research is used to construe the attitudes, belief and motivations within a subject; it is more about obtaining an inside view in order to collect resonant, fertile data to enable the development of a social framework through the dynamic process of research.

In the first phase, qualitative methods (Berg, 2001, Miles and Huberman, 1994, Strauss and Corbin, 1998, Yin, 1994) are used to validate and enrich the theoretical framework developed in the previous chapters. The theoretical framework developed in the preceding chapters thus serve as the theoretical lens or “frame of reference” in the model development inquiry.

The quantitative phase builds and tests the model based on the theoretical framework at the project level. The units of analysis are the World Bank transportation projects. Data from these projects is used to test the model development.

In general, the qualitative phase enriches the theory by grounding it with relevance and meaning. It does not test or verify the theory (Dangherty, forthcoming). Instead, we test the theory with a large of sample projects in the quantitative stage. This is a multiple-methods approach of empirical investigation used in order to arrive at a more comprehensive understanding of the phenomenon under study. The following figure shows the general research approach.

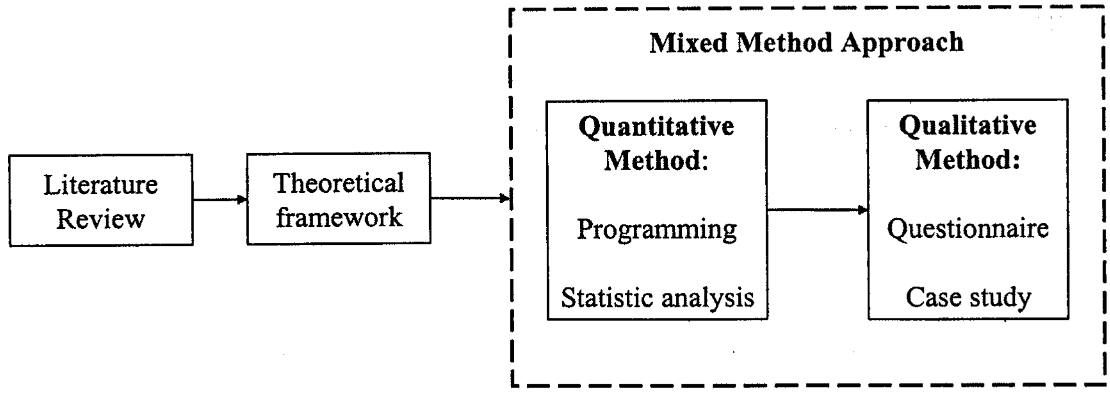

Figure 4.5 Research roadmap

As shown in Figure 4.5, the mixed methods approach is a research strategy that mixes both qualitative and quantitative data in a single research study or program (Creswel, 2003). Although this mixed methods approach only emerged in recent years (e.g. Brown, Willis, and Prussia, 2000, and Edmondson, 1999), it can be traced to the multitrait-multimethod approach of Campbell and Fiske (1959) and Jick's (1979) triangulation. Mixed methods research can be conducted in several ways (Creswell, 2003) depending on: (1) Implementation, i.e. is the data collected sequentially or concurrently? (2) Data priority, i.e. is the data collected sequential concurrently, or both? (3) Integration, i.e. does the mixing of data take place at data collection, analysis, interpretation or some combinations of these stages? And (4) theory, i.e. is there an overall theoretical perspective that guides the design?

Our approach begins with a questionnaire to verify the data used for data development. Data is collected sequentially in phases rather than concurrently. It then follows with the model design by using a variety of programming tools. Later on, a systematic verification and validation procedure is carried out to test and improve the developed model. We mix both types of data at the analysis stage where we use case study data to develop the refined framework and reuse previous project data to test the resulting framework. In addition, a broad theoretical perspective guides this research.

4.3 RESEARCH METHODS

The actual research methods used to carry out the research include both quantitative methods, such as computer programming and statistic analysis, and qualitative methods, such as a questionnaire and case study. This section introduces these methods and explains how the actual tools and approaches are used. As statistical analysis is more complicated to explain without an example, it will be fully explained in Chapter 9, verification and validation of the model.

4.3.1 Programming

One of the objectives of this research is to develop an IT based model to improve project risk identification. The tool to develop the prototype model is chosen based on the underlying philosophy and required functionality. The model design process is based on typical software engineering procedures including: concept design, data collection, detail design, coding, testing and maintenance.

The main shell to develop the system is carefully chosen through a range of comparisons and analyses. Packages used include:

- Borland C++ and Access are used to develop the model developing;
- SPSS, Minitab, Excel are used for statistical data analysis; and
- HTML, PHP and Dreamweaver are used in building the online questionnaire for verification and validation purposes.

4.3.2 Questionnaire

Questionnaires offer an objective means of collecting information about people's knowledge, beliefs, attitudes, and behaviour. This is used to collect opinion for different components of the model and comments on the model performance after the model are developed.

The questionnaires are used at three stages in this research. First of all a questionnaire is designed to collect opinions for choosing appropriate attributes for case indexing and case representation. Secondly, a simple questionnaire is used to collect the views of user friendliness tests in interface validation. Finally, for the case study a questionnaire is used to obtain feedback of the model's performance. Further opinion of the model itself is sought from the questionnaire including achieving objectives of risk identification, accuracy, the potential benefit, and ways for improvement.

4.3.3 Case Study

A case study is used for the qualitative approach. This research method is typically effective for research that attempts to understand the "how" and "why" questions, such as those stated in this proposal (Yin, 1994). Case study is also especially appropriate for studying relatively new topic areas (Eisehardt, 1989), such as the one proposed here. For case study research Meredith (1998) argues that "understanding what is achieved is only meaningful within a framework of assumptions, belief, and perspectives specified by the researcher".

'Case study' refers to the collection and presentation of detailed information about a particular participant and frequently includes account by the subjects themselves. Case study research excels at bringing us to an understanding of a complex issue or object and can extend experience or add strength to what is already known through previous research. Case study examines the interplay of all variables of a limited number of events in order to provide as complete an understanding of an event or situation as possible. This type of comprehensive understanding is arrived at through a process known as thick description, which involves an in-depth description of the entity being evaluated and the circumstances under which it is used.

The case study method is utilised because data collected can be presented in a more comprehensive manner and can be treated more carefully in the required context. In brief, the case study allows an investigation to present the key data collected and meaningful characteristics of simulation in real life event.

In the case study, a previous project is used in the developed model. A series of results from the model performance is then analyzed and explained. The system output is then used to compare the real situation to evaluate with the performance of the model. Meanwhile the potential problems and parts of it that can be improved are examined. The prototype then is refined according to the results of the case study.

4.4 SUMMARY

The research process for this study goes through seven steps: formulating a research problem; conceptualising a research design; writing a research proposal; managing research data; refining a conceptual model and generating a prototype; validation and verification; and refining of the prototype and reporting the output. The structure of how the knowledge chain is formatted is explained; the data management strategy from data collection to data analysis is introduced; and a mixed-methods design consisting of both qualitative and quantitative methodologies to carry out the research is described. Several methods of research are then outlined. The programming processes are based on the typical artificial intelligent system design procedure and are chosen for functionality. The questionnaire is used to select the attribute, user friendliness tests, and finally the gathering of feedback. Subsequently, verification and validation evaluates the performance and functionality of the underlying model, and finally, the case study demonstrates the model performance process, further evaluates the developed model, and refines the prototype.

CHAPTER 5

CASE-BASED REASONING

5.0 INTRODUCTION

According to the discussion in previous chapters, case-based reasoning is an appropriate approach for improving risk identification in construction projects. This chapter explains what case-based reasoning is and, through a comparison of other information techniques, answers why it has been chosen as the research tool. It then describes the advantages of case-based reasoning and explains CBR algorithms. Subsequently, applications of CBR in both the construction industry and construction management are introduced. In order to choose the appropriate system development shell, a range of CBR packages were evaluated and compared. The latter section of this chapter introduces the history, features and applications of the selected package.

5.1 WHAT IS CASE-BASED REASONING?

Case-based reasoning is a problem solving paradigm that solves a new problem by remembering a previous similar situation and by reusing the information and knowledge gained from that situation (Aamodt and Palaz, 1994). More specifically, CBR uses a database of problems to resolve new problems. The database can be built through the knowledge engineering process or it can be collected from previous cases.

In a problem solving system each case would describe a problem and a solution to that problem. The reasoner solves new problems by adapting relevant cases from the library (Riebeck and Shrank, 1989). Moreover, CBR can learn from previous experiences. When a problem is solved, the case-based reasoner can add the problem description and the solution to its case library. The new case that is generally represented as a pair <problem, solution> is immediately available and can be considered as a new piece of knowledge.

5.1.1 CBR Definition

Case-based reasoning has been given many definitions. Some of these definitions follow along with their source:

- Case-based Reasoning is reasoning by remembering (Watson and Marir, 1994);

- Case-based Reasoning is both the ways people use cases to solve problems and the ways we can make machines use them (Kolodner, 1993);

- Case-based Reasoning is a problem solving method that offers potential solutions to the limitation of expert systems (Lewis, 1992);

- Case-based reasoning solves new problems by adapting solutions that were used to solve old problems (Riesbeck and Schank, 1989);

- CBR is a methodology to model human reasoning and thinking, for building intelligent computer systems, and for storing previous experience (cases) in memory (Bergmann, 1996).

Broadly speaking, case-based reasoning is concerned with solving new problems by adapting solutions that worked for similar problems in the past. As with Artificial Intelligence in general, case-based reasoning deals with two different aspects of intelligence: the first motivation is to establish cognitive models in order to understand human thinking and behaviour; the second motivation is to build systems, which help to solve real-world problems.

According to Doyle (1998), case-based reasoning is different from other Artificial Intelligence approaches in the following ways:

- Traditional AI approaches rely on general knowledge of a problem domain and tend to solve problems on first-principles while CBR systems solve new problems by utilizing specific knowledge of past experiences.

- CBR supports incremental, sustained learning. After CBR solves a problem, it will make the problem available for future problems.

5.1.2 History of CBR

In 1977, Schank and Abelson's work transferred CBR from research into cognitive science (Watson, 1997). They proposed that our knowledge of situations was recorded as scripts that allowed us to set up expectations and perform inferences (Schank and Albelcon, 1977). Schank then investigated the role that the memory of previous situations and situation patterns (scripts) or memory organisation packets (MOPS) play in both problem solving and learning (Schank, 1982).

Almost at a similar time, Gentner (1983) investigated analogy reasoning that is related to CBR, while Carbonell (1983) explored the role of analogy in learning and plan generalization. Kolodner (1983) developed the first CBR system called CYRUS. CYRUS contained knowledge, as cases, of the travels and meetings of ex- US Secretary-of-State Cyrus Vance. CYRUS was an implementation of Schank's dynamic memory model. Subsequently, increasing numbers of research papers and applications were published, and CBR has grown into a field of widespread interest. It has proved to be a methodology suited to solving 'weak theory' problems where it is difficult or impossible to elicit first principle rules from which solutions may be created (Kolodner, 1993).

5.2 WHY CBR?

Chapter 3 discussed the demand of improving information/knowledge sharing and transfer in risk identification. This section answers why CBR is the appropriate approach to solve this problem through a comparison of CBR with other information techniques. It then concludes with the advantages of CBR systems.

5.2.1 Comparison with other Techniques

It is necessary to understand the strengths and weaknesses of a technique before choosing it as an appropriate research tool. Table 5.1 provides a comparison of the suitable application environments of each methodology.

Table 5.1 Methodology comparison (Watson 1997)

Methodology	When to Use	When not to Use
Databases	well-structured, standardized data and simple precise queries	complex, poorly structured data and fuzzy queries
Information Retrieval	large volumes of textual data	non-textual complex data types, background knowledge available
Statistics	large volumes of well-understood data with a well-formed hypothesis	exploratory analysis of data with dependent variables
Rule-based Systems	well-understood, stable, narrow problem area and justification by rule-trace acceptable	poorly understood problem area that constantly changes
Machine Learning	generalisability rules are required from a large training set and justification by rule-trace is acceptable	rules not required, and justification by rule-trace unacceptable
Artificial Neural Networks	noisy numerical data for pattern recognition or signal processing	complex symbolic data or when justification is required
Case-based Reasoning	poorly understood problem area with complex structured data that changes slowly with time and justification	when case data is not available, or if complex adaptation is required, or if an exact optimum answer is required

Having understood the suitable application situation of each technique, the following section compares CBR in detail with other computational techniques, including statistical techniques, rule-based systems, model-based systems, artificial neural networks, and machine learning.

CBR vs. Statistical Analysis

Several studies regarding CBR and various statistical techniques have been performed. Musgrove et al. (1996) compared the nearest-neighbour case-based approaches and linear discriminant analysis. The nearest-neighbour CBR system correctly classified treatments 87% of the time, compared to 67% for discriminant analysis. Similarly, Daimler-Benz also carried out a comparative study between linear discriminant analysis and CBR (Nakhaeizadeih, 1993). The result showed that the percentage accuracy of CBR was around 93%, while linear discriminant analysis was only about 61%. Both results suggested that CBR outperforms linear discriminate analysis from induction and nearest-Newburgh. Techniques like discriminant analysis are not suited to exploratory analysis where the independence of variables may not be known. Statistical analysis is usually applied to large volumes of well-understood data; CBR can handle the unclear data sets.

CBR vs. Rule-based Systems

Rule-based reasoning is one of the most popular techniques in intelligent systems. A rule-based system represents knowledge as a group of rules which inform us of what should be done, or what could be concluded in different situations. A rule-based system consists of a bunch of IF-THEN rules, a bunch of facts, and an interpreter controlling the application of the rules, given the facts. Rule-based programming is motivated by the observation that humans often detect a set of conditions in their environment and respond with a particular set of behaviour. In this sense, rules can be viewed like the IF-THEN statements in procedural programming: if a particular set of conditions exist, then respond with a particular set of behaviours. The typical structure of a rule can be described as follows:

$$
\begin{array}{l}
LSH \Rightarrow RHS \\
Whenever \Rightarrow THEN \\
\Pr oposition \Rightarrow Conclusion \\
Rule \begin{cases} Conditional - Fact1 :\Rightarrow Actions \;\&\; Assertions \\ Conditional - Fact2 :=> Actions \;\&\; Assertions \\ \qquad\qquad \ldots \end{cases}
\end{array}
$$

Figure 5.1 Rule-based system structure

There are two broad kinds of rule-based systems: forward chaining systems, and backward chaining systems. A forward chaining system starts with the initial facts, and uses the rules to draw new conclusions (or take certain actions) given those facts. A backward chaining system starts with some hypothesis (or goal) to prove, and looks for rules that would allow that hypothesis to be concluded, and perhaps set new sub-goals to prove. Forward chaining systems are primarily data-driven, while backward chaining systems are goal-driven.

Table 5.2 Rule-based Systems vs. Case-based Reasoning

Rule-based Systems	Case-based Reasoning
used in narrow, well understood, strong domain theory, stable over time	used in wide, poorly understood, weak domain theory, dynamic over time
knowledge representation with facts and IF-THEN rules	knowledge representation with cases
system provides answers	system provides precedents
explanation by trace of fired rules	explanation by precedents
system cannot learn, usually requires manual addition of new rules	system can learn by case acquisition
rules in rule base are patterns	cases in case library or constants
rules are retrieved that match the input exactly	cases are retrieved that partially match the input
rules are applied in an interactive cycle of micro events	cases are initial retrieved, approximating the entire solution at once, adapted and refined to a final solution
rules are small, ideally independent but consistent pieces of domain knowledge	cases are large chunks of domain knowledge, quite likely redundant

From Table 5.2, it can be seen that rule-based systems rely upon a knowledge engineer to define and encode domain specific rules from experts in the field. CBR reasons from explicitly stored knowledge in their case-base, and can explain their decisions by citing these cases. They are much better with symbolic or free-text data.

Rule-based systems require an algorithmic representation of the first principles governing good decision making within a domain. Adaptive systems generally only require pairs of experiences and outcomes (problem-solution pairs), or even real-time access to the environment itself. This makes adaptive systems ideally suited for use in poorly understood domains or within dynamic and constantly changing conditions. One such technique, CBR, is gaining more and more support in the AI and information technology communities as a common sense, though sometimes imperfect, solution to some very complex problems.

Rule-based systems use rules to guide their decision processes. Typically a knowledge engineer works with a domain 'expert' to derive the heuristics that the expert uses when solving a problem, whereas, case-based reasoning 'looks' for similarities between the current needs and previous examples of similar problems and their attendant solutions. Rule based programming is currently very popular and well developed. Most subject matter experts will clearly explain the rules they use to solve either everyday or very difficult and detailed problems. However research into human problem solving has determined that in almost all situations the 'rules' used by experts have been in part derived from a cause and effect relationship derived from previous experiences/cases.

Thus, if one has a narrow, well-understood problem that does not change with time, it may be sensible to use a rule-based system. When a problem is less well understood or is dynamic, CBR may be better (Watson, 1997).

CBR vs. Model-based Reasoning

Model based reasoning systems aim to provide the end user with a means of selecting and utilising models appropriate to their needs. The Model-Based Reasoning (MBR) process itself can be viewed as the symbolic processing of an explicit representation of the internal workings of a system in order to predict, simulate, and/or explain the resultant behaviour of the system from the structure, causality, functionality and behaviour of its components. Thus, the first step in MBR is building an accurate representation (model) of the system. Once there is a complete model and an algorithm which correctly reproduces the relationships and behaviour of the system, one can get complete, precise solutions based on a set of inputs.

The differences between model-based reasoning and case-based reasoning are listed below:

- Model based reasoners store casual models of devices or domains;
  Case-based reasoners store cases describing the way things work.

- Model-based reasoning is applicable when a casual model exists, that is, when a domain is well enough understood to enumerate a causal model;
  Case-based reasoning is applicable in both ways. The set of cases plays the role of generalized model when a domain is not well understood.

- Model based reasoning provides a means of verifying solutions, but solution generation is unguided;
  Case-based reasoning provides a means of efficient solution generation, and evaluation is based on the best case available.

CBR vs. Artificial Neural Network

Artificial Neural Network (ANN) is defined as an information processing system whose architecture is inspired by the structure of biological systems (Caudill and Butler, 1990). The artificial neuron is an approximately simulated model of a biological neuron. These artificial neurons are used to develop an artificial neural net with many interconnections among different tasks such as pattern recognition, linear optimization, speech recognition, and prediction (Mukherjee and Despande, 1995)

Like the human brain, ANNs learn from experience, generalise from previous examples and new ones and abstract essential characteristics from inputs containing irrelevant data. They deal with problems where there are complex relationships between inputs and outputs and where the input data is distorted by high noise levels. ANNs also allow self-learning, self-organisation, parallel processing, and are well suited for problems involving matching input patterns to a set of output patterns where deep reasoning is not required.

As both techniques are artificial intelligent algorithms, there are some similarities between CBR and ANN, such as their reliance on past experiences. The major disadvantage of

ANN, compared with CBR, is that ANN is more like a black box in that only input and output are visual; the reasoning process cannot be retrieved. Neither explanation nor justification of any sort can be given by ANN. In addition, ANN deals only with numeric input data but in real situations there is much data that cannot be transferred to numerical data.

Table 5.3 Artificial Neural Network vs. Case-based Reasoning

Artificial Neural Network	Case-based reasoning system
requires expert sample	requires expert cases
requires the knowledge of the samples to include, and how to scale and represent them	Does not require any domain knowledge other than the cases to be included
can update and maintain themselves	may update and maintain themselves
inferences may be growing and changing	inferences are continually growing and changing
reason by abstracting and distributing the knowledge in a biologically plausible way	reasons from past experiences or cases, in a psychologically plausible way
cannot explain the reasoning process	can explain decisions citing cases
has no access to past cases, only the network created by them	has access and knowledge of past cases
extremely robust for incomplete and noisy data	very robust for incomplete and noisy data
Good for: difficult and intractable recognition, identification, or pattern manipulation tasks	**Good for:** poorly understood or highly dynamic domains where past decisions may form the basis for sound future decision making
Used in: recognition tasks such as financial monitoring, signal processing, or any task with highly codified data or lots of numerical information	**Used in:** diagnostic, design, or optimisation problems where good future solutions are likely to be composed of previously seen cases such as in medicine, law, complex machinery diagnosis, help desks, etc.

CBR vs. Machine Learning

Machine learning is a computer program which is said to learn from experience E, with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E (Mitchell, 1997) The field of machine learning studies the design of computer programs able to induce patterns, regularities, or rules from past experiences. Learner, a computer program, processes data D, representing past experiences, and tries to either develop an appropriate response to future data, or describe in some meaningful way the data seen.

General machine learning involves analyzing past cases to derive rules that apply to the set of cases. These rules may then be applied in solving new problems. Machine learning clearly separates the processes of learning rules and solving problems. Nevertheless, CBR uses induction algorithms to work out an index tree, which is used to match a new case against existing cases. As a result, the division between learning and problem solving is reduced.

Justification for a decision or answer is another distinction between Machine learning and CBR. Through machine learning, the reasoner will justify a decision by quoting the rules established by the training examples; any link or reference to individual training examples is completely lost. Conversely, CBR systems use the retrieved cases as precedents to support a decision (Watson, 1997).

5.2.2 The Advantage of CBR

CBR provides a number of advantages over alternative approaches, including:

- CBR doesn't require extensive analysis of domain knowledge. CBR permits problem solving even if domain knowledge is incomplete. The most important thing is to know how to compare two cases (Kolodner, 1993 and Watson, 1997).
- CBR allows shortcuts in reasoning. If a suitable case is found, a solution can be proposed quickly (Kolodner, 1993 and Watson, 1997).
- CBR can lead to an improved explanation capability in situations where the most comprehensible explanations are those that involve specific instances (Branting and Aha, 1995).

- CBR can help avoid past errors and learn from the errors and successes. In CBR, the system keeps a record of each situation that occurred for future reference, both positive and negative. These cases can warn of the potential for problems that have occurred in the past, and alert a reasoner to take actions to avoid repeating past mistakes.

- CBR allows a reasoner to propose solutions in domains that are not completely understood.

- CBR gives a reasoner a means of evaluating solutions when no algorithmic method is available for evaluation.

- Cases help a reasoner to focus its reasoning on important parts of a problem by pointing out the crucial features of a problem are.

- CBR can be used in highly complex, incompletely understood domains, as it creates a library of concrete experiences. The experience do not need to be completely understood, only recorded.

5.3 CBR TECHNIQUES

5.3.1 The CBR Cycle

The process involved in CBR can be represented by a schematic cycle (see Figure 5.4). Aamodt and Plaza (1994) have described CBR typically as a cyclical process comprising the four REs:

- RETRIEVE the most similar case after matching the current case with those in the library of past cases;

During this process, the case-based reasoner searches the database to find the case that most resembles the current situation.

- REUSE the retrieved case to try to solve the current problem;

This process includes using the retrieved case and adapting it to the new situation. At the end of this process, the reasoner might propose a solution.

- REVISE and adapt the proposed solution as required;

Since the proposed solution could be inadequate, this process can correct the first proposed solution.

- RETAIN the final solution as part of a new case.

This process enables CBR to learn and create a new solution and a new case that should be added to the case base.

It should be noted that the RETRIEVE process in CBR is different from the process in a database. To query data, the database only retrieves data using an exact match, while CBR can retrieve data using an approximate match.

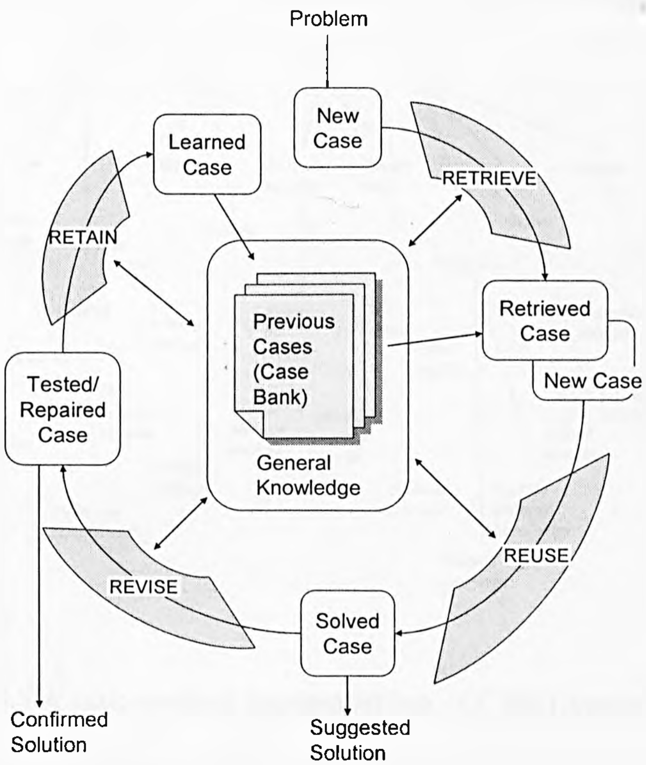

Figure 5.2 The CBR cycle (Aamodt and Plaza, 1994)

As shown in Figure 5.2, an initial description of a problem (top of figure) defines a new case. This new case is used to RETRIEVE a case from the collection of previous cases. The retrieved case is combined with the new case - through REUSE - into a solved case, i.e. a proposed solution to the initial problem. Through the REVISE process this solution is tested for success, e.g. by being applied to the real world environment or evaluated by a teacher, and repaired if failed. During RETAIN, useful experience is retained for future reuse, and the case base is updated by a new learned case, or by modification of some existing cases.

A new problem is matched against cases in the case base and one or more similar cases are retrieved. A solution suggested by the matching cases is then reused and tested for success. Unless the retrieved case is a close match the solution will probably have to be revised producing a new case that can be retained.

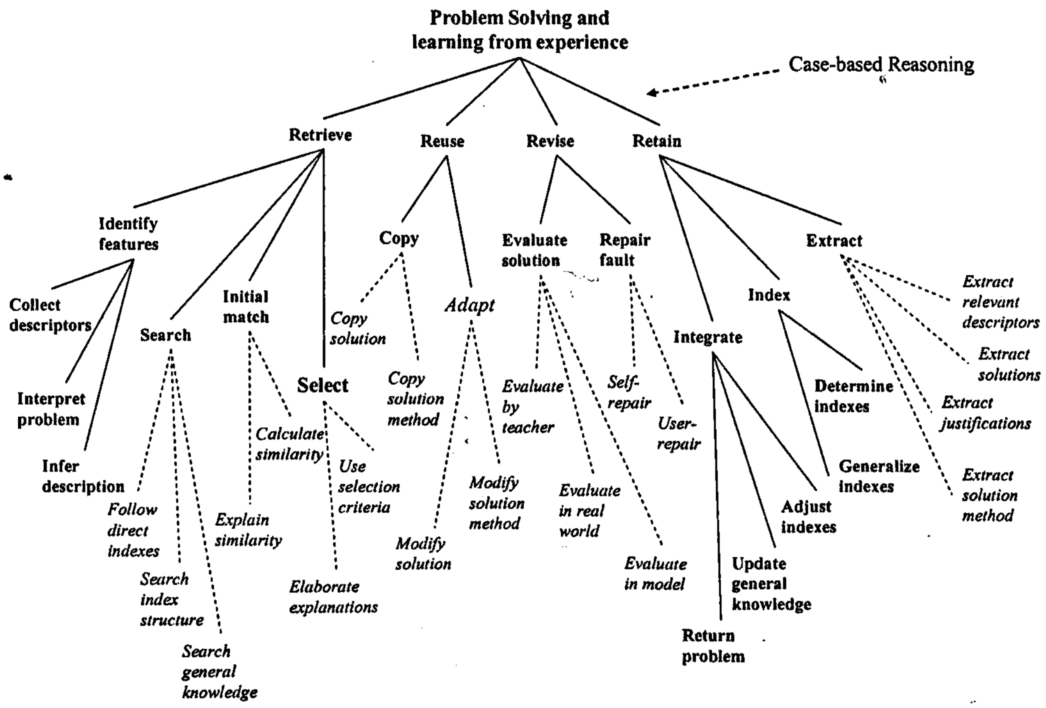

Figure 5.3 A task-method decomposition of CBR (Aadom, 1994)

Figure 5.3 provides a visual image of how each process in CBR works and the procedures to be followed.

5.3.2 Case Representation

Case representation includes the functionality and the ease of acquisition of the information (Kolodner, 1993). It must be expressive enough for users to accurately describe a problem or case.

A case is a contextualised piece of knowledge representing an experience. It contains the past lesson that is the content of the case and the context in which the lesson can be used (Alterman, 1989, David, 1991 and Kolodner, 1993). Typically a case comprises :

- the problem that describes the situation when the case occurred, which can be used to derive solutions to new problems, as in CASEY (Koton, 1989);
- the solution which states the derived solution to that problem, which can be used to evaluate new situations; and/or
- the outcome which describes the situation after the case has occurred, which can be used to evaluate the outcome of proposed solutions and prevent potential problems, as in MEDIATOR (Simpson, 1985).

Table 5.4 The content of the major parts of a case (Kolodner, 1993)

Major Parts	Contents
Problem Description	1. Goals to be achieved 2. Constraints on the goals 3. Features of the problem situation and the relationship between its parts
Solution	1. Solutions 2. Reasoning steps 3. The set justifications for decisions
Outcome	1. The outcome itself 2. Explanation of the expectation violation and/or failure 3. Repair strategy 4. Pointer to next attempt at solution

Table 5.4 lists the content of the above major parts of a case. The two major functional parts of a case are the lessons it teaches and the context in which it can teach (indices),

designating the circumstances in which it would be appropriately retrieved.

Another way to describe case presentations is to visualize the structure in terms of the problem space and the solution space (Watson, 1997 and Doyle, 1998); Figure 5.4 illustrates the structure. According to this structure, the description of a problem resides in the problem space. The retrieval process identifies the features of the case with the most identical problems. When the best match is found the system uses similarity matrices to find the best matching case. The solution of the case with the most similar problem may have to be adapted to solve the new problem.

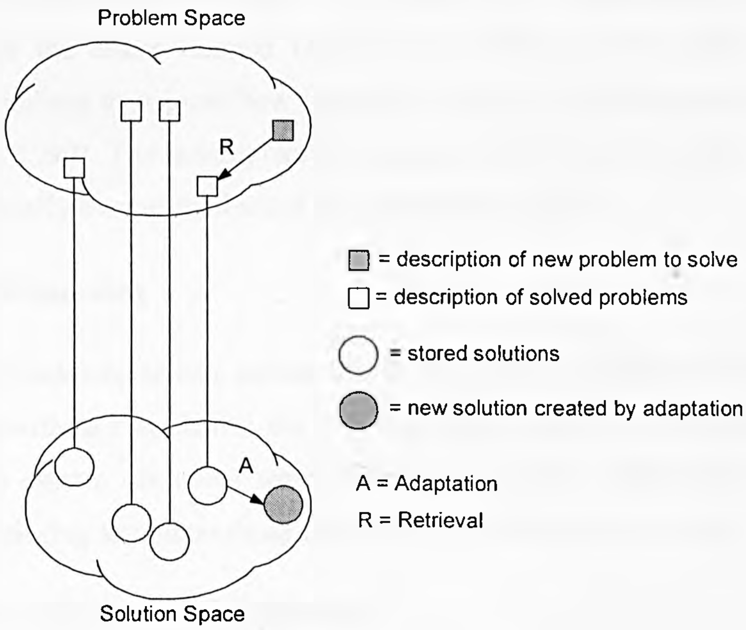

Figure 5.4 Problem and solution spaces (Watson, 1997 and Doyle, 1998)

5.3.3 Case Index

Case indexing involves assigning indices to cases to facilitate their retrieval. The aim of indexing is to make sure a case is accessed whenever appropriate. Indices may represent surface features or more abstract characteristics that describes its usefulness for a specific task. Indices should (Birnmaum and Collins, 1989 and Hammond, 1989):

- be predictive,

- address the purposes for which the case will be used;
- be concrete enough to be easily recognized in future; and
- be abstract enough to allow for widening the future use of the case-base and to retrieve a relevant case in a variety of future situations.

There is an ever increasing number of automated indexing methods including:

Checklist-based indexing

Checklist-base indexing indexes cases by features and dimensions that tend to be predictive across the entire domain (Acorn and Walden, 1992). These features are associated with indices to express how important they are when computing the matching scores (Aamodt, 1990). The indices can be assigned globally to all cases, or they can be assigned more locally over a small set of cases (Kolodner, 1993).

Difference-based indexing

Difference-based indexing selects indices that differentiate a case from other cases so that the retrieval algorithms can choose the best-matching cases from the case-base. During this process the system discovers which features of a case differentiate it from other similar cases, choosing as indices those features that differentiate cases best.

Similarity and explanation-based indexing

Similarity and explanation-based generalisation indexing produces an appropriate set of indices for abstract cases created from cases that share some common sets of features, whilst the unshared features are used as indices to the original cases (Hammond, 1987 and 1989). It is aimed at choosing indices appropriate to individual cases (Barletta and Mark, 1988).

Inductive learning indexing

Inductive learning methods (Lebowitz, 1987) identify predictive features that are then used as indices (Goodman, 1989). These techniques are widely used (e.g., in Cognitive

System's ReMind) and variants of the ID3 algorithm are used for rule induction.

However, despite the success of many automated methods, Kolodner (1993) believes that people tend to do better at choosing indices than algorithms, and therefore for practical applications indices should be chosen by hand.

5.3.4 Case Storage or Case Memory Models

Case storage is an important aspect in designing efficient CBR systems in that it should reflect the conceptual view of what is represented in the case and take into account the indices that characterise the case. The case-base should be organised into a manageable structure that supports efficient search and retrieval methods (Watson, 1997).

A balance has to be found between storing methods that preserve the semantic richness of cases and their indices and methods that simplify the access and retrieval of relevant cases. These methods are usually referred to as case memory models. The two most influential case memory models are the dynamic memory model of Schank and Kolodner, and the category-exemplar model of Porter and Bareiss.

In Schank and Kolodner's dynamic memory model, MOPs are either instances representing cases, events or objects, or abstractions representing generalised versions of instances or of other abstractions.

While in Porter and Bareiss' category-exemplar model, the case memory is a network structure of categories, semantic relations, cases and index pointers. Each case is associated with a category. Different case features are assigned different levels os importance in describing a case's membership to a category. Three types of indexes are provided, which may point to a case or a category:

- feature links that point from problem descriptions to a case or category;
- case links that point from categories to its associated cases; and
- difference links pointing from categories to the neighbouring cases that only differ in a small number of features.

5.3.5 Case Retrieval

Retrieval, given a description of a problem and using the indices in the case-memory, should retrieve the most similar cases to the current problem or situation. The retrieval algorithm relies on the indices and the organisation of the memory to direct the search to potentially useful cases. Template retrieval returns all cases that fit within certain parameters. Unlike database searches that target a specific value in a record, retrieval of cases from the case-base must be equipped with heuristics that perform partial matches, since in general there is no existing case that exactly matches the new case. The well known methods for case retrieval are: nearest neighbour retrieval, induction retrieval, knowledge guided retrieval and template retrieval. These methods can be used alone or combined into hybrid retrieval strategies.

Nearest Neighbour Match

Nearest neighbour involves the assessment of similarity between stored cases and the new input case, based on matching a weighted sum of features. Nearest neighbour matching finds the closest matches to a new case by calculating the similarity. The similarity of a target case to a source case for each attribute is determined. This measure may be multiplied by a weighting factor. Then the sum of the similarity of all attributes is calculated according to the Equation 5.1 as below (Watson, 1997):

$$Similarity(T,S) = \sum_{i=1}^{n} f(T_i, S_i) \times w_i \qquad \textit{Equation 5.1}$$

where T is the target case; S is the source case; n is the number of attributes in each case; i is an individual attribute from 1 to n; f is a similarity function for attribute i in cases T and S; and w is the importance weighting of attribute i.

A typical algorithm for calculating nearest neighbour matching is the one used by Cognitive Systems ReMind software as reported in Kolodner (1993).

$$\frac{\sum_{i=1}^{n} w \times sim(f_i^I, f_i^R)}{\sum_{i=1}^{n} w_i} \qquad \textit{Equation 5.2}$$

where w is the importance weighting of a feature (or slot), sim is the similarity function, and f_i^I and f_i^R are the values for feature *i* in the input and retrieved cases respectively.

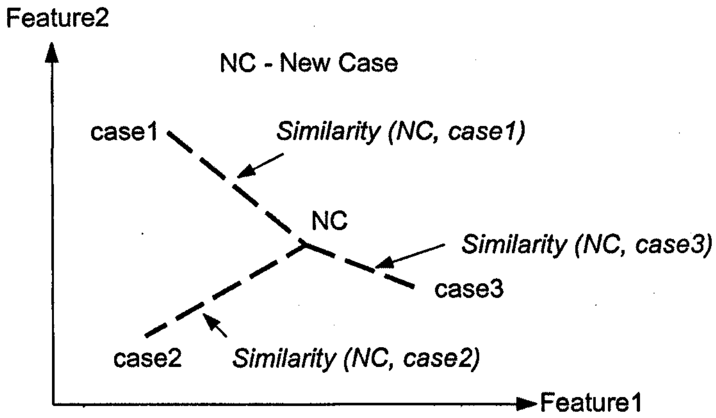

Figure 5.5 How to find the nearest neighbour of the new case

Figure 5.5 displays a simple scheme for nearest-neighbour matching. In this 2-dimensional space, *case3* is selected as the nearest neighbour because *similarity (NC, case3)> similarity (NC, case1) and similarity (NC, case3)> similarity (NC, case2).*

The biggest problem here is to determine the weights of the features. The limitation of this approach includes problems in converging on the correct solution and retrieval times. In general the use of this method leads to the retrieval time increasing linearly with the number of cases. Therefore this approach is more effective when the case base is relatively small (Watson and Marir, 1994). This approach is suitable if flexibility is required because one case-base can be used for several retrieval tasks (Barletta, 1991).

Inductive retrieval

Induction is a technique developed by machine learning researchers to extract rules or construct a decision tree from past data. Inductive retrieval algorithm is a technique that determines which features do the best job in discriminating cases and generates a decision tree type structure to organize the cases in memory (Watson, 1997). Inductive retrieval is performed by comparing the new problem against the decision tree defined for difference-based indexing. The features of the new case are propagated down the appropriate path of the identified decision tree until a leaf node in the tree is found. The leaf node indicates a group of similar cases to be retrieved. As the cases that meet the features of the new situation are accumulated, a group of similar cases can be easily retrieved by the system.

This approach is very useful when a single case feature is required as a solution, and when that case feature is dependent upon others. The most widely used induction algorithm in CBR is ID3. ID3 builds a decision tree from the database of cases. It uses a heuristic called 'information gain' to find the most promising attribute on which to divide the case.

Knowledge guide induction

Knowledge guide induction applies knowledge to the induction process by manually identifying case features that are known or thought to affect the primary case feature. Rules about the domain can be used to determine if the features of one case match the new situation. This approach is frequently used in conjunction with other techniques, because the explanatory knowledge is not always readily available for large case bases (Watson and Marir, 1994).

Template retrieval

Similar to SQL-like queries, template retrieval returns all cases that fit within certain parameters. This technique is often used before other techniques, such as nearest neighbour, to limit the search space to a relevant section of the case-base (Watson and Marir, 1994).

5.3.6 Case Adaptation

Once a matching case is retrieved a CBR system should adapt the solution stored in the retrieved case to meet the needs of the current case. Adaptation looks for prominent differences between the retrieved case and the current case and then applies formulae or rules that take those differences into account when suggesting a solution. An ideal set of adaptation rules must be strong enough to generate complete solutions from scratch, and an efficient CBR system may need both structural adaptation rules to adapt poorly understood solutions and derivational mechanisms to adapt solutions of cases that are well understood (Watson and Marir, 1994). In general, there are two kinds of adaptation in CBR: structural adaptation and derivational adaptation.

Structural adaptation

Structural adaptation applies rules or formulae directly to the solution stored in cases (Kolodner, 1993). Riesbeck and Schank (1989) have suggested three types of structural adaptation strategy namely null adaptation, parameterised adaptation and critic-based adaptation.

- Null adaptation is a direct, simple technique that applies whatever solution is retrieved to the current problem without adapting it. Null adaptation is useful for problems involving complex reasoning but with a simple solution.
- Parameter adjustment is a structural adaptation technique that compares specified parameters of the retrieved and current case to modify the solution in an appropriate direction.
- Critic-based adaptation is where a critic looks for combinations of features that can cause a problem in a solution. Importantly, the critic is aware of repairs for these problems.

Derivational adaptation

Derivational adaptation reuses the rules or formulae that generated the original solution to produce a new solution to the current problem. Unlike structural adaptation it operates not on the original solution, but on the method used to generate that solution (Maher et al, 1995). Derivational adaptation, sometimes referred to a re-instantiation, can only be used for cases that are well understood. Several techniques have been used in CBR for derivational adaptation. These include:

- Re-instantiation is used to instantiate features of an old solution with new features;
- Derivational replay is the process of using the method of deriving an old solution or solution piece to derive a solution in the new situation.

5.4 THE APPLICATION OF CBR

Having understood the concepts and processes that underlie CBR, this section introduces the application of CBR. CBR techniques have proven useful for implementing intelligent software components and CBR systems have been deployed successfully for many types of tasks: for electronic commerce (Watson, 1997, Stolpmann and Wess, 1998, Vollrath, 1998); for decision support applications such as helpdesks (Goker and Roth-Berghofer, 1999); for planning tasks, such as design and configuration (Navinchandra, 1988, Sycara and Navinchandra, 1989, Stroulia et al., 1992, Hennessy and Hinkle, 1992); and other classification tasks, such as diagnosis, prediction, and assessment (Koton, 1989, Bareiss, 1989). Furthermore, Watson (1997) listed more than 130 major companies using CBR for a range of tasks, including: hardware and software technology, finance and insurance, telecommunications, manufacturing and transportation, utilities, retail/consumer, outsourcing, and miscellaneous.

5.3.1 A Classification of Applications

The application of CBR is divided into two sub-sections: classification tasks and synthesis tasks. Figure 5.6 lays out a classification of CBR applications by Althoff (1995).

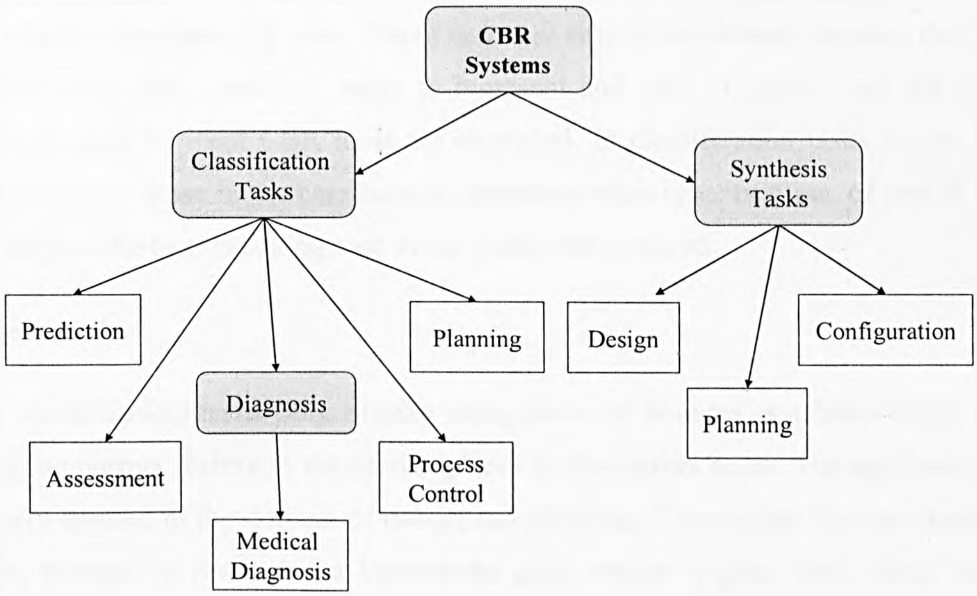

Figure 5.6 A classification hierarchy of CBR application after Althoff et al. (1995)

Classification tasks

Classification tasks cover a wide variety of applications which are listed in Table 5.5 below.

Table 5.5 Classification tasks: application area and examples

Application Area	Examples
Diagnosis	Medical diagnosis or development of the domain of clinical audiology
Prediction	The forecasting of finance market performance
Assessment	Risk analysis for banking or insurance or the estimation of project costs
Process control	The control of manufacturing equipment
Planning	The reuse of travel plans or work schedules

Classification tasks can be recognized by the need to match an object against others in a library from which an answer can be inferred. The outcome is usually an attribute of each case, which is structured by users. These tasks are easy to implement, because they match the CBR-cycle well; cases are easier to represent and easy to collect; and the retrieval algorithms used by most CBR tools are classified. In classification tasks a new case is matched against those in the case-base to determine what type, or class, of case it is. The solution from the best matching case in the class is then reused.

Synthesis tasks

While classification tasks only require recognition of features, synthesis tasks require placing the correct feature in the correct places in the correct order. The synthesis system is usually applied in the domain of design and planning. The system tries to simplify the creative process by producing a know-to-be-good design or plan from which the final design or plan can be generated. The assumption is that modifying a good initial design or plan will be easier than creating a new one from scratch. The synthesis tasks are inherently complex because of the constrains between elements used during synthesis. CBR systems performing synthesis tasks should use adaptation and combine with other techniques. The synthesis tasks involve can be:

- design the creation of a new artifact by adapting elements of previous ones;
- planning the creating of new plans from elements of old ones; and
- configuration the creation of new schedules from old schedules.

5.3.2 CBR Application in Construction

Examples in Civil Engineering

In the construction industry there has been a variety of applications of case-based reasoning. Case-based design has been the largest research field such as: architectural design (Faltings et al., 1991; Shih, 1991); office design (Pearce et al., 1992); integrated design and construction (Schmitt 1993); building design (Soibelman, 1999); industrial building design (Borner, 1995); structural design (Gero, 1990; Maher and Garza, 1997); and material selection (Dutton and Maun, 1997).

Furthermore, a group in Wales is applying CBR to the design of motorway bridges (Moore et al., 1994). Moore and Lehane (1999) detail a system which allows design data for small and medium-sized bridges to be sorted in a case-base for future retrieval. Other examples of CBR applications include:

- **KICS** (Yang and Robertson 1994) is being implemented in the domain of building regulations. This system accumulates case histories of the interpretation of regulations used to establish precedents. These precedents can be used to revise statutory regulations and to enrich the resulting new versions of regulations.
- **KOP**, Kit of Parts, is a case-based system for the design of medium-sized post offices, which is composed of pre-designed modules. Facility Service Center in Memphis, Tennessee developed KOP in 1986.

- **ARCHIE** (Kolodner, 1992) is a CBR aid for the conceptual design of building. It provides users with an effective tool with which to navigate through a maze of description of previous design cases.

- **CADCASE** (Maher and Balanchandran, 1994) is derived from CADSYS (Maher and Zhang, 1993). In this model information held in design cases is represented using multiple media: attribute value pairs, test strings, and drawings. CADCASE was found to be more flexible in case retrieval than CADSYS.

- **SEED** (Flemming, 1994) is aimed at supporting tasks in the early phases of building design. It provides systematic computational support for the rapid generation of design with respect to recurring building types.

Examples in Construction Management

CBR research in construction management covers a range of fields including: planning, estimating, prequalification, cost forecasting, procurement and others. Examples include strategic cost estimating for building refurbishment (Marir and Watson, 1995); housing renovation grant assessment (Brandon and Ribiero, 1997); the prediction of the outcome of construction litigation (Arditi and Tokdemir, 1999); and a variety of applications as listed below.

- **CasePlan** (Dzeng, 1995) is a CBR model that automates the generation of construction schedules for power plant boiler erection. This uses a generic product model to establish the basis for comparing projects with one another, and is achieved using a model to represent a project's design and relate it to the construction schedule.

- **EQUAL**, Expert QUALifier, (Ng, 1996), developed a prototype model for contractor pre-qualification. It used ReMind as the programming shell.

- **COMMIT** (Rezgui et al., 1997): this project investigated and proposed new forms of project information management based on recent developments in Information Technologies. The CBR system is particularly used for the construction process activity specification.

- **CBRMIND** (Morcous et al., 2002) is an implement for CBR in modelling infrastructure deterioration. CBRMIND contains special classes and algorithms for representing and matching time-dependent data.

- **CAMP,** Case Matching Prediction, (Yi, 2003), is a dynamic forecasting model for periodic expenditures of residential building projects. It enables the project manager to forecast monthly expenditure with less time and effort.

5.5 SELECTION OF MODELLING PACKAGE

In 1983, Janet Kolodner developed one of the first CBR systems, CYRUS, based on Schank's dynamic memory model (Schank, 1982). Since then, a number of CBR applications have become available in the field of industry and business. Many Case-Based Reasoning development tools are commercially available with many applications. Each has different functionality and will suit different requirements. Table 5.4 compares some of the popular CBR packages in the commercial area, and presents a summary of the key features of the major CBR shell.

Table 5.6 Comparison of CBR Packages

Package Name	Vendor	Neighbour matching	Text handling	Algorithm	C/C++ Library	Platform	Application	Further features
CBR Express	Inference Corporation	Mediate	Yes	CBR and RBS	No	Windows 95/98 and NT and Web Server	Help desks, Intelligent task assistant, knowledge publish, information access system	The most successful CBR package
Case Point	Inference Corporation	Very fast	No	CBR	No	Windows 95/98 and NT	Diagnostic, Classification Problems	Memory efficient, integrates with other applications
ART* Enterprise	Mindbox Company	Fast	Yes	CBR, MBR and RBS	Yes	Windows 95/98/2000,NT, OS/2, Solaris, Web Server	Mortgage organizer, financial services industry, small businesses, Client-server MIS applications	GUI builder, ability to link to data repositories in most proprietary DBMS formats for developing client-server applications
Eclipse-the Easy Reador	Haley Enterprise	Mediate	Yes	CBR and RBS	Yes	Windows 3.1, 95/98, NT and a range of UNIX	Help desks, customer support, trouble shooting, complex diagnosis, data mining and other CBR applications	Lacks the ability to integrate data from disparate heterogeneous databases
ReMind	Cognitive System Inc.	Good	Part	CBR	No	Macintosh, Windows 3.x, 95, NT, available as a C library for embedding	Trouble shooting, complex diagnosis, data mining and other CBR applications	The most flexible CBR tool, has the wildest range of case retrieval methods
SpotLight®	CaseBank Technologies Inc.	Mediate	No	CBR and KBS	No	Windows 95/98 and NT	Complex diagnosis and trouble shooting of machines	Solutions capture and manages knowledge through reasoning engine
ReCall	Isoft Company	Mediate	No	CBR	Yes	Windows 3.1/NT/95, SUN, IBM RISC System/6000, BULL DPX20, HP 9000 series 700, and DEC Alpha	Trouble shooting, complex diagnosis, data mining and other CBR applications	Graphical user interface, easy interface to external applications in databases, and integrates with other applications.
Esteem	Esteem Software Inc.	Mediate	No	CBR	Yes	Windows 3.1, 95/98 and NT	Help desks, customer support, trouble shooting, diagnosis, and other CBR applications	Supports various similarity assessment methods and can automatically generate feature weights
Casse Power	Inductive Solutions Inc.	Mediate	No	CBR	No	Windows 3.1, 95 and NT	CBR applied to data within Microsoft Excel spreadsheets	Builds its cases within the spreadsheet environment of Microsoft Excel
KATE	Acknosoft	Mediate	No	CBR	No	Windows 3.1, 95, NT	Help desks, customer support, trouble shooting, complex diagnosis, data mining and other CBR applications	Graphical editing facilities to build the case-index structure and to generate dialogues

Vafaie *et. al.* (2002) introduced a selection process for intelligent decision support tools for real-time system monitoring. Over 70 packages were initially investigated. Through expert evaluation, the reviewing of relevant websites and from personal contacts, eight packages were chosen for further consideration: NACoDAE, Kaidata Advisor, Knowledge Builder, ART*Enterprise, Easy Reasoner, Brokat Advisor, Exsys CORVID, and G2 Classic. A questionnaire with 25 criteria and 8 dimensions (as shown in Figure 5.7, step 1) was sent out, and in accordance with replies, a list of the three most recommended packages was drawn up.

Step1. Criteria selection

Evaluation Criteria	Criteria Weight	Dimension Weight
Data Connectivity		
Data Connectivity Methods	0.43	
Multiple Data Sources	0.14	
Export to a Database	0.43	
Total	1.00	
Functionality		0.24
Reasoning approaches	0.10	
Sources used for reasoning	0.10	
Types of searching	0.10	
Forms of Conflict Resolution	0.10	
Uncertain, Imprecise data	0.10	
Real time response	0.10	
Data retrieval Mechanism	0.10	
Tracing capabilities	0.10	
Operational Statistics	0.10	
Multiple scenario resolution	0.10	
Total	1.00	
Architecture		0.17
Hardware/Operating Systems	0.20	
Other required components	0.06	
Complexity/required training	0.27	
Scalability	0.20	
History log file	0.20	
Administration	0.06	
Total	1.00	
User presentation		0.11
Quality of the IDE	0.50	
Standard interface	0.00	
Quality of the API	0.50	
Total	1.00	
Technical Support		0.06
Company Profile		0.06
Installed Customer Base		0.06
Total		1.00

Step2.

Vendor	User Presentation	Data Connectivity	Architecture	Functionality	Technical support	Installed Base	Company Profile	Final Weighted Score
Weight	0.11	0.30	0.17	0.24	0.06	0.06	0.06	
Kaidata	3.5	3.7	2.9	3.2	4.0	3.0	3.0	**3.3**
Empolis	2.0	2.7	2.2	1.9	3.0	2.0	3.0	**2.1**
MindBox	3.5	3.7	2.8	3.4	4.0	3.0	4.0	**3.3**
Borkat	2.0	3.1	1.9	2.0	3.0	2.0	3.0	**2.4**
Gensym	4.0	3.7	2.8	3.0	3.0	4.0	3.0	**3.3**

Step3.

Criteria Score	Kaidara Advisor (Kaidara)	ART*Enterprise (MindBox)	G2 (Gensym)
Basic Required Functionality	3	3	3
Capacity for add-on Functionality	1	2	3
Compatibility of Reasoning Approach with Domain Expertise	2	3	1
Impact of the Solution on the Organization	3	3	1
Final Weighted Score	2.25	2.75	2

Figure 5.7 Selections of CBR tools

The three packages incorporated a rule-based approach, or a case-based approach, or a combination of the two. Kaidata Advisor uses a pure case-based reasoning approach, G2 uses on a pure rule-base approach and ART*Enterprise is considered a hybrid tool because it contains both rules and case based approaches. By comparing the different characteristics when selecting rule- and case-based reasoning software, a final list of criteria was developed (as shown in Figure 5.7, step 2).

With the above analysis, ART*Enterprise is outstanding with better overall functionality and flexibility than the other packages. As a result Mindbox's ART*Enterprise was chosen as the tool for this research.

5.6 ART*ENTERPRISE

5.6.1 The History of ART*Enterprise

ART*Enterprise is an integrated knowledge-based applications' development tool that supports rule-based, case-based, object-oriented and procedural representation and reasoning of domain knowledge. This package is originally a product of Inference Corporation, one of the oldest established vendors of AI tools. It was developed from the family of rule-based development environments originating with ART in the mid-1980s. At that time ART and then ART-IM were marketed as AI development tools. Inference have dropped the AI label and market ART*Enterprise as an integrated, object-oriented applications' development tool designed for Management Information Systems' (MIS) developers. In 1995 the division responsible for this package split from the Inference Corporation and established Brightware Inc. ART*Enterprise 3.0 onwards is the product of Brightware. On 31st March 2000 a division of Brightware launched MindBox, and then released the latest version of ART*Enterprise 10. Therefore, documents comparing CBR packages, usually provide three different citations for the vendor of ART*Enterprise.

ART*Enterprise offers a variety of representational paradigms including: objects supporting multiple inheritance, encapsulation and polymorphism, and rules and cases. This is all packaged with a GUI builder, version control facilities, and the ability to link to data repositories in most proprietary DBMS formats for developing client-server applications. Moreover, ART*Enterprise offers cross-platform support for most operating systems, windowing systems and hardware platforms. The CBR component in ART*Enterprise is essentially the same as that in CBR-Express (or rather vice-versa since CBR-Express uses code from ART to provide its CBR functionality). Thus, ART*Enterprise offers nearest neighbour matching and impressive text handling ability. The CBR functionality of ART*Enterprise is more controllable than in CBR-Express. Moreover, the integration with other knowledge representational paradigms means that this offers an excellent environment to integrate CBR with other techniques and to use MBR techniques for case adaptation.

5.6.2 The Application of ART*Enterprise

Mindbox's ART*Enterprise has already used successfully in a range of applications. Ford Motor Credit has used the ART family of products since 1993. In Ocwen, ART*Enterprise was used to deploy a system that would work transparently with REALServicing™, Ocwen's state-of-the-art mortgage servicing system (CRE). This system was used to improve responses to customer inquiries. In 2001, the GMAC Mortgage Corporation implemented ART*Enterprise to provide intelligent automation of the mortgage lending process and to build the company's end-to-end automated underwriting system.

Recent examples of ART*Enterprise as applied to academic research include:

MUKCA, Multi-User Knowledge Capture Administrator, (Fagan and Corley, 1998) is a tool incorporating CBR for supporting the reuse of past SQL queries to solve real problems in BT.

CALIBRE, Candidate LIBrary Retrieval, (Fagan and Heath, 1999) combines CBR and data mining as an innovative tool to support the extraction of candidate lists for BT's targeted marketing campaigns.

CBRidge Planner (Tah and Carr, 1998 and Tah et al., 1999) is a CBR application used in the planning of highway bridge construction. This system links with an external planning package Microsoft Project, which enables task data, durations, network diagram, resources and costs to be exported and viewed graphically.

CPAS, Case-based Procurement Advisory System for construction, (Luu et al., 2003) was developed for construction procurement selection and suggestions for procurement methods. This system, powered by a fuzzy similarity retrieval mechanism, gives a greater accuracy than normal similarity retrieval processes. It also combines both critic-based adaptation and parameterized adaptation as its adaptation strategy.

5.6.3 Features of ART*Enterprise

ART*Enterprise offers cross-platform support for most operating systems, windowing systems, and hardware platforms (Watson, 1997). It integrates a broad range of database connections, such as Oracle, Sybase, Informix, Q+E, and DB2, MS Excel, Access etc. ART*Enterprise can handle both structured and unstructured data. Users access data (documents, etc.) without knowing a file name or location, but naming simply the type of information they require.

This is all packaged with a GUI builder, version control facilities and an ability to link to data repositories in most proprietary DBMS formats for developing client-server applications. Moreover, ART*Enterprise offers cross-platform support for most operating systems, windowing systems and hardware platforms. It allows seamless integration with industrial standard programming languages like C/C++ and offers CORBA and Web features that allow the rule engine to communicate with components written in Java or any other language. The ART*Enterprise environment provides a forward chaining engine where backward chaining can be implemented, though it is not supported directly. ART*Enterprise also provides a CBR kernel for those who are interested in incorporating it into their applications. In addition, ART*Enterprise's object-oriented architecture streamlines software development, reduces maintenance and upgrade costs, and maximizes legacy hardware and software investment (Steelhammer and Maxxoni, 2002).

Primary features that ART*Enterprise provides are listed below:

- **Object-Oriented Programming:** Provides a full featured multiple inheritance object model for clear representation of business objects and their associated methods. Objects provide a central mechanism for integration of data from user interfaces, data bases, rule processing, and case-based reasoning.

- **Rules:** Captures formal policies and heuristic logic in modular sets of criteria and action pairs. Rules respond automatically to changes in the state of the object model or associated facts.

- **Case-based Reasoning:** Stores a "case-base" of previously solved problems, and solves new problems by searching for similar past cases and adapting their solution to the current situation.

- **Procedures:** Includes hundreds of built-in functions to perform computations, manipulate sequences of items and strings, and provide bases for building sophisticated user defined functions to automate application actions.

- **Hypothetical Reasoning:** Supports "what if" analysis and maintains knowledgebase consistency.

- **Database Integration Tools:** Provide point and click integration of multiple database sources into ART*Enterprise objects. Can automatically generate SQL and database table definitions.

- **Document Integration Tools**: Provide viewer classes for text documents. CBR can be used to retrieve documents using a combination of text and database attributes. Support importing CDF document-base index files.

- **User Interface Tools**: A flexible object-oriented graphics library provides native look and feel by wrapping around native window system calls. Fully portable across Motif and Microsoft Windows. Screen Painter provides point and click GUI design with productivity tools such as menu editors, and forms creation and layout wizards.

- **System Integration Tools:** Developers can easily integrate ART*Enterprise applications into existing systems to add intelligence to strategic applications. In addition to support for peer-to-peer communication using TCP/IP sockets, ART*Enterprise delivers an embeddable architecture with an open API that supports both dynamic linking and static linking with C++ call-in and call-out capabilities. The ART*Enterprise web add-on library provides World Wide Web integration.

- **Development / Debugging Tools:** The highly integrated development environment allows rapid prototyping and fully incremental testing. It includes a full set of debugging and logic tracing tools that dramatically increases productivity during development and the maintenance of intelligent applications.

5.7 SUMMARY

Case-based reasoning mimics the sense of human reasoning. CBR looks for previous cases that are similar to the current problem and reuses them to solve the problem. Compared with other information techniques such as statistical analysis, rule-based reasoning, model-based reasoning, artificial neural networks, and machine learning, case-based reasoning is particularly suitable for risk identification. This is because CBR is good for poorly understood or highly dynamic domains where past decisions may be used for future decision making. A typical CBR cycle has four processes: retrieve, reuse, revise and retain. The main aspects of the CBR model include: case representation, index, storage, case retrieval, and case adaptation.

CBR has been used successfully in solving a wide range of problems, including diagnostic, prediction, assessment, process control, planning, design and configuration. In civil engineering applications, CBR has achieved particularly in design. Furthermore, a variety of applications in construction management has been attempted, for example: planning, estimating, prequalification, procurement, cost forecasting and litigation. Having compared a range of CBR shells, ART*Enterprise is selected as the tool for CBRisk development because, it stands out with its overall functionality and flexibility compared to other packages. ART*Enterprise integrates object-oriental programming, case-based reasoning and rule-based reasoning. It has been successfully implemented in a range of commercial applications as well as in academic research, and is likely to be compatible with the applications of this research.

CHAPTER 6

PROTOTYPE FRAMEWORK

6.0 INTRODUCTION

After formulating the current problem of risk identification through a literature review detailed in Chapters 2 and 3, and having chosen an appropriate technique and selected a suitable system development tool in Chapter 5, from this chapter onwards, the thesis introduces the development of a risk identification model utilizing case-based reasoning. This chapter demonstrates the conceptual framework of the risk identification model. Firstly, the concept design dictates the requirements for the model, the preliminary framework of the model is constructed, and the programming paradigms used for definite applications are described. The later sections of this chapter discuss each module that within the framework of the model. These include: criteria formulation, knowledge base, graphic user interface, CBR application, and data mapping.

6.1 CONCEPTUAL DESIGN

It has been asserted that formulating a conceptual framework is useful for organising ideas and suggesting action (Scott-Morton, 1984). Designing a case based intelligent model for risk identification may involve a number of alternative ideas and actions, which include determining the data used in a particular risk identification process and decision structure. Devising a suitable conceptual framework can lead to a more effective and efficient design for the system.

6.1.1 Requirements Specification

To design an IT based application model, the first stage is to specify its requirements. A requirement specification abstracts descriptions of the services which the system should provide and the constraints under which the system must operate (Sommerville, 1996).

A list of requirements was formulated (as listed below) to guide the development of the prototype model and to describe the necessary characteristics that should be manifested in the design. These requirements were identified based on three main objectives: firstly, eliminating the short comings of current risk identification methods; secondly, by accommodating the complexity, evolving nature and subjectivity of the risk identification domain; and thirdly, by extracting as much knowledge as possible from similar projects, which saves significant time and effort and increases the efficiency in carrying out risk identification.

These requirements are:

- Representation of project component interaction: Construction projects are complex as they are made up of many components and subcomponents. In the risk identification domain, the condition of some components significantly affects the risk level. Therefore the prototype model must provide a way to represent the interaction among different deterioration mechanisms of project components in order to account for its contribution in identifying risks.

- Versatility and extensibility in case and knowledge representation: It is impossible to develop a system that satisfies every need for every user at the outset, but it is possible and recommended to develop a system that satisfies the current basic needs and allows future extensions to take place (Ricard and Fenves, 2000). Hence the design of the risk identification model must be flexible enough to allow: (1) the representation of cases with versatile structure and contents; (2) the extending of case descriptions both in structure and contents, (3) the customizing of retrieval knowledge, and (4) the updating and augmenting of the adaptation knowledge.

- Fuzziness of retrieval knowledge: Retrieval knowledge consists mainly of attribute weights and their degrees of similarity. This knowledge is usually obtained from domain experts who assess the relative importance of attributes and the relative similarity among attribute values based on their own experience, intuition, and judgment (Kolodner, 1993). Thus, the prototype model must be able to support the representation of retrieval knowledge in a way that accounts for the subjectivity and imprecision of human judgment.

- Data reusing and sharing: The prototype model should enable the user to organize cases and retrieve them for reuse by modelling according to the CBR process, and enable the user to capture and interpret the relationship between the work-packages in each case, in a computable and reusable form so that the reuse process can be automated.

- Friendly user interface: A friendly user interface is the medium for communication with the user. A good user interface can attract and keep the user's attention. A goal driven interface is easy for users to understand, follow and then carry out the application. It should not only be friendly in appearance, but follow logical processes and interface transactions.

The requirement specification provides the concept content and functionality that the model should achieve. Based on these requirements the components needed to construct the model can be then defined and developed.

6.1.2 Conceptual Framework

ART*Enterprise has been chosen as the tool to develop the initial intelligent risk identification model. An ART*Enterprise application typically has three layers:

- A database mapping

- A graphic user interface (GUI)

- An application model

These layers emphasize one of the distinguishing features of ART*Enterprise: all three layers represent entry points into application development.

Within the application of a risk identification model, the key role of the knowledge developer is to create knowledge management frameworks and intelligent application generation tools for use by project experts in creating the end-user applications. The ART*Enterprise dynamic language lets project experts update the project model, the rules that monitor this model, and screen presentations as the application is running, allowing instantaneous feedback on those changes.

According to the structure above, Figure 6.1 describes components made up of the prototype model and the relationships and sequences among them. It consists of five modules: a criteria formulation module, a knowledge base module, a graphic user interface module, a data mapping module, and a CBR application module.

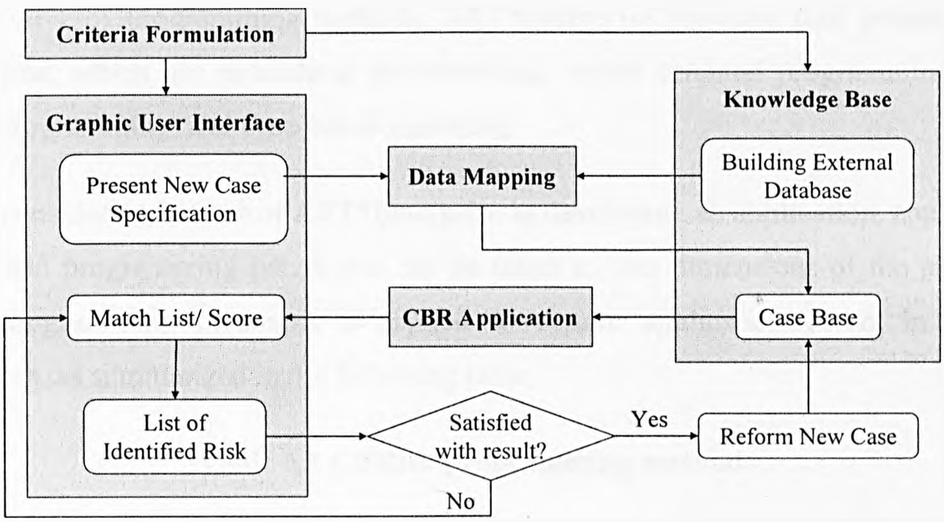

Figure 6.1 Preliminary framework of CBRisk

Criteria formulation specifies the main features that should be considered in the clarification of a project and the method of picking up the information relevant to project risks from the textual description. Later on the chosen criteria were used to build the structure of an external database. The information input into this database was imported into the ART*Enterprise application and then loaded to the case base through data mapping. Data Integrator (DI) connects both the GUI objects and external database objects with case objects in a coordinate way, in order to transfer data among them. The data captured both from DI and GUI was saved in the case base. On the other hand, the confirmed features were also used to construct the Graphic User Interface (GUI). The user input the project information through GUI and presented it as the new problem. CBR application then retrieved the case base and found the most similar cases. These cases were listed in a match list along with their match scores respectively. Each matched case was accompanied by a relevant identified risk list to which the user could refer. Apparently the adaptation process within the CBR application can also formulate a recommended identified risk list back to the user interface. This list is then reviewed by the user. If it is deemed to be satisfactory, the risk list along with the case description will be stored in the case base. Otherwise it will return to the risk list of matched cases. Meanwhile proposes an appropriate risk list for the project.

6.1.3 Programming Paradigms

In the view of programming methods, ART*Enterprise provides four programming paradigms, which are procedural programming, object oriented programming, rule-based programming, and case-based reasoning.

When considering the use of ART*Enterprise in developing an application, application layers and programming paradigms can be taken as two dimensions of the problem. Each programming paradigm is applied to certain application layers in certain situations, as summarized in the following table:

Table 6.1 CBRisk programming methods

Programming	Procedural	Object Oriented	Rule-based	Case-based
Graphic User Interface	×	Event-driven interface	Mixed initiative interface	
Application Model	×	Object model	CBR adaptation rules	Information retrieval
Database Mapping	×	×		

The Graphic User Interface component is defined as objects based on the project data. Theses objects will be connected to the case base and external database through data mapping. Applications are defined as objects for collecting the individual objects from GUI and also through data mapping. They are then transferred to the present case format to carry on the CBR application. Rule-based programming contributes to main functions: (1) to build up the transaction of user interfaces according to the different application demands, and (2) to carry out a systematic CBR adaptation process, which will be explained in detail in Chapter 8.

6.2 CRITERIA FORMULATION

In risk identification project managers are concerned with analyzing the risks that impact a decision. There are many risks which potentially affect the progress and benefit of a project. Those risks can be identified by scoring each of the contributing features, and then combining those scores in various ways to arrive at a case matching and so a suggested risk list for the project.

CBR systems vary in the way the case database is structured. The representation can be flat, where all cases are represented at the same level, or it can be hierarchical, expressing relationships between cases and sub-cases. The hierarchical organization is useful when the CBR system is used for taxonomic tasks.

Criteria formulation decides what features a case should have. This module aims to determine a set of decision criteria that best reflect the organization, project and risk identification objectives. To answer this question a range of situations should be considered. Firstly, the attributes to describe the project are typical enough to express the project's features, both common and diversified. Secondly, the number of criteria should be more than enough to carry out the CBR retrieval. Thirdly, the criteria should be regular enough to be picked up from most of the project documents. Finally, the module must be flexible enough to cover a range of normal situations even if similar cases can not be found in the case base.

To design the case structure, the case's relevant features must be identified. This module decides the case features including the category and scope of work for which an existing risk identification model was used. Three kinds of features exist. The first feature is that descriptive features help identify and describe the case and constitute the problem part of the case. The second set of features represents the solution. The solution part of the case was represented by the criteria and criteria adopted in the risk identification model. The third one affects the adjustments. A difference in the value of one of these features should lead to a difference, even if slight, in the solutions. A useful way to discover theses adjustment features is to have experts compare and contrast different cases. The features they describe as contributing to differences in answers are the adjustment features. These adjustment features will be used for similarity retrieval and adjustment of the retrieved case.

6.3 KNOWLEDGE BASE

Knowledge base is defined as helping the user of the application select the right resources in terms of decision models, reports, databases and knowledge bases (Klein and Methlie, 1995). A methodological knowledge base can support the decision analysis methodology and the model building process. The knowledge base provides the space to store facts collected from the project documents and the solutions that are inferred from these facts.

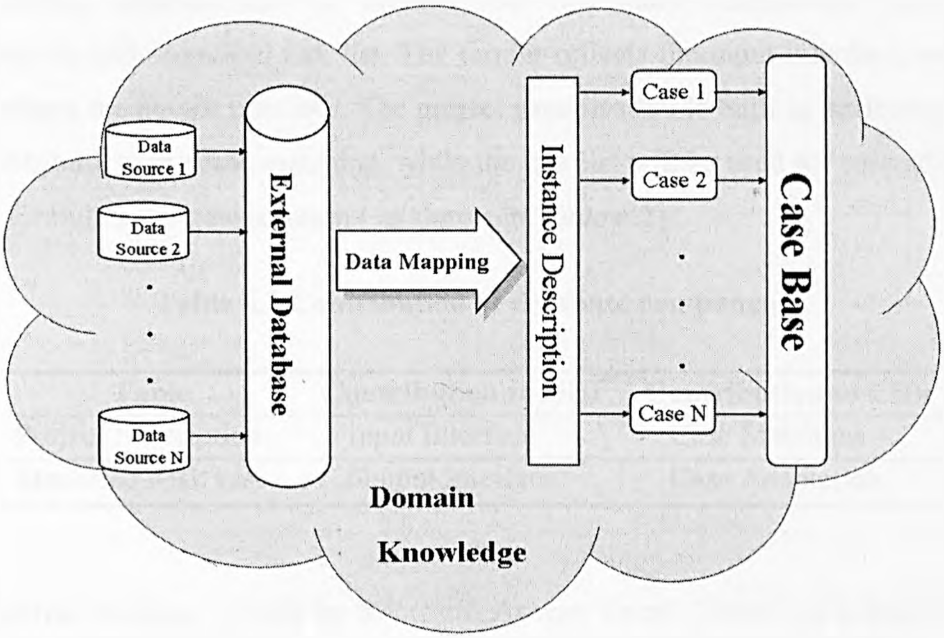

Figure 6.2 A flat structure of knowledge base

As shown in Figure 6.2, the knowledge base contains two main components: an external database and a case base. The individual project document is analysed and input into a structured external database. This data set was imported into the ART*Enterprise application domain and transferred to the relevant individual case through data mapping. These cases are then loaded into the case base to contribute to case matching and retrieval.

6.3.1 External Database

Once the set of features that define a case has been decided, the external database can be built up based on the set of a project's features. The database is organized into tables. Each table has a record for every project. Each record has fields to propose different features of the project. These fields may be of different types, depending on the kind of information they hold.

The entry of the database is the easiest step in CBR application development. Much of the data may reside in project documents, which must be translated into the appropriate format. During entry of project information care should be taken not only to avoid introducing new data errors but also to find errors in the existing data.

The external database can be divided into two main components: the project specification and identified risk list. The former reflects the input interface, while the later reflects the output interface. The project specification is built as attributes, which will contribute to the case matching, while the risk list will be used for case adaptation and to formulate the new solution (as shown in Table 6.2).

Table 6.2 Contribution of database components

Table	Contribution to GUI	Contribution to CBR
Project Description	Input Interface	Case Matching
Identified Risk List	Output Interface	Case Adaptation

The external database is built by Microsoft Access/ Excel. There is a linkage table to connect project descriptions and the identified risk list. In this manner, when match list is provided, the user can access an individual risk list directly. The relationship of the external database is shown in Figure 6.3.

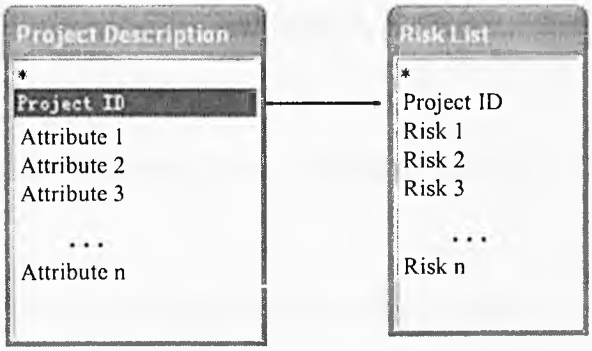

Figure 6.3 The relationship of external database

However, information contained in an external database cannot be used as knowledge directly in a CBR application. It should be transferred in a certain format for case description and saved to the case base.

6.3.2 Case Base

Case base is a special sort of database in which individual or multiple cases are treated as units for query. Case base as the memory is organized into an appropriate structure in order for its retrieval process to find appropriate cases. The organization of case memory is concerned with the study of organizing cases into a structure in order to support an efficient retrieval of cases (Maher et al., 1995).

The organization in case memory provides a means for accessing individual cases when searching for a relevant case. The issue of case memory organization is closely linked with case indexing schemes, retrieval strategies and the update of case memory. A good case memory organization might facilitate the efficiency and accuracy of retrieval. Retrieval of relevant cases is performed by a certain algorithm that searches the organizational structure of case memory for the right place for relevant cases to reside. Thus, the organizational structure provides platforms for how the retrieval process proceeds and how a new case is inserted in the case memory. Therefore, along with the discussion of organizational structures of case memory, the methods of building the structure are also taken into account (Maher et al., 1995).

Based on the features of construction projects and the risks identified, it is necessary to construct appropriate objects to map the available information. To construct the case base, the following aspects should be considered:

- risks are divided into categories and factors. A class is required for each of these concepts;
- the factors would be scored using symbolic scores. A class is required for symbolic scores;
- the categories, factors, and scores naturally form a tree;
- relations are required between classes to form class instances in a tree structure; and
- finally, since the manner in which to assess the overall risk for a given client must propagate the risk matching scores up the tree, attributes are required so classes can propagate these scores.

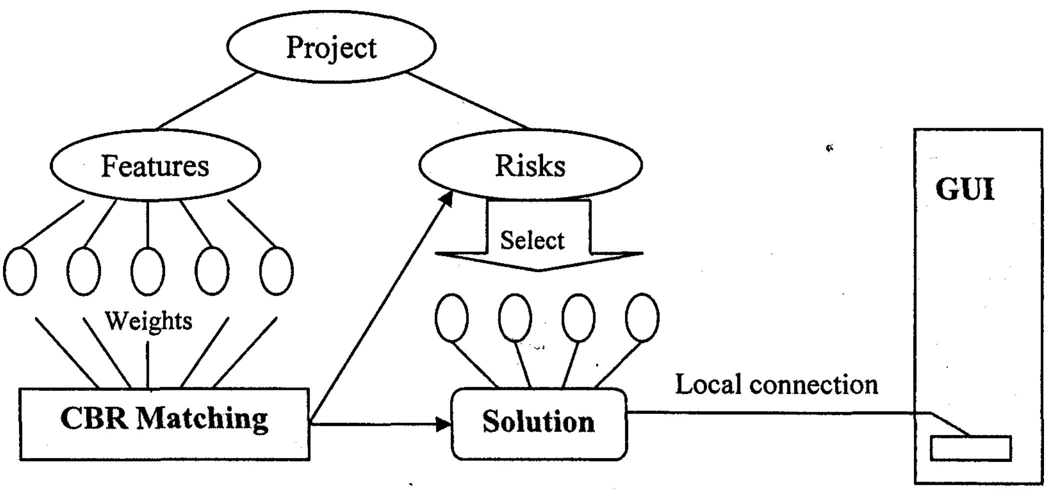

Figure 6.4 The structure of case base for prototype model

Figure 6.4 demonstrates the hierarchy data set of the case base. When objects involved in this system have been decided, the database can be established according to this hierarchy. Ellipses and circles represent objects, and rectangles represent methods. The risk hierarchy is constructed out of a number of features and weight objects. Each of these has a CBR matching method which, in its simplest form, sums up the matching the associated object. In this way, matching is carried out and a solution is found out from the other part of the project data: identified risks. When a user input the features for a given project in the user interface, that relevant weight is automatically transferred from the user interface objects to the object model objects via a location connection. It then provides user a list of identification risks for that project with the overall risk level.

6.4 GRAPHIC USER INTERFACE

Graphic user interfaces intend to improve data presentation and user interaction. Beyond the general description, in order to be usable by project experts it necessarily takes different forms, depending on the project's characteristics. To make the concept clearer, firstly design a general interface and a set of forms that will allow the project expert to enter and maintain the specific risk factors for each project.

ART*Enterprise offers the capability to dynamically generate and lay out GUI screen elements at runtime based on an underlying model. This feature provides generators, which create end-user application screens directly from the project model.

The Graphic user interface (GUI) is developed through Screen Painter (SP) and Graphic Toolkit (GTK). Screen Painter provides a point-and-click interface for accessing all the power of the graphic Toolkit. It allows you to design and test graphical user interfaces directly on-screen, instead of writing the ART*Enterprise programming code. While using Graphics Toolkit, the set of ART*Enterprise objects and procedures are used for constructing the graphic user interfaces.

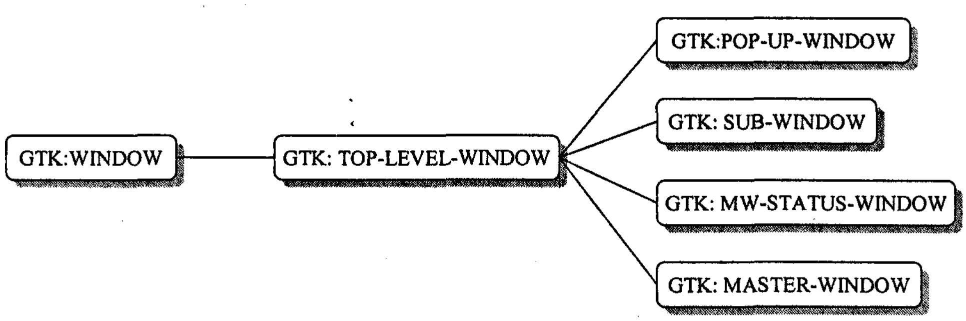

Figure 6.5 Window classes supported by the GTK

Figure 6.5 illustrates the hierarchy of the window classes supported by the GTK. To build up the model interface, GTK is basically used for building up the GUI frame, then Screen Paint takes the role of building up the draft interface, finally it goes back to GTK for application of some advanced graphic coding to construct a more friendly and perfect interface. The user interface is generally divided into two parts: the input and the output, which is fully explained in Chapter 8.

6.5 CBR APPLICATION

The CBR application is the core of the prototype model. Figure 6.6 demonstrates how the CBR system works in project risk identification. Information from previous projects is processed and stored as cases in a knowledge base. When a new project is initiated, it is first entered into the project description form as the cases stored in the knowledge base. It then searches the knowledge base to find similar previous projects, reuses the risks identified, and revises these risk lists to come up with suggested risks for the project in question. This preliminary suggestion list is reviewed by human experts, and then a new final report of identified risks for this project is generated. This solution is associated with the project description and is retained in the knowledge base as a piece of new knowledge for future use. The final process is called "refine" and is more about maintenance of the knowledge base and updating the information, such as when the project will finish, so that the identified risks and response strategy can be evaluated and the knowledge base revised.

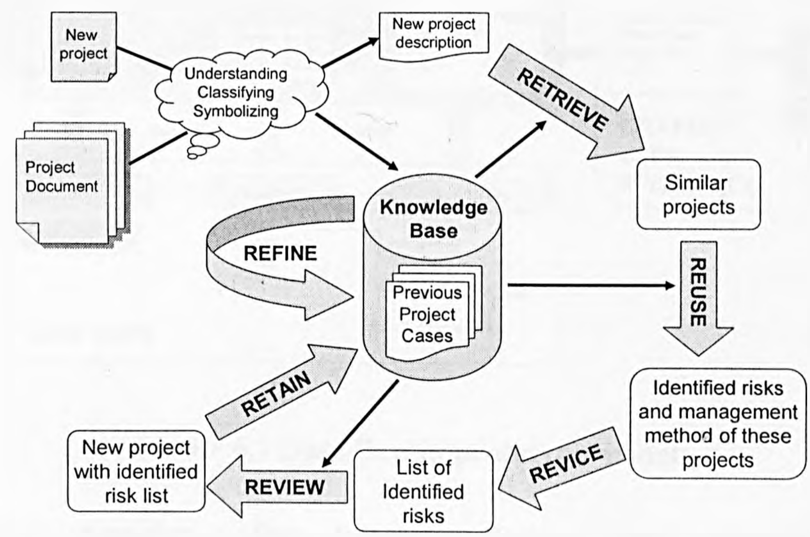

Figure 6.6 Project risk management CBR procedure

Basically, in the CBR system, project is described by attributes. The result of the retrieving process relies on past cases with attributes that match the current case. The matching scheme depends on varying degree of importance as indicated by attributes of different matching weights. A search engine then searches through all the cases in the case base and retrieves those that are a close match. As new cases are added, the CBR becomes increasingly powerful and accurate. The detailed design process of the CBR application will be proposed in Chapter 8.

6.6 DATA MAPPING

Data mapping is an important module to construct the prototype model. Relational database and case memory is the gateway to ease data accessibility and flow.

Figure 6.7 illustrates the relationship of the external Access database, user interface and case base in terms of data transaction, the internal relationship among themselves, and the data flow within the prototype model. Within this figure, plain lines indicate one-to-one relations and lines terminated with a dot represent one-to-many relations, with the dot on the many side.

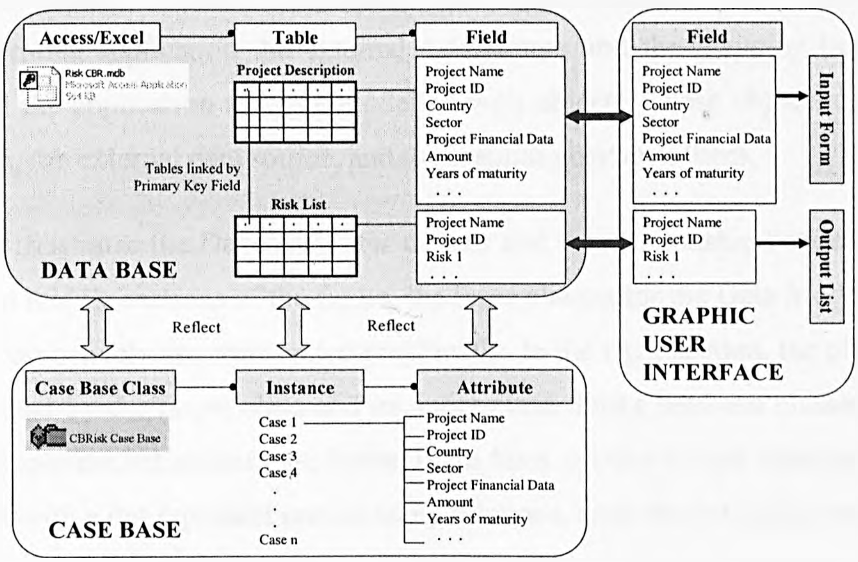

Figure 6.7 Data flow in prototype model

ART*Enterprise provides a Data Integrator (DI) to allow the user to create ART*Enterprise objects to represent data in external database and use those objects to transfer data to and from the external database.

6.6.1 Data Integrator

In side ART*Enterprise, data is organized into objects. There are two kinds of objects, classes and instances. Classes contain structural information and their instances store data. The instances of a class are similar to the records of a table. Each attribute holds a piece of information about the instance. Just as records and a table form a database, all the existing instances and classes together from the object base.

Creating an interface between an ART*Enterprise application and an external data source, using the Data Integrator, consists of two basic steps:

- Creation of an association between the ART*Enterprise application objects and the external data. This association is called a mapping;
- Use of the mapping created in step 1 to transfer data either from the external data source to ART*Enterprise or vice versa.

6.6.2 Data Integration Method

In Data Integrator, in order to transfer data from external data sources to the ART*Enterprise application the external data source and the mapping between that source and the application must be modelled with objects. These objects describe the application, the external data source, and the mapping between them.

Figure 6.8 illustrates the Data Integrator classes and the relationship between them. In the left and middle sections of the figure, the three classes for the Data Model and Data Mapping respectively are represented graphically. In the right section, the object model is represented by the target class and its super-class. Links between classes represent relations implemented as class attributes. Plain lines are one-to-one relations and lines terminated with a dot represent one-to-many relations, with the dot on the many side.

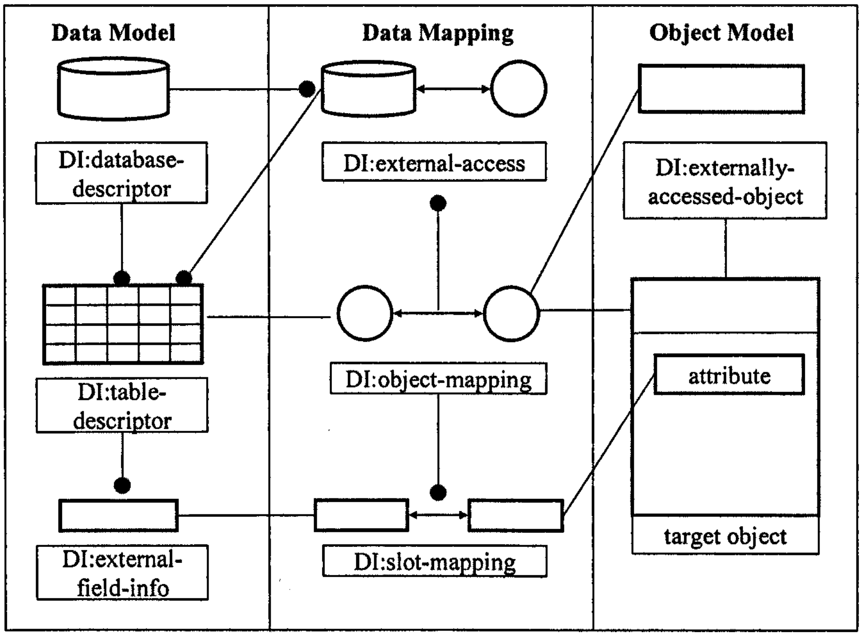

Figure 6.8 The Data Integrator classes and the relationships (Brightware, 1999)

As shown in Figure 6.8, creating a mapping between ART*Enterprise objects and external project data involves three kinds of objects: data model, data mapping and object model. Data model express the external database, data mapping states ART*Enterprise objects, and object model joins them together.

6.7 SUMMARY

Concept design is the first step in creating a computing system. This chapter demonstrates the framework used to construct the prototype model of risk identification. Based on the requirements' specification set up at the beginning, a concept framework was developed, which included five functionality modules. (1) Criteria formulation processes information from raw project documents, designs the frame for the data set, and builds up appropriate rules. (2) Graphic user interface constructs the interface, allowing communication between user and the model. (3) Knowledge base provides space in which to hold the domain information in order to support the model operation. (4) CBR application carries out case-based reasoning to find similar projects and provide primary risk identification. (5) Finally, Data mapping creates the objects link for the case base, user interface, and external data source.

CHAPTER 7
DATA CAPTURE

7.0 INTRODUCTION

This chapter outlines how data collection and interpretation is achieved. Firstly, the selection of the available data source is explored and the features of the data set are highlighted. The processes for data extraction from the raw project documents are then discussed and the procedure of data capture is drawn up. It reviews the available project documents to identify parameters that can be assumed to be typical for projects in each process: user input, case retrieval, reviewing cases, and risk identification solutions. To facilitate further understanding of domain knowledge, the later sections of this chapter explain some complicated parameters and analyse the core features of the data set.

7.1 DATA CAPTURE METHOD

Before actually capturing the research data it is necessary to clarify which method and in what procedure the data capture is used. This section develops an outline to guide the selection of the data source from a wide variety of information; extracting useful information from the selected data source; and structuring a data set to suit the intended functions of the model.

Figure 7.1 outlines the stages involved data collection. In order to collect suitable data for this research, several parties were contacted. From these contacts, the most appropriate data source was selected and the reason for the choice of data source is explained. Data samples clarify the relationship between data concepts and systemize the unstructured data into an appropriate order, range and classification. The required data for the study was extracted from the collected samples by detailed manual examination. The data then was channelled into different structures for different functionalities in the risk identification model. To fully understand the method undertaken, the data profile explains some of the complicated data in detail. Finally, data analysis provides further understanding of the overall data set and presents potential biases.

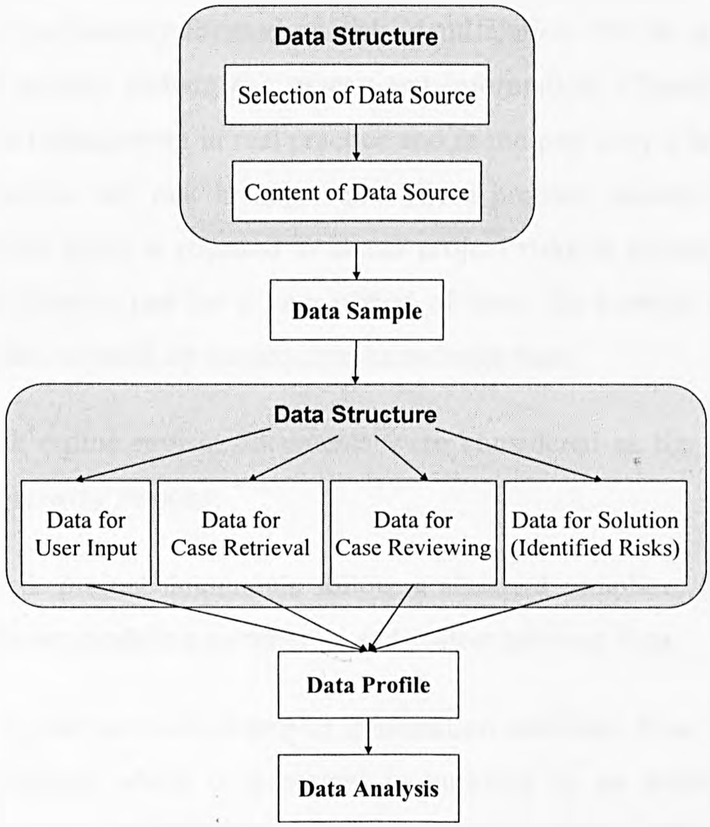

Figure 7.1 Data capture process

7.2 DATA COLLECTION

Data is a collection of disorganized facts that have not been processed into useful information. Since a large amount of data was required when building the risk identification model, extracting data for the knowledge domain was considered important. Data collection includes three components: data source, data sample and data gathering, which will be explained in this section.

7.2.1 Data Source

All research involves the collection and analysis of data. In the case-based risk identification model, the source of data decides the domain knowledge and the main functionality of the model.

Selection of Data Source

This research is particularly focused on risk identification. So the project document selected should already include risk assessment information. Chapter 2 explains the limitation of risk management in real practice and in the past only a limited number of projects have carried out risk management. These projects usually involve a large investment and the client is required to assess project risks in advance. However, by and large these projects last for a long period of time. As a result it is not easy to collect enough data to build up the required knowledge base.

The World Bank online project documents were considered as the appropriate data source for the following reasons:

- World Bank project documents follow a standard template, from which it is easy to extract modeling parameters and collect relevant data.
- There is a great amount of project information available from the World Bank online database, which is important in building up an artificial intelligence system. It provides sufficient cases to develop, test and maintain the system.
- More importantly, the identified risk of each project is listed in the Project Appraisal Documents. These identified risks were generated by project experts from different countries with different backgrounds. On one hand the risk identification domain is based on a very strong knowledge base. On the other hand the reasoning processes of CBR will integrate the experts' knowledge based on wider and stronger experience, which may provide better performance.

Introduction of Data Source

The "World Bank" is the name that has come to be used for the International Bank for Reconstruction and Development (IBRD) and the International Development

Association (IDA). Together, these organizations provide low-interest loans, interest-free credit, and grants to developing countries.

The oldest World Bank agencies, the International Bank for Reconstruction and Development (IBRD), was set up in 1944 at a conference convened in the town of Bretton Woods, New Hampshire, at the end of the Second World War. Its original intention was to provide low interest loans to Europe and Japan to help rebuild their infrastructure after the devastation of the war. Over the next few decades IBRD rewrote its original mandate to provide cheap loans to the Third World instead. In fiscal year 1999, IBRD loaned US$22.2 billion, up from US$21 billion the previous year, making it the biggest source of development capital for Third World countries and the former Soviet Union. Many of these loans are for major industrial development projects such as dam building, power plants, transportation, and mining for non-renewable resources. In addition IBRD dispenses loans for social matters such as education and health, but these loans are often linked to strict economic policies such as Structural Adjustment Programs that have often exacerbated local problems. (World Bank website, 2005).

The International Development Association (IDA) is the part of the World Bank that helps the earth's poorest countries reduce poverty by providing interest-free loans and some grants for programs aimed at boosting economic growth and improving living conditions. IDA funds help these countries deal with the complex challenges they face in striving to meet the Millennium Development Goals. While IBRD raises most of its funds on the world's financial markets, IDA is funded largely by contributions from the governments of the richer member countries. Their cumulative contributions since IDA's beginning up to the end of June 2003 totalled the equivalent of US$118.9 billion. Additional funds come from IBRD's income and from borrowers' repayments of earlier IDA credits.

By the end of January 2005, the World Bank online database contained documents for 10419 projects. However, project documents were not fully available for all these projects. There are usually more detailed electronic documents available for projects approved after 1997. A survey of the available project documents in the World Bank database was made in October 2003. Table 7.1 shows the number of available cases in different sectors and regions. However, because of the change of the database layout, to update these numbers becomes incredibly time consuming. According to Table 7.1, transportation projects in East Asia & Pacific and South Asian areas were considered when building the case base because the author is familiar with the background, policy and environment of these areas and the number projects of available are suitable for developing a case-based reasoning model.

Table 7.1 Available World Bank project cases up to Oct. 2003

Sector	East Asia & Pacific	Africa & sub Sahara	European & Central Asia	Latin America & Caribbean	Middle East & North Africa	South Asia	Total
Transportation	35	22	27	21	8	17	129
Environment	15	18	21	42	4	4	104
Private sector development	0	0	0	0	0	0	0
Public sector development	19	33	36	31	11	6	136
Electric Power and other energy	12	8	12	5	3	8	48
Oil and Gas	4	8	5	1	0	2	20
Total	82	89	101	100	26	37	437

Much of the initial qualitative and quantitative data was collected by exploring a variety of available project documents which provided a number of project details. A World Bank project document normally includes: Project Appraisal Document (PAD), Project Information Document (PID), Integrated Safeguards Data Sheet (ISDS), Resettlement Plan (RPL), Environmental Assessment (EA), Indigenous Peoples Plan (IP), and Board Report (BR). A whole project document can be more than one thousand pages. By sourcing the data from structured project details available, the Project Appraisal Document, Project Information Document, and the 2004 report of status of projects in execution (FY04) were mainly referenced. If information could not be found in these three documents, other relevant documents were then referenced.

The Project Appraisal Document describes an overall project summary in nine sections: Project Development Objective, Strategic Context, Project Description Summary, Project Rationale, Summary Project Analyses, Sustainability and Risks, Main Loan Conditions, Readiness for Implementation, and Compliance with Bank Policies. There are also a number of appendices providing supportive details of each project, such as the Project Design Summary, Project Description, Estimated Project Costs, Cost Benefit Analysis Summary, Financial Analysis, Procurement and Disbursement Arrangements, Institutional Development and Reforms, Management Information System, Financial Management, Environmental Assessment and Action Plan Summary, Project Processing Budget and Schedule, Documents in the Project File, Statement of Loans and Credits, and Country at a Glance.

7.2.2 Data Sample

As the data accumulates, a valuable step is to organize the shapeless mass of data by building typologies and taxonomies, that is, by identifying differences in the data and thereby forming subgroups within the general category (William, 2001). Data sample helps to clarify the relationships among concepts.

The hard data required consists of explicit project details as well as risk description from a sample of several projects. The sample is a pool of transport projects funded by the World Bank. Using projects from within the same sector enables the assumption to be made that the project conditions and required standard of the facility and/or service are common to these projects, or at least very similar. On examination of the sample projects, trends and typical ranges of the relevant values were identified. From these trends and ranges, characteristic details were generated for the data structure of the different stages of a CBR application. The details of project features could then be used for further analysis. This procedure ensures that the data used for a CBR application is representative of the actual project within this sector.

Whilst definitions and criteria (such as project cost, financial source, and economics and social impact) for selecting similar projects were initially drawn up, the level of data actually accessible made adherence to some of them impossible. Several modifications have been made to the data collection. Due to limitations in communication between World Bank experts, the criteria eventually used in the selection procedure was, primarily, that the sample was to consist of projects procured by the project features listed in the project description, and some common features are found in most of the project documents. Transportation sector projects were selected as cases were available and the sector itself is more related to the construction industry. There is also more similarity between transportation projects then may be the case in other sectors. Other criteria stipulated that the projects be from similar regions and, as the data would have to be extracted largely from available project documents, that the layout of the model facilitated the extraction of data.

7.2.3 Data Gathering

During data gathering, one should decide what data might be required, what sources are to be used and the availability and possible methods for collecting the data. Literature sources (Bull, 1989, Hussain and Hussain, 1995) have recommended establishing detailed data requirements by gathering sample data, files, forms and documents from various functional areas.

In choosing suitable data sets for this research there were five steps:

- **Step1:** Searching the World Bank Database to find 1982 projects relevant to transportation, 343 of them in the East Asia and Pacific area and 201 of them in the South Asian area.

- **Step 2:** To reduce the work load in reviewing the data source, only projects approved after 1997 were considered. As a result the cases available were reduced to 94 in EAP and 54 in SA.

- **Step 3:** Visiting each project's website to check the percentages of transportation projects, and removing with less than 50% of transportation. At this stage, there were only 68 projects left.

- **Step 4:** Checking the available project documents and deleting the ones without project appraisal documents. 6 cases fell into this category.

- **Step 5:** Double checking these cases with the 2004 Execution Project Report, and only keeping the ones with transportation as the sector description. Two more cases in rural development and social development were then removed.

As a result of the above processes, only 60 cases were left to construct the knowledge domain of the intelligent risk identification model. With this project list in hand (see Appendix A), the data capture could then be carried out.

Figure 7.2 outlines the stages involved in data gathering. Firstly, documents of selected projects were downloaded. According to the PAD, PID and FY04, the criteria for the construction of the data structure for the knowledge domain were drawn up. Based on this structure, the data was gathered from relevant documents. Then the project data was evaluated. If the project has already been completed, the data will be further verified by the Project Performance Assessment Report before applying criteria to the project data, otherwise the criteria can be applied to the project data directly. Then the project detail was added to the database as a record. Finally the sufficiency of the data set was evaluated. If the number of cases was enough, the case base would be built up according to the database; otherwise more project data should be collected.

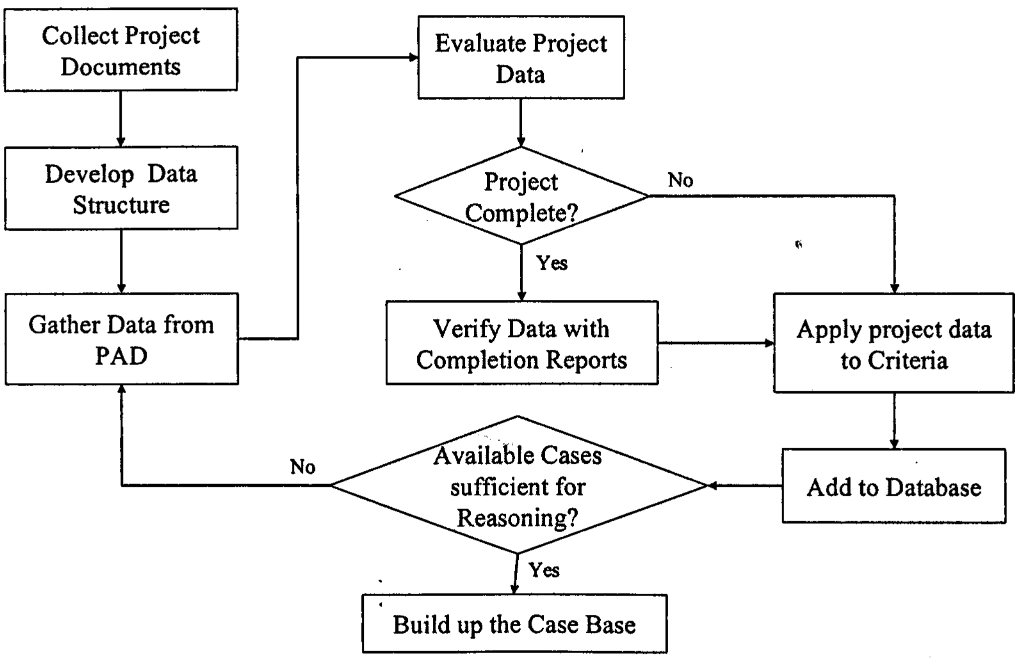

Figure 7.2 Data gathering flow chart

There is a great deal of work involved in building up the data base. Several problems were encountered with influenced the speed of this process:

Units not standard. 16 countries were involved in the areas that were chosen for the collection of data. Each of them has a different currency, and all these currencies had to be converted to US$. In addition, with regard to land acquisition, China uses three different units of measurement: Mu (1.5Mu=1km^2), Hectare (1Hectare=10km^2), and km^2. India usually uses Ha (1Ha=10 km^2).

Information not available directly. Some features could not be gathered directly, and usually needed to be checked further in relevant documents. For example, the finance FNPV and FIRR were not always available in the project description, and it was necessary to reference further finance analysis reports.

Risk interpretation. The identified risk list of raw data usually consists of quite a lot textual description. A great deal of time is spent on picking out the main and common points of the risk description in order to improve accuracy for adaptation of the CBR model.

7.3 DATA STRUCTURE

Fournier (1991) has stated that the data for producing the necessary output must be clearly defined. However, not all data is relevant to a decision making process (Martin and Powell, 1992). Often, data that is relevant to one process may not be relevant to another. The use of irrelevant data not only distracts decision makers' attention from important issues but also leads to poor decisions (Reynolds, 1992).

The purpose of constructing a data structure is to achieve the functionality of CBR retrieval and screen input and output. The collected data can be divided into two main categories: project summary and identified risks. As a result, the data structure reflects this in 4 main sections: one for user input screen, one for case matching, one for solution output screen, and one for case review output screen.

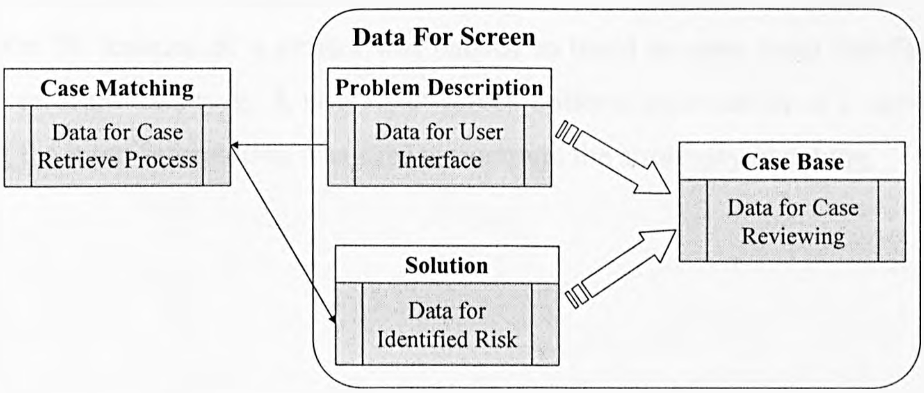

Figure 7.3 Framework of data structure

Figure 7.3 illustrates the framework of the overall data structure. The data required is summarized in the problem description, system provided solution and case base. The user input data plus identified risk data will reform case data to save in the case base for case reviewing purposes. Data for the case retrieval process determines the similarity of a case. Qualitative data is described in text, while quantitative data, such as the project's cost, is described as numerical data. Both textual data and numerical data can be used to produce the similarity score. User inputting data was used for case retrieval and depending on the matching result, the identified risks are listed.

Data for User Input Interface

Although 60 sets of transport project documents were collected, none of them had explicitly specified the exact relationship between the requested data and the risk identification processes. To determine the data structure for each process, a sequential elimination method for analysing semi-structured decision support systems was adopted (Kendall and Kendall, 1988). The principle of this method is to eliminate any irrelevant data by setting constraints. Data which does not satisfy the predefined constraints will be discarded.

Two constraints were established for identifying the relevance of data for the screening and reviewing process. Firstly, the data must be significant to the decision undertaken. Secondly, the data should not only be able to represent the main features of a project (data is relevant to the environmental, financial, technical, and functional characteristics of projects, and diversity of a project), but also to express the diversity aspects of a project. Thirdly, the data should be easy enough to obtain. Table 7.2 listing the 39 features of a project was chosen to build up user input interface along with its relevant data type. A user input screen collects information of a current case. Most of the input information was used to compute the similarity matching.

Table 7.2 Data for User Input: Problem Representation

Attribute	Type	Attribute	Type
Basic Information		**Financial**	
Project ID	T	Finance Source	T
Project Name	T	Amount	N
Region	T	Commitment Fee	T
Country	T	Front End Fee of Bank Loan	N
Sector % of Transportation	N	Service Charge	N
Sub Sector	T	Proposed Terms	T
Program of Targeted Intervention	T	Main Loan/Credit	T
Project Objective Category	T	Government Commitment	N
Environment Category	T	IBRD Commitment	N
Grace Period	N	IDA Commitment	N
Year to Maturity	N	Total Project Cost	N
Project Approval Year	N	Product Line	T
Implementation Period	N	Lending Instrument	T
Effective Date	N	Financial_NPV	N
Project Status	T	Financial_FIRR	N
Land Acquisition	N	**Economics**	
Resettlement People	N	Economical_NPV	N
Borrower	T	Economical_EIRR	N
Implementing Agency	T	Economics Evaluation Method	T
Targeted Thematic Outcomes	T	Economical Analysis	T

N-Number T-Textual description

7.3.2 Data for Case Retrieve

The attributes that contribute to the case retrieval is derived from Table 7.2. Table 7.3 removes some features that might not be related to other projects, such as project ID, project name, and main load credit code. The remaining 36 features contributed to the case retrieval process.

As the data in ART*Enterprise (AE), CBR application can be divided into a variety of types to meet different match requirements. The relevant AE cbr type for each feature was also decided. The numerical data was transferred to cbr:range. cbr:range match is based on numeric distance between numeric values in the presented and stored case. Whilst the textual description was divided into three types: those with short words (less than 3 words) were defined as cbr: symbol; those with medium words (less than 8 words) were defined as cbr:string; and those with long words were defined as cbr:

word. cbr:symbol converts symbols to strings and then uses exact string matching; cbr:string uses exact string matching; and cbr:word follows text pre-processing to remove misspellings, ignored words, and suffixes. In addition, attribute values from stored cases are recorded as words to match against.

Table 7.3 Data for CBR retrieve process

Attribute	Type	Attribute	Type
Basic Information		**Financial**	
Region	cbr:string	Finance Source	cbr:string
Country	cbr:string	Grant Amount	cbr:range
% of Transportation	cbr:range	Commitment Fee	cbr:range
Sub-sector	cbr:string	Front End Fee of Bank Loan	cbr:range
Program of Targeted Intervention	cbr:string	Service Charge	cbr:range
Project Objective Category	cbr:string	Proposed Terms	cbr:word
Environment Category	cbr:string	FP_Government	cbr:range
Grace Period	cbr:range	FP_other	cbr:range
Year to Maturity	cbr:range	IBRD Commitment	cbr:range
Project Approval Year	cbr:range	IDA Commitment	cbr:range
Implementation Period	cbr:range	Total Project Cost	cbr:range
Effective Date	cbr:string	Product Line	cbr:string
Project Status	cbr:string	Lending Instrument	cbr:string
Land Acquisition	cbr:range	Financial_NPV	cbr:range
Resettlement People	cbr:range	Financial_FIRR	cbr:range
Borrower	cbr:word	**Economics**	
Implementing Agency	cbr:word	Economical_NPV	cbr:range
Targeted Thematic Outcomes	cbr:word	Economical_EIRR	cbr:range
		Evaluation Method	cbr:string

7.3.3 Data for Reviewing

Data for the reviewing process contains the full project details, not only the project description but also the overall risk level and risk list in company with the individual risk rating and mitigation strategy. 52 parameters of the project were listed in Table 7.4.

This data set provides more supportive information of matched cases for decision maker reference. When there are no similar cases in the case base, or the solution of system output was unsatisfactory for the project manager, this information would help the project manager to identify the risks based on the matched cases.

Table 7.4 Data for user reviewing details of relevant project

Basic Information	**Financial**
Document ID	Finance Source
Project ID	Grant Amount
Project Name	Commitment Fee
Region	Front End Fee of Bank Loan
Country	Service Charge
% of Transportation	Proposed Terms
Sub-sector	Main Loan/Credit
Program of Targeted Intervention	IBRD Commitment
Project Objective Category	IDA Commitment
Environment Category	Total Project Cost
Project Status	FP_Government
Grace Period	FP_other
Year to Maturity	Product Line
Implementation Period	Lending Instrument
Effective Date	Financial_NPV
Approval Date	Financial_FIRR
Closing Date	**Economics**
Land Acquisition	Economical_NPV
Resettlement People	Economical_EIRR
Borrower	Economics Evaluation
Implementing Agency	Economical Analysis
URL of Project	**Potential Risks**
Project Introduction	Risk Description
Development Outcomes And Goals	Individual Risk Level
Targeted Thematic Outcomes	Mitigation Strategy
Supporting this Broader Development Goal	Overall Risk Level

7.3.4 Data for Solution

All data relevant to risk identification and the overall risk assessment (see Table 7.5) was used for outputting the solution. This data is reflected in the CBR adaptation process.

As shown in Table 7.5, this data set consists of four components. The risk description lists key risk factors together with an impact rating and a suggested imitation measure respectively. Then an overall risk level of the whole project was also presented. The risk rating was divided into four levels: H represents high risk; M represents modest risk, S represents substantial risk, and N represents negligible or low risk.

Table 7.5 Data for identified risk: solution

Potential Risks	Note
Risk Description	Risk List
Individual Risk Rating	H, S, M, N
Mitigation Measure	Textual
Overall Risk Level	H, S, M, N

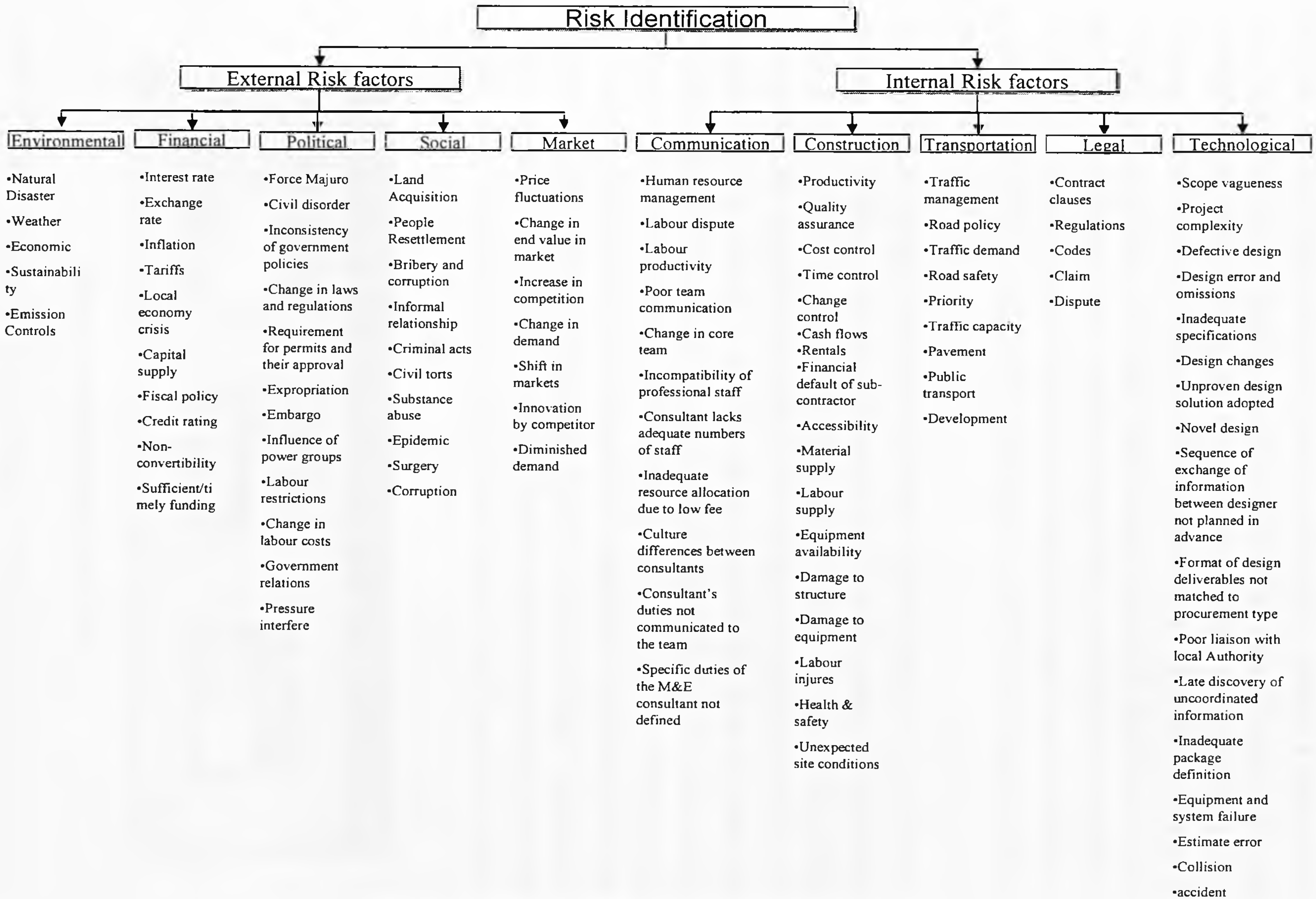

Figure 7.4 Risk Identification Breakdown Structure

To organize the collected risk data in an appropriate format for the adaptation purpose, the 556 risk items from 60 selected projects are extracted to generate a standard risk log. Each risk item is evaluated and grouped with similar items. The author uses the risk breakdown structure as the way to classify these risk items, as shown in Figure 7.4. Various external risk factors in terms of five major classes are identified: environment, finance, politics, society, and market. Similarly, the internal risk factors are divided into five major classes, which include communication, construction, transportation, legal, and technology. The project risks can be internal to the organization in which the project is carried out, or external to the organization, within the environment in which the organization exists.

If an individual risk item contains various detailed risk descriptions, it will be divided into separate individual risk items and then put into a relevant category. The classified risks are then sorted and picked up to be assigned a risk code. The risk item, along with relevant risk code is then generated as the risk log containing 126 risk items (see Appendix B). The risk code, is then allocated to each risk item in the risk base, and then the relationship between the risk code and risk base is built up. As illustrated in Figure 7.5, the risk base consists of three tables: overall risk level, risk list and risk log. The overall risk level and risk list is connected through project ID, while the risk list and risk log is linked through risk code. To review the risks of each individual project, a query can be generated based on project ID. Based on this relationship, the project ID, project name, overall risk level, risk items and relevant risk rating can be easily displayed.

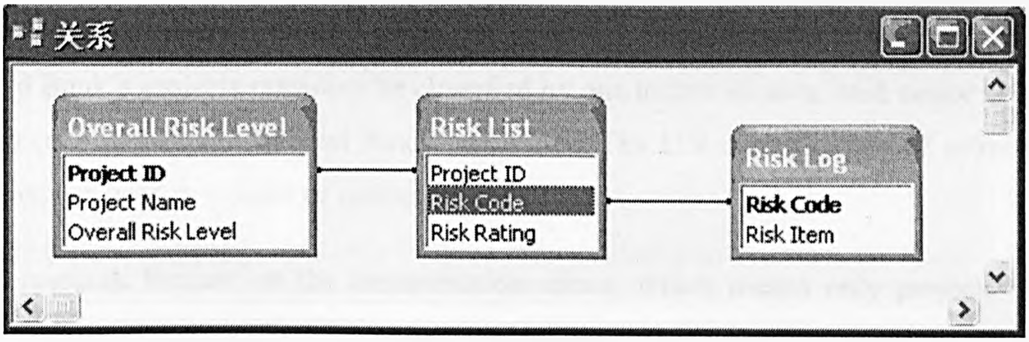

Figure 7.5 Relationship of risk base

Based on this structure, the learning processes of case-based reasoning introduced in Chapter 8 can be carried out.

7.4 DATA PROFILE

In this research, data from sixty transportation projects funded by the World Bank were analyzed. All the data was collected from the East Asia & Pacific and South Asia areas and was constructed between 1997 and 2004. Thirty-six attributes were decided to contribute to the case retrieval process. Here, some of key features were summarized, including: grace period, sector, front-end fee, commitment fee, NVP, EIRR, FIRR, and lending instrument.

7.4.1 Grace Period

IDA credits have maturities of 20, 35 or 40 years with a 10-year grace period before repayments of the principal begins. IDA funds are allocated to the borrowing countries in relation to their income levels and record of success in managing their economies and their ongoing IDA projects. There is no interest charge, but credits do carry a small service charge, currently 0.75 percent on funds paid out.

7.4.2 Sector

Sectors are a high-level grouping of economic activities based on the types of goods or services produced. The World Bank sector includes nine categories: agriculture, fishing and forestry; Education; Energy and mining; Finance; Health and other social services; Information and communication; Law and justice and public administration; Transportation; and Water, sanitation and flood protection. In November 2001, a decision was taken by the Bank's Management Committee to use a new system of sectors and themes in order to better measure the Bank's operational work. All of the World Bank's projects may now be classified by one to five sectors; each sector is now assigned a percentage of total funds committed. The UN classification of economic sectors was used as a point of reference.

This research focuses on the transportation sector, which means only projects with over 50% of transportation are under consideration. There are five sub-sectors under transportation sector: aviation, general transportation sector, ports waterways and shipping; railways; and roads and highways.

7.4.3 Front-end Fee

Front-end fee refers to the fee paid by a borrower to a lender at the beginning of a loan facility, or the cost to an investor of buying into a unit trust such as an equity or property trust. Front-end fees vary substantially, so it is advisable to check the amount applied in each case.

7.4.4 Commitment fee

A commitment fee refers to a charge by a lender for holding credit available for a borrower. It is fee lenders charge their borrowers for unused credit or credit that has been promised at a specified future date. A commitment fee is different to interest although the two are often confused. A lender charges a borrower a commitment fee to keep a line of credit open, or to guarantee a loan at a certain future date even though the credit is not being used at that particular time.

7.4.5 NPV - Net Present Value

The Net Present Value (NPV) of an investment project indicates the expected impact of the project on the value of the investment. Projects with a positive NPV are expected to increase the value of the investment. When choosing among mutually exclusive projects, the project with the largest (positive) NPV should be selected. Otherwise, if negative, the investment should not be made.

The NPV is calculated as the investment's future net cash flows minus the initial investment. This relationship is expressed by the following formula (Mathis, 2005):

$$NPV = \sum_{t=0}^{T} \frac{CF_t}{(1+r)^t} = CF_0 + \frac{CF_1}{(1+r)^1} + \frac{CF_2}{(1+r)^2} + \ldots + \frac{CF_T}{(1+r)^T}$$ *Equation 7.1*

where CF_t is the cash flow at time t; and r is the cost of capital.

EIRR- Economic Internal Rate of Return

In public project appraisal, social or economic benefit must account for externalities by incorporating social or shadow prices which are based on international or accounting prices in World Bank (WB) manuals. When externalities and price distortion are allowed for in the cost and benefit stream, the social criteria are the same, i.e., NPV and economic internal rate of return (EIRR). Since the EIRR criteria are ratios, they are scale-neutral, whereas the NPV being a difference increases with the size of the project. NPV and EIRR are related in Table 7.6.

Table 7.6 Relationships between NPV and EIRR

NPV	EIRR
If > 0 then	> r [cost of capital]
If < 0 then	< r
If = 0 then	= r

These criteria can give conflicting results in ranking projects, and in deciding whether to accept or reject a project. Projects with less than 10 percent EIRR, the usual threshold for acceptance, may be accepted because of special social contributions now fashionable and politically correct. Different risks between projects do not seem to be a factor taken into account of in relation to rates of return. The marginal cost of capital is assumed to be the same for all standard projects. A complicated decision matrix exists for selecting projects ranked differently by NPV or rate of return, depending on whether there is more than one project and if projects are mutually exclusive. In practice, projects are selected several times in relation to an acceptable in-house rate of return in both the public and private sectors.

When making decisions on loans current practice among international financial institutions is to analyze costs and benefits in terms of the economic internal rate of return (EIRR). The internal rate of return is a discount rate calculated to equalize the net present value of cost with that of benefit. In practice the minimum viable level of EIRR will depend on the circumstances of each country at each chronological stage.

In developing countries, generally, projects with an EIRR estimated in excess of 12% tend to carry a high priority for realization. Economic theory shows that using the Net Present Value (NPV) criterion usually leads to more robust rankings than for the EIRR.

FIRR - Financial Internal Rate of Return

The FIRR is an indicator to measure the financial return on investment of an income generation project and is used to make the investment decision. The FIRR is obtained by equating the present value of investment costs (as cash out-flows) and the present value of net incomes (as cash in-flows). This can be shown by the following equation.

$$\sum_{n=0}^{m} \frac{In}{(1+r)^n} = \sum_{n=1}^{m} \frac{B_n}{(1+r)^n}$$ *Equation 7.2*

where I_0 is the initial investment cost in the year 0 (the first year during which the project is constructed) and $I_1 \sim I_m$ is the additional investment cost for maintenance and rehabilitation for the entire project life period from year 1 (the second year) to year m. $B_1 \sim B_m$ is the annual net income for the entire operation period (the entire project life period) from year 1 (the second year) to year m.

By solving the above equation, we can obtain the value of r and this r is the Financial Internal Rate of Return (FIRR).

With increasing private finance initiative (PFI) in the public sector, and policies for cost recovery in urban infrastructure projects, FIRR has become more important. In the private sector, FIRR is compared to the weighted average cost of capital (WACC) in real terms. Whereas the opportunity cost of capital may be fixed for any country, the cost of capital may differ between projects depending on sources of capital (Hufschmidt, et al., 1983). See Table 7.7.

Table 7.7 FIRR vs. WACC, five possibilities

FIRR > WACC	Project covers cost and yields profit
FIRR = WACC	No private profit
FIRR < WACC	Project requires subsidy
FIRR = ZERO	Project covers costs except cost of finance
FIRR < ZERO	Cost recovery not possible

This analysis reveals the subsidy requirements of projects which are deemed socially desirable but do not cover costs.

7.4.5 Lending Instrument

The World Bank has two basic types of lending instruments: investment loans and adjustment loans. Investment loans have a long-term focus (5 to 10 years), and finance goods, works and services in support of economic and social development projects in a broad range of sectors. Adjustment loans have a short-term focus (1 to 3 years), and provide quick-disbursing external financing to support policy and institutional reforms. Both investment and adjustment loans are used flexibly to suit a range of purposes and are occasionally used together in hybrid operations (World Bank website, 2005).

Investment loans provide financing for a wide range of activities aimed at creating the physical and social infrastructure necessary for poverty alleviation and sustainable development. There are a variety of lending instruments, such as Specific Investment Loans (SILs), Sector Investment and Maintenance loans (SIMs), Adaptable Program Loans (APLs), Learning and Innovation Loan (LIL), Technical Assistance Loan (TAL), Financial Intermediary Loans (FILs), and Emergency Recovery Loans (ERLs). In this research, the projects are mainly using two of the above: SIL and APL.

- Specific Investment Loans (SILs) support the creation, rehabilitation and maintenance of economic, social, and institutional infrastructure. In addition, SILs may finance consultant services and management and training programs.

- Adaptable Program Loans (APLs) provide phased support for long-term development programs. They involve a series of loans that build on the lessons learned from the previous loan(s) in the series.

7.5 DATA ANALYSIS

Data analysis is about the tools, techniques and resources required. Various types of data are continuously generated as the project proceeds. A wide range of statistical analysis on gathered data will be carried out in order to obtain a full understanding of them. Three popular statistical analysis packages were employed in this analysis: SPSS 12, Minitab 14, and Microsoft Excel XP.

The data required for the model were the risks as well as the project summary. The structure and elements of the original data were based on World Bank project documents. Data from 60 incomplete or completed projects were analysed for developing the model. More detailed explanation on some of the key components that were not explained in the data profile section are depicted in following sections.

7.5.1 Amount

The financial scheme of a project consists of several sources: the IBRD Loan, IDA Credit, Grant amount, and government/ borrower commitment.

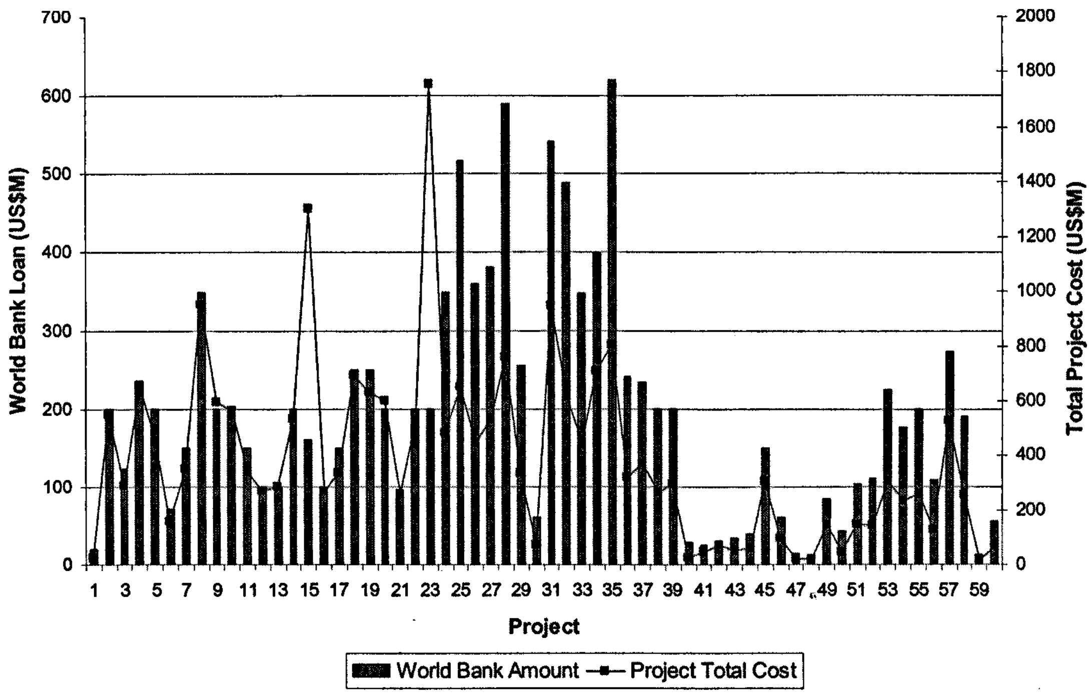

Figure 7.6 The World Bank Loan vs. total project cost of individual project

The sum of the IBRD loan and IDA credit makes up the World Bank Loan. The total of the project cost includes World Bank commitment, Government commitment, and any other grants. The Figure 7.6 illustrates the amount of the World Bank loan and total project cost for each project in the knowledge base.

7.5.2 Country

Projects from 16 countries are considered in this research. Six of them are from the South Asian area: Afghanistan, Bangladesh, Bhutan, India, Nepal, and Pakistan. The remaining 10 are from the East Asia and Pacific areas, including: Cambodia, People's Republic of China, Indonesia, Lao Demonstration of Republic, Mongolia, Papua New Guinea, Philippines, Samoa, Thailand and Vietnam. The number of cases selected from each country was illustrated in Figure 7.7.

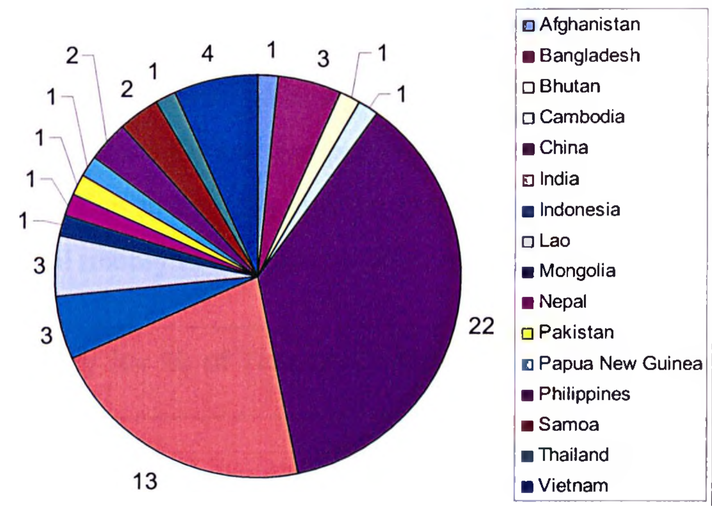

Figure 7.7 Number of projects in Country

7.5.3 Year

As stated in previous sections, the selected projects were took place between 1997 and 2004. Figure 7.8 illustrates the number of projects in the knowledge base each year between 1997 and 2004.

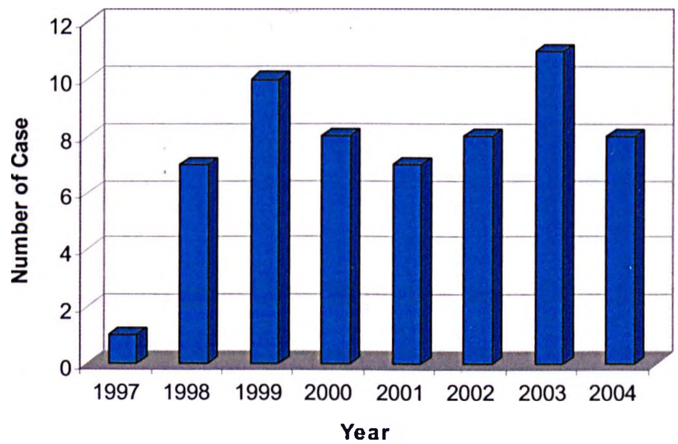

Figure 7.8 Number of projects in the year 1997-2004

7.5.4 Percentage of Transportation

A project usually consists of several separate parts. For example, a rural development project can constitute two thirds transportation project and one third environment project. As a result, the percentage of transport work involved in a project is the main parameter to identify the main sector and functionality of a project.

Figure 7.9 summarizes the statistical analysis of the chosen project's transportation percentage. The results show the average percentage to be 89.8%, the median value to be 94%, the lowest value to be 55% and the highest value 100%, which means the projects chosen are all mainly in the transportation sector.

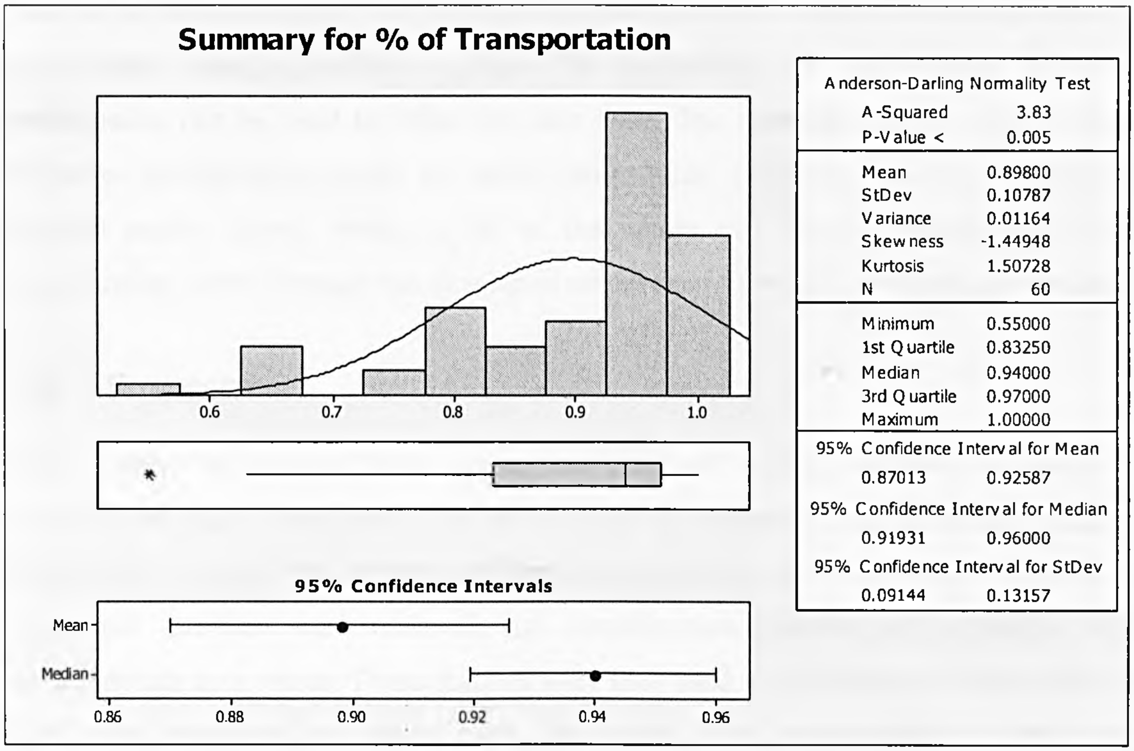

Figure 7.9 Statistical summary for percentage of transportation

7.5.5 Project Status

Project status demonstrates whether the project was complete, proposed or still active. In the project database of the model, there is no proposed project; only 3 projects have been completed (5%); and most of the projects are still active (95%), as shown in Figure 7.13.

PAGE NUMBERING AS ORIGINAL

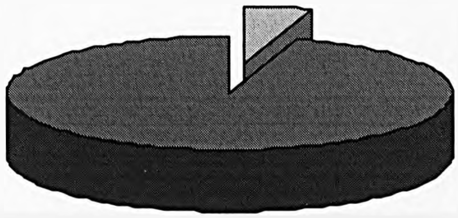

Figure 7.10 Projects' status of case base

The project status is one of the important factors to reference when evaluating system performance because, when a project is completed, the information of real performance can be used to refine the case base. The more the case is refined, the better the identification output the model can provide. Currently there are only three projects totally closed, which is 5% of the whole case base. Therefore, the risk identification output through the developed model might not be as accurate as expected.

7.6 SUMMARY

This chapter has outlined the processes undertaken to collect, interpret and analyse data for this thesis. The data collected was used to determine typical characteristics that might be considered for each process of the risk identification model, including user input interface, case retrieval, risk identification solution and reviewing the project details as a whole. These features were then used as parameters around which a CBR risk identification model to suit the World Bank transportation sector was developed.

CHAPTER 8

CBRISK

8.0 INTRODUCTION

This Chapter demonstrates the design process of CBRisk, a prototype **C**ase-**B**ased **R**easoning model for **Risk** identification. During the system development phase, based on the framework introduced in Chapter 6 and using the data captured in Chapter 7, the detailed design specifications were coded into the selected shell ART*Enterprise. It begins with the architecture of CBRisk, which provides an overview of how the model works. The process of data mapping integrates the external database with ART*Enterprise. Each component of the CBR application is then described, including case index, case representation, case retrieval and case adaptation.

8.1 ARCHITECTURE OF CBRISK

As explained in Chapter 6, CBRisk consists of five interrelated modules: criteria formulation, knowledge base, data mapping, graphic user interface, and CBR application. Figure 8.1 shows the architecture of CBRisk and the interactions between its modules. These modules are represented as different colour shaded boxes. System input and system output together are part of the graphic user interface module. Because data mapping integrates the components of knowledge base, they share a common box. The white boxes and cylinders within each module denote the processes of data stores respectively, while the boxes shown in the system input and output symbolise the information required and generated by the system. Information flowing within the system is shown by arrows.

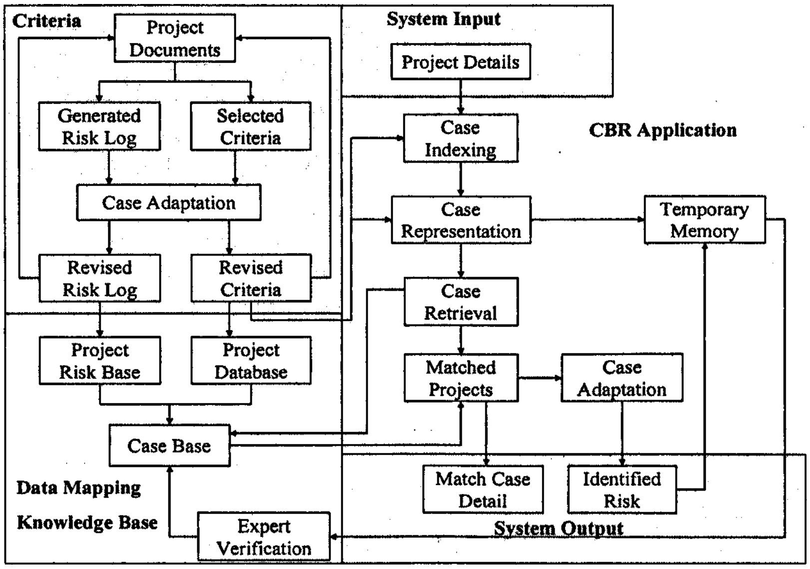

Figure 8.1 Architecture of CBRisk

The system originates from the criteria formulation module where a set of appropriate criteria for the subsequent modules is determined. From the project appraisal documents, the risk log and initial criteria are generated. They are then tested through case adaptation. Based on the adaptation result, the relevant modifications are made and the appropriate risk log and criteria are used to build the project risk base and project database respectively. The project risk base and project database are then developed to case base through data mapping. The risk base, project database and case base together form the knowledge base.

The user is required to enter the relevant project information into the screening. This information, along with satisfactory criteria, is carried out by case indexing and then case representation. Based on the represented information of the new project, case retrieval compares the similar case with the case in the case base and then a match project list along with match score is provided. Case adaptation is then carried out according to the match result and the identified risks are output as a solution for this project. Meanwhile, the details of matched cases can be output for the user to review. The identified risk and represented case are initially firstly stored in a temporary space as a potential new case. After passing the verification of domain experts, they can be saved to the case base as a new piece of knowledge for future use.

8.2 DATA MAPPING

CBRisk was developed using Mindbox's ART*Enterprise (AE) development environment operating under Microsoft Windows. This was chosen because of its flexible CBR facilities, integrated seamlessly with rule-based and object-oriented capabilities. Another important feature of the ART*Enterprise is its ability to create case-bases from existing data sources in various formats, such as Microsoft Excel and Access. In addition, the environment provided user interface construction utilities which allowed demonstrating case indexing, retrieval, adaptation, and reusing to be built quickly.

According to the data mapping scheme introduced in section 6.6, the external Access/Excel database and CBRisk application can be linked by using the Data Integrator shell of AE. In order to transfer data between on external database and CBRisk it is necessary to build a new user data source in Windows' ODBC. As shown in Figure 8.2, ART Access and ART Excel connect with relevant types of external Access database and Excel spreadsheet respectively.

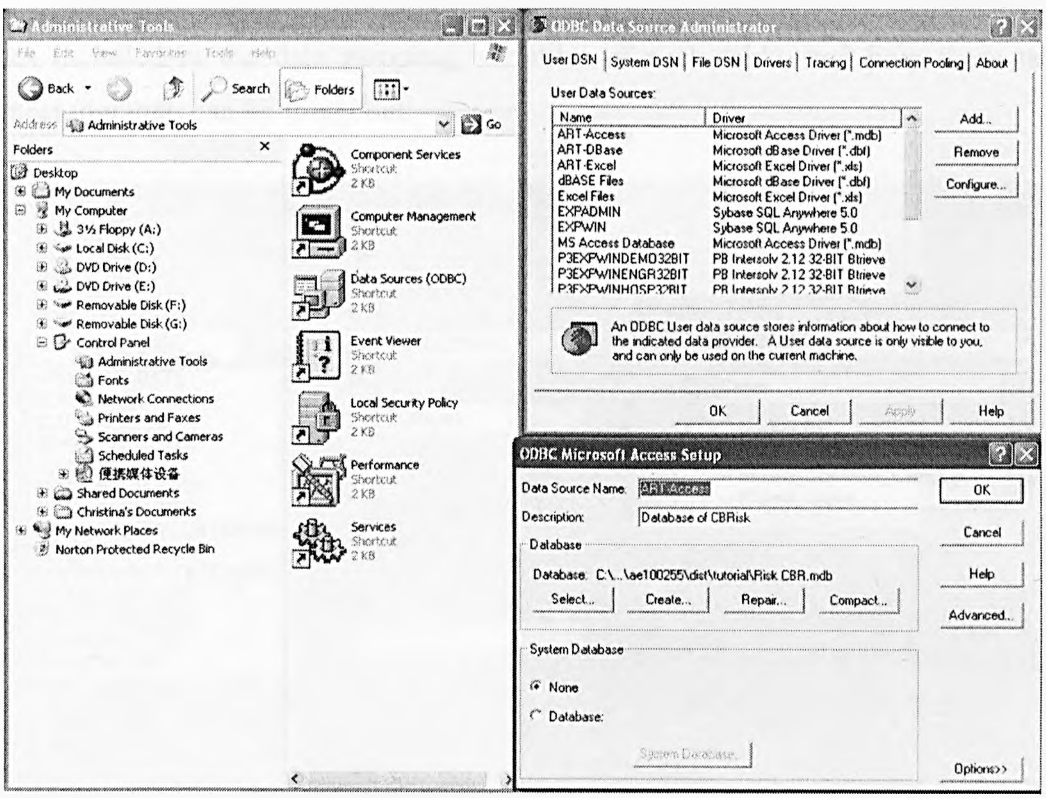

Figure 8.2 Building ODBC connection

Once the ODBC connection has been established it is ready to integrate the external Access/Excel data to CBRisk. The user input interface is designed to provide an easy means by which an end-user can input data into the system. The first step of the program is to build the connection between the external Access database and the DI objects.

Figure 8.3 shows the interface of the data integration. The target is a class, also called the target object or target class. Normally each attribute of the target corresponds to a field in a database table. When creating a new target it is actually creating an entirely new Data Integrator object set. As you connect to database tables, a target attribute is created and mapped to each field of each table. Only mapped attributes are displayed in the Target area. Every mapping must connect to at least one database. When connecting a mapping to a database, external data objects are created and associated with that mapping. These objects are ART*Enterprise representations of databases, tables, and fields.

The left side drop down box lists all attributes in the target data objects. The right hand side illustrates the status of data connection. There the connected external database the tables included in the data mapping, variables of each table, and how these table connect together, can be found out.

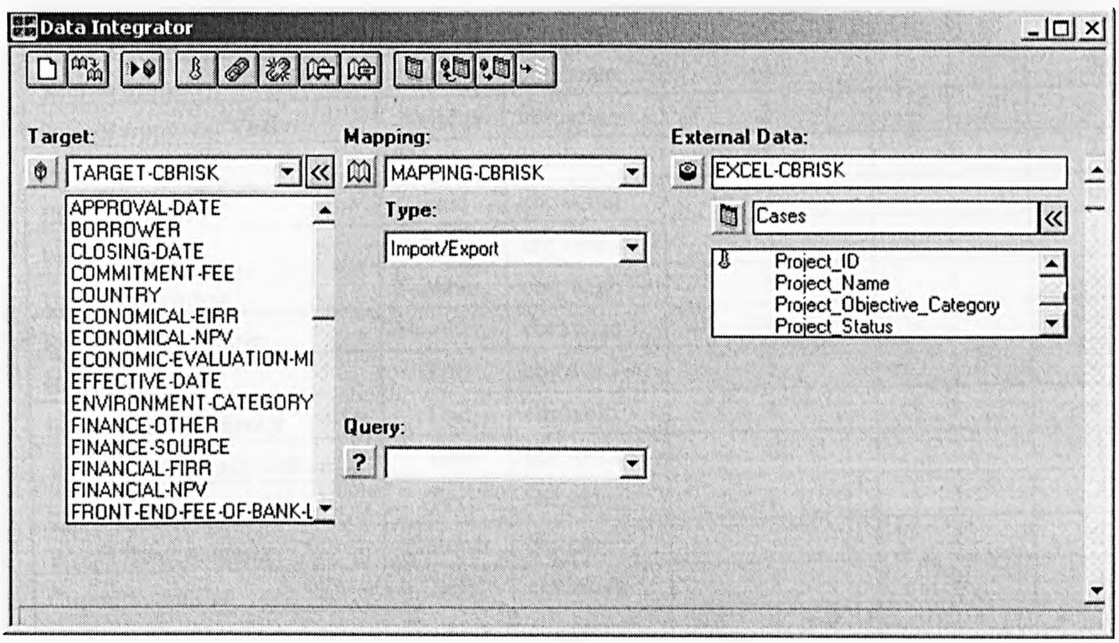

Figure 8.3 Data mapping of CBRisk

After building up the connection between the external Access database and the CBRisk application, the initial user entry form according to this data mapping can be developed.

8.3 CASE INDEX

According to Section 8.3, forty attributes are used for case indexing, and thirty-six of them contribute to the case match. The respective data type, match type and match scores are illustrated in Table 8.1.

Table 8.1 Match attributes of CBRisk

Data	Data Type	Match Type	Match Contribution	Mismatch Penalty	Absent Penalty
Project ID	Text				
Project Name	Text				
Main Load Credit	Text				
Closing Date	Text				
Region	Text	cbr:string	3	1	1
Country	Text	cbr:symbol	5	2	2
Sub Sector	Text	cbr:string	6	3	1
% of Transportation	Number	cbr:range	5	2	2
Program of Targeted Intervention	Text	cbr:symbol	2	1	1
Project Objective Category	Text	cbr:string	2	1	1
Environment Category	Text	cbr:symbol	5	3	1
Grace Period	Number	cbr:range	4	2	2
Year to Maturity	Number	cbr:range	4	2	2
Project Approval Year	Number	cbr:range	3	1	1
Implementation Period	Number	cbr:range	5	2	2
Effective Date	Text	cbr:string	2	1	0
Project Status	Text	cbr:symbol	3	1	1
Land Acquisition	Number	cbr:range	5	2	1
Resettlement People	Number	cbr:range	4	2	1
Borrower	Text	cbr:word	5	3	2
Implementing Agency	Text	cbr:word	4	1	1
Targeted Thematic Outcomes	Text	cbr:word	6	2	1
Finance Source	Text	cbr:string	6	2	2
World Bank Amount	Number	cbr:range	7	2	2
Commitment Fee	Text	cbr:string	5	2	1
Front End Fee of Bank Loan	Number	cbr:range	3	1	0
Service Charge	Text	cbr:string	3	0	0
Proposed Terms	Text	cbr:word	5	1	1

FP_Government	Number	cbr:range	4	2	1
FP_other	Number	cbr:range	2	0	0
IBRD Commitment	Number	cbr:range	3	1	0
IDA Commitment	Number	cbr:range	3	1	0
Total Project Cost	Number	cbr:range	7	3	2
Product Line	Text	cbr:string	3	1	0
Lending Instrument	Text	cbr:string	3	1	1
Financial_NPV	Number	cbr:range	4	1	1
Financial_FIRR	Number	cbr:range	5	2	1
Economical_NPV	Number	cbr:range	5	1	0
Economical_EIRR	Number	cbr:range	6	2	2
Economical Analysis Method	Text	cbr:string	3	1	1

There are four types of cbr data listed in the table. cbr:symbol copes with the single character match; cbr:string deals with string match, cbr:word is used for long textual information, and cbr:range is used to counting the difference of numbers within a range. Different from the other three cbr types, cbr:range needs defining into three more control values: lowest attribute value, highest attribute value, and match interval value.

The match score is accumulated from three types of matching weight: the match contribution score, the absent penalty score and mismatch penalty score. There are two stages by which the matching weights of each of the attributes can be allocated: initial allocation based on attribute importance, and further justification.

In the first stage, attributes are reviewed according to the data analysis results and the overall importance to represent a project. The initial weights for each attributes are then allocated.

By considering the match results it should be possible to identify the differences of two cases relevant to the current case. It is not expected that a match should show more than two cases sharing the same score are always in the top ten matched cases, unless the match score is 0. Several constraints were used to improve matching weight allocations.

- Absent Penalty ≤ Mismatch Penalty and ½ Mismatch Penalty ≤ Match Contribution. In order to control the match score in a reasonable range (not to be a minus value, and to be able to reflect a good space according to similarities), the total match score is designed to be 200, the mismatch score is 60, and the absent penalty score is 40.

- If up to 50% of cases in the case base have attribute values which are empty, then the absent penalty will be set as 0.
- The more important the attribute is, the closer the absent penalty to the mismatch penalty.
- The lowest and highest attribute values are defined for cbr:range, based on the highest and lowest value of each attribute, and give out 10% grace range for each value, if the lowest value is not 0. They are then ranged to the closest value scale and set as the final values.

Once the case indexing is completed, it is time to build the object of criteria to represent the knowledge of each project in relation to the CBRisk model.

8.4 CASE REPRESENTATION

The main objective of the knowledge acquisition was to develop useful representations of the related information processing model in risk identification (Ugwu, *et al.*, 2004). Case representation is the process for determining the contents of cases and their organization. In its simplest form, a case is an entity described by a list of features (attribute value pairs). In CBRisk, the case is retrieved through an attribute label, so a list of attribute associated case labels is stored as the value of a variable this is called an indexing list.

Case index builds the structure for the instances of a case base, and case representation will actually construct these instances (or cases). In general, a case is a unique knowledge entity describing a problem and a solution. A case can be represented as a single database "object" or broken into two or more associated objects. A typical case has the following fields:

- Title

- Problem (or situation) description
- Solution

In order to support the questions and answer sessions, each case is associated with question objects. The association can be direct, i.e. expressed in the case, or implied, in which case the CBR system calculates which questions to ask. This is described in the following section.

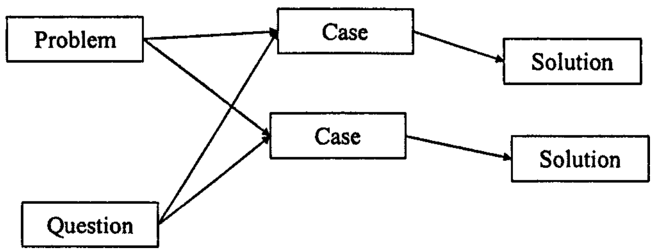

Figure 8.4 Case representation formats

Figure 8.4 shows a generic structure of a CBR knowledge representation. Usually the representation is:

- a problem points to one or more cases;
- a case has a single solution; and
- a question can influence one or more cases.

As mentioned in Figure 6.10, the problem reflects the project description, while the solution reflects the identified risks. Each entry in relation to the project description was made uniform. Similarly, after reviewing the 558 risk items of the projects in the case base, a risk log was generated to provide a uniform description for each risk.

Figure 8.5 depicts an example of case representation. Each individual case is described by features of each project. The top level is the primary information for each project. The middle level is the user data entry to provide the project description information. The bottom level is the risk lists.

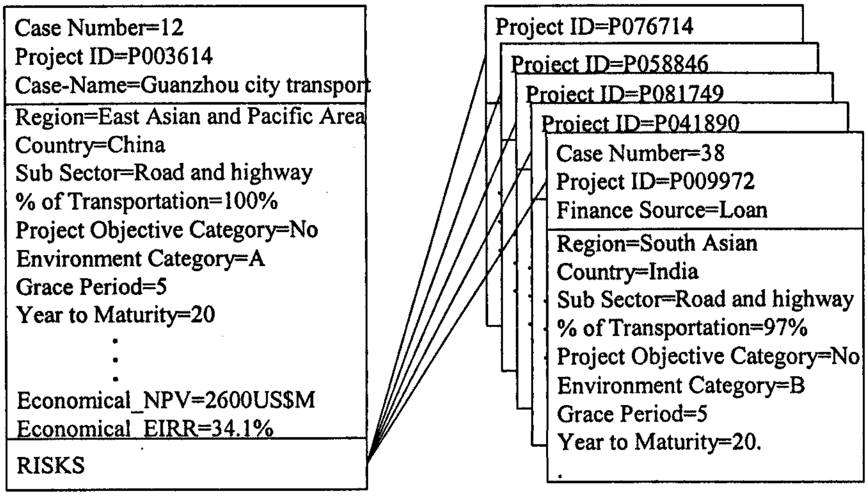

Figure 8.5 An example of case representation

After case representation has been completed, cases can be loaded to the case base. This case base is then used for case retrieval.

8.5 CASE RETRIEVAL

The main performance in case retrieval is to compare the current project to previous ones in the case base and retrieve the similar cases for a more precise identification. Given a description of a problem's requirements, retrieval functions must find a small set of useful similar cases. Matching is the process of comparing two cases or case attributes to each other and determining their match. Ranking is the process of ordering partially matching cases according to the number of matches (Maher, *et al.,* 1995). While retrieval is based on feature similarity, adaptation is based on feature differences.

Case matching is the core of a CBR application. It computes similarity between a new case (the user's input) and cases in the case base. Most CBR system retrieval relies on a nearest-neighbour method, as does CBRisk. In this method, the weighted sum of features in the input case is compared to historical cases. The nearest-neighbour algorithm uses statistical methods to determine the optimal set of features and the number of cases that should be used to calculate similarities. The cases are associated with qualitative and quantitative parameters called attributes, and relevant scores. The CBR algorithm determines the similarity between cases based on these values.

A similarity measure should have the following attributes:

- Reflective: A case is similar to itself
- Symmetric: If A is similar to B, then B is also similar to A

Similarity between two cases is based on the similarity between the features of the two cases'. Similarity depends on the calculation of both numeric values and non-numeric values. Similarity also depends on the range of permitted values. For example, in the range of 1 to 100, 10 is not very similar to 20, but in the range of 1 to 10,000, the two may be considered similar.

To build up a case matching scheme, the relevant classes and objects are defined. The matching process is a straight forward use of the in-built capabilities of ART*Enterprise. Each attribute (see Table 8.1) is compared and scored according to its type: symbol, string, word and range. Each attribute was allocated relevant match values, which then contributed to the case matching process. Taking into account the weights these are combined to given an overall score for the case. The following formula is used in calculation for case matching.

8.5.1 Case Score

ART*Enterprise provides three means for calculating the case score: score based on the case representation criteria, score based on the attributes that the presented case contributes to score; and score based on both. Equation 8.1 below provides a universal formula for calculation of the case matching score.

$$Case-Score = \frac{\sum_{i=1}^{n} attribute-score_i}{\sum_{i=1}^{n} MatchContributionScore_i} \times 100 \qquad \textit{Equation 8.1}$$

When the n in Equation 8.1 represents all the criteria for the presented case, Equation 8.1 can be expressed as Equation 8.2. The sum of the match contribution for each attribute in the presented case regardless of its value, is called maximum-presented-score.

$$Case-Score=\frac{\sum_{i=1}^{n} attribute-score_i}{\max imun-presented-score}\times 100 \qquad \textit{Equation 8.2}$$

Scoring with maximum presented will find the best matching cases for incomplete data. This means that the score will decline as you specify additional data.

When the n in Equation 8.1 stands for all the criteria that generated the case base, Equation 8.1 can be expressed as Equation 8.3. The sum of the match contribution for each attribute in the presented case regardless of its value is called maximum-stored-score.

$$Case-Score=\frac{\sum_{i=1}^{n} attribute-score_i}{\max imun-stored-score}\times 100 \qquad \textit{Equation 8.3}$$

Scoring with stored max criteria will find the most complete matches for the presented case. It assumes that the presented case is complete, otherwise cases with a great deal of data which perfectly match the presented case will match poorly.

To integrate both scoring methods above, the case score can be calculated with Equation 8.4 below.

$$Case-score=\frac{\frac{\sum_{i=1}^{n} attribute-score_i}{\max imum-stored-score}\times 100+\frac{\sum_{i=1}^{n} attribute-score_i}{\max imum-presented-score}\times 100}{2}$$

Equation 8.4

This scoring method works for two kinds of problems: firstly, when both the presented case and the stored cases are likely to have extra information in matching cases; secondly, when it is not necessary for scores to be perfect, prior to finding the right cases, but when more guidance is needed in selecting specified attributes. According to features of the data collected in Chapter 7, and taking into account that CBRisk is more about support decision making, this method is used for the calculation of the case matching score.

8.5.2 Attribute Score

Having chosen the appropriate method for calculation of the case matching score, the next step is to decide the attribute score. According to cbr type listed in Table 8.1, the calculation can be classified in two groups: numeric attribute score and textual attribute score.

Numeric Attribute Score

In this section, a variant on the numeric attribute type included with CBR is considered. This attribute type is more complex because it requires more parameters than those provided with CBR. With the standard numeric type, the ratio of the difference between the stored and presented values and a match interval determines the score (see Equation 8.2).

$$Attribute - score = match - \frac{|SV - PV|}{MatchInterval} * (match + mismatch) \qquad \textit{Equation 8.5}$$

where SV is the stored case attribute value, and PV is the presented case attribute value

Figure 8.6 illustrates the relationship of each component in Equation 8.5. There are two match intervals. One is for the case in which the stored value is less than the presented value and the other is for the case in which the stored value is greater than the presented value. There are two thresholds, one corresponding to each match interval. The thresholds determine the value of the difference between the stored and presented values beyond which the score is automatically set to the mis-match score (Brightware, 1999). For instance, if the stored value is greater than the presented value, as the stored value increases the score decreases linearly as a function of the match interval until it reaches the corresponding threshold, at which point it is set to the mismatch score. Thresholds, like match intervals, are expressed in terms of attribute values rather than scores.

For example, for the attribute "Total Project Amount", if the match contribution score is 7; mismatch penalty score is 2; the match interval is 100 million US$; the value of the current case is 200; and the value of the presented case is 300. Then the score of "Total Project Amount" can be calculated as: $7 - \frac{|200-300|}{100} \times (7-2) = 2$.

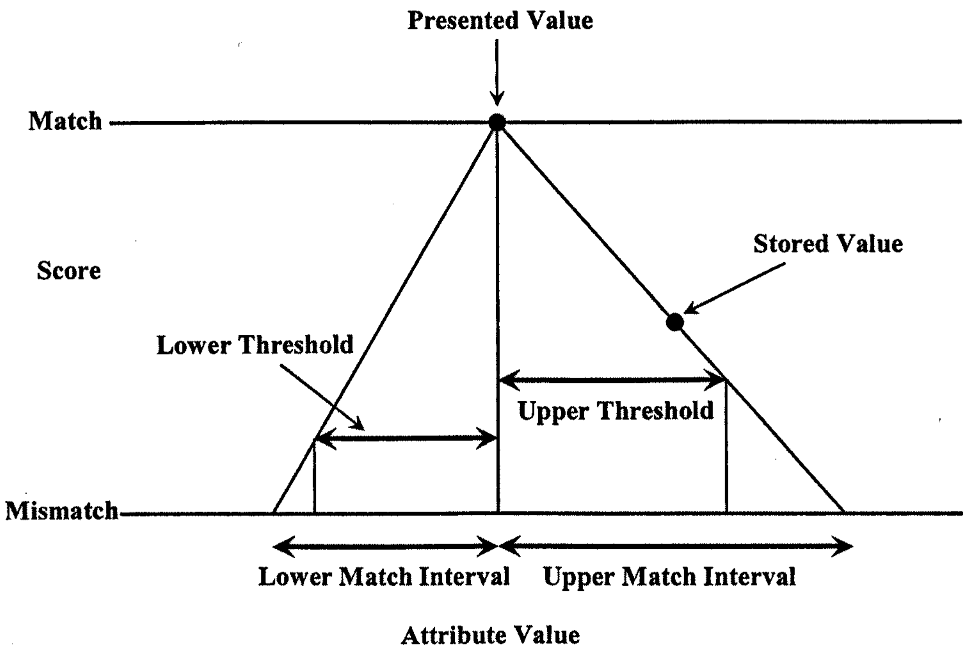

Figure 8.6 Matching scoring for numeric attributes (Brightware, 1999)

Textual Attribute Score

Textual attributes in CBRisk include the attributes with cbr type of symbol, string and words.

If the attributes are defined as cbr:symbol type, the attribute score is judged by comparing the relevant symbol with the one in pretended case. The relevant attribute weighting score is assigned based on the result of the comparison.

If the attributes are defined as cbr:string and cbr: word type, the attribute score is calculated with the percentage of the word exactly matched against the presented case.

If the attributes are defined as cbr:word type, automatical word processing is invoked. A spell checker serves as the basis for word pre-processing in CBR. The spell checker makes use of 10 lexicons to perform spell checks, removal of ignored words, synonym replacement, and suffix removal. In addition, the cbr:word attribute type provides pre-processing of numbers and non-alphabetic elements in text (Brightware, 1999).

Finally, each attribute match score will then be calculated to make up the total case match score and then the top ten match cases will be listed by CBRisk. The final step of case retrieval is to rank the matching cases in a temperate memory and select the most similar cases. Case ranking ranks all retrieved cases according to the accumulated matching scores and prints out the project IDs in order, along with their match scores.

8.6 CASE ADAPTATION

Once a matching case is retrieved, CBRisk attempts to reuse the solution suggested by the retrieved cases. Case adaptation is also related to the parameter value in the case matching stage. Because of the complex nature of construction projects, CBRisk aims to provide decision making support evidence for human experts. The adaptation methods in CBRisk have three options: a statistic based adaptation scheme, the use of the solution of best match cases, and human expert decision.

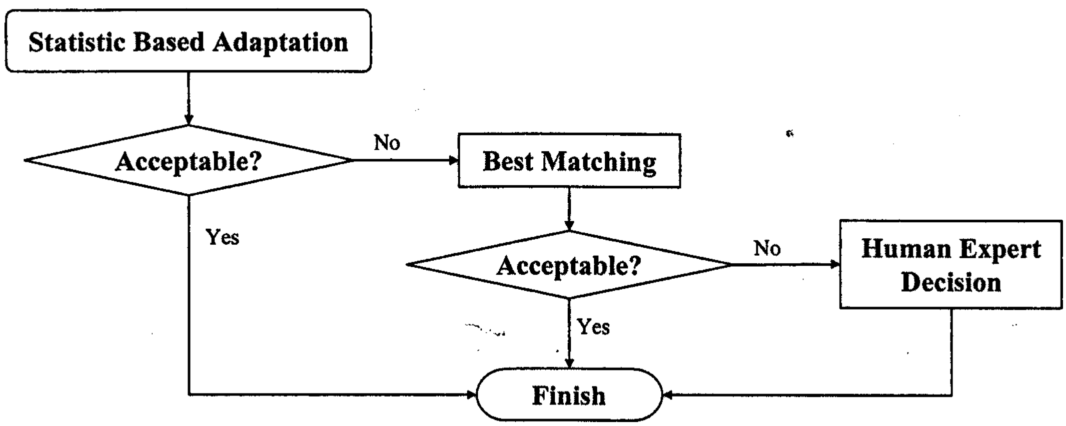

Figure 8.7 The flow chart of CBRisk adaptation

Figure 8.7 demonstrates the case adaptation processes of CBRisk. After case retrieval is completed, the statistic based adaptation is first used to generate the possible risk list for the new project. If the solution is acceptable the case adaptation will finish; otherwise, the risk list of the best matching cases will be used as the possible solution. If this solution is still not acceptable, project experts will be asked to identify potential risks according to their knowledge.

8.6.1 Statistic based adaptation scheme

Similarly with the risk analysis that can be found in the World Bank project appraisal document the case adaptation should include three components: the overall risk level, the potential risk lists and the risk rating for each individual risk item.

To make decisions in project risk management, the type of risk that should be considered is important. Figure 8.8 compares the probability of occurrence of an event with its impact on the construction project. Events with a low impact are not important and can be divided into the elements of trivial and expected. For the high impact and low probability events these are a hazard which could arise but are too remote to be considered (Smith, 1998). The risks with high probability but low impact are expected and usually under control, whilst the risks with high impact but low probability can be considered as hazard. Only risks with both high probability and high impact fall into the focus of risk management.

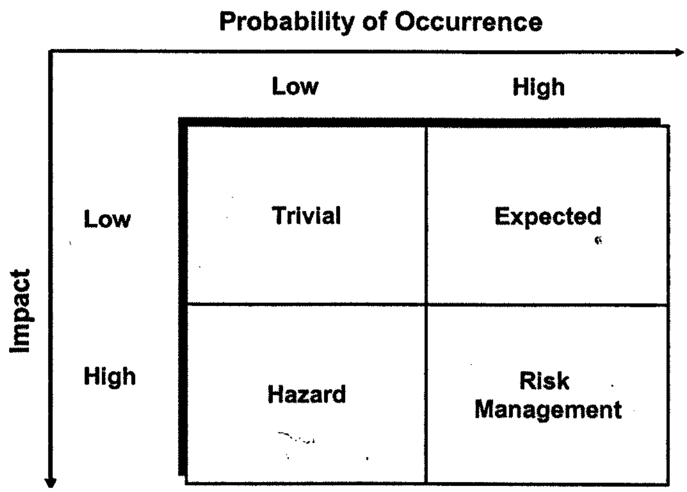

Figure 8.8 Classification of risk souces (Smith, 1998)

In CBRisk, a risk list provides the potential opportunities of risk items; risk rating illustrates the potential impact of a relevant risk item; and the overall risk rating integrates both probability and impact to provide an overview of the potential risk impact. According to the different characteristics of these components, different methods are used.

Overall Risk Level

Based on the projects in the original case base, a risk log is established to list potential risks along with a relevant risk code for each individual risk item. A risk description for each project is allocated with a relevant risk code and risk rating, which is quantified as numerical data as shown in Table 8.2 below.

Table 8.2 Overall risk evaluation list

Risk Level	High	Substantial	Modest	Negligible or Low
Quantified value	4	3	2	1

Similarly, the overall risk level is based on the match score and each risk level,

$$OverallRiskLevel = Round\left(\frac{\sum_{i=1}^{n} RiskLevel_i \times MatchScore_i}{\sum_{i=1}^{n} MatchScore_i}\right) \qquad \textit{Equation 8.6}$$

where, *i* is the individual case, *n* is the total matched cases.

The result will then be rounded to the closest integer value, which is then used to decide the actual overall risk level and is translated back to textual information to be stored in the case base.

Table 8.3 Identify overall risk level based on matching cases

Project ID	Project Name	Risk Rating	Match Score	$Level_i \times Score_i$
P058845	Second Jiangxi Highway Project	2	63	126
P003614	Guangzhou City Center Transport Project	1	56	56
P069852	Wuhan Urban Transport Project	2	54	108
P041890	Liaoning Urban Transport Project	1	51	51
P058844	Third Henan Provincial Highway Project	2	47	94
P056596	Shijiazhuang Urban Transport Project	2	40	80
P058843	Guangxi Highway Project	2	40	80
P070421	Karnataka State Highways Improvement Project	3	38	114
P050036	Anhui Provincial Highway Project	2	37	74
P070459	Inner Mongolia Highway Project	2	35	70
	Overall Risk Level	2	461	853

Table 8.3 demonstrates an example of a matching list against an assumption project. According to the risk rating and match score of each match project, the overall risk level is calculated, using Equation 8.3, as modest.

Risk Items

In statistic based adaptation, the system adapts identified risks of matching cases based on the statistic analysis of the individual risks of matching cases and generates a suggested risk list. The identified risks of projects in the case base was evaluated and picked up to form a risk list. Each risk item in this list is assigned with a risk code. According to the case match score and the risk items in the matched cases, the system will calculate the risk item score, and then compare it with a particular parameter (P) to decide if the risk item will be chosen for the output.

All risk items listed in the top ten matched cases are then used to calculate the risk item score based on Equation 8.7.

$$RiskItemScore = \frac{\sum_{i=1}^{n} ItemScore_i \times MatchScore_i}{\sum_{i=1}^{n} MatchScore_i} \begin{cases} \geq P, Choose \\ < P, notChoose \end{cases} \qquad \textit{Equation 8.7}$$

where, if the Risk Item is listed in the matched case, then $ItemScore_i$ =1, otherwise, $Item_i$=0; n is total matched cases.

To decide whether a risk item is chosen, it is necessary to consider its potential for inclusion in the overall top 10 matching cases. The individual risk items considered for the top 10 matching cases should be more than the average risk item frequency. Suppose the average item frequency is N_F, and the risk items in a risk log is N, the P value can be calculated with the Equation 8.8 below.

$$P = \frac{N_F}{N} \qquad \textit{Equation 8.8}$$

To find out how many risk items over there are the average frequency, it is necessary to find out the average item frequency. The average risk frequency (F) can be calculated by dividing number of risk items in the risk base with the number of risk items listed in the risk log, as shown in Equation 8.9.

$$F = \frac{TotalRiskItems}{NumberofRiskItems} \qquad \textit{Equation 8.9}$$

As mentioned in Chapter 7, the risk log contains 126 risk items which is generated from 625 risk items in 60 projects. According to Equation 8.8, the risk item frequency can be worked out as $F = \frac{625}{126} = 4.96$.

Subsequently, a frequency analysis of risk items in a risk base is carried out. The result shows 44 risk items over 4.96. Consequently, the parameter P can be decided as $P = \frac{44}{126} = 0.35$.

Risk Rating

Having decided which risk item is chosen, the next step is to find out the potential impact of these risk factors. From P value, only 35% risk items reach the average risk item frequency. Therefore, unlike an overall risk level value assigned to each project to calculate the rating for each individual risk item only the projects containing such risk items are considered.

$$RiskRating = Round\left(\frac{\sum_{i=1}^{n} RiskRating_i \times MatchScore_i}{\sum_{i=1}^{n} MatchScore_i} \right) \qquad \textit{Equation 8.10}$$

Where, *i* is the individual project that contains relevant risk item, and *n* is the number of projects that contain such a risk item.

However, the risk rating calculated from this equation will be slightly higher than is expected. Because, in real practice, the overall risk rating is based on each individual risk item rating, it is reasonable to use the generated overall risk rating to control the individual risk rating. Here the overall risk level is used as the reference value for each individual risk item. If the rating of individual risk items is higher than overall risk level, the relevant rating will reduce one level; otherwise it will remain the same, as illustrated in Equation 8.11.

$$RiskRating_i = \begin{cases} RiskRating_i, if \leq OverallRiskLevel \\ RiskRating_i - 1, if > OverallRiskLevel \end{cases} \qquad \textit{Equation 8.11}$$

Table 8.4 Statistic based risk identification result

Risk Code	Risk Item	Risk Rating	P
158	Staff will not use new skills	2	1.06
178	Bus operators do not make enough use of communication and information systems and priority measures to improve services.	2	1.06
179	Weak traffic regulation enforcement	2	0.99
203	Vehicle emissions control strategy is not enforced	2	0.87
206	Political pressures interfere with implementation and adversely affect the quality of construction	2	0.79
146	Delays in necessary land acquisition /site readiness	2	0.77
132	Cooperation between the various stakeholders of the project does not continue	2	0.7
205	Lack of local government support for proposed works	1	0.67
131	Effective coordination is not established among project agencies	2	0.65
176	Future congestion is not mitigated by demand management measures	2	0.6
140	Quality of construction will not comply with required standards	2	0.58
157	Training program irrelevant to needs and objectives	2	0.55
147	Delays in resettlement	2	0.52
217	By not implementing some of the agreed policy prescriptions and enforce regulations, Government contracting policies not adjusted or improved	2	0.51
156	Trained staff lack the opportunity to apply acquired skills in the workplace	1	0.47
109	Insufficient accountability of the use of internal and external resources	2	0.41
143	Deficient engineering design	2	0.41

Table 8.4 lists the identified risk items, relevant risk rating and the control parameter P for item selection. The items with a P value of over 0.40 are selected as potential risks to support decision making.

8.6.2 Best Matching Adaptation

The alternative simple adaptation options include the best matching adaptation and the human experts' own solution. If the new project is not very similar to any case in the case base, even the best matching case might have a low matching score and cannot truly represent the situation of the current case, the solution of best matching case may not be suitable for this particular project.

Figure 8.9 demonstrates the process of the best matching adaptation. In brief, if the case matching score is between method chosen control parameter V and best matching parameter Vc, the adaptation is carried out similar to statistic adaptation, but the mast list is reduced to the cases with the matching score over V. If the matching score is less than V, then decisions from human experts are required. If the matching score is over Vc, the risk list in the relevant case can be borrowed as the identification result for the current case.

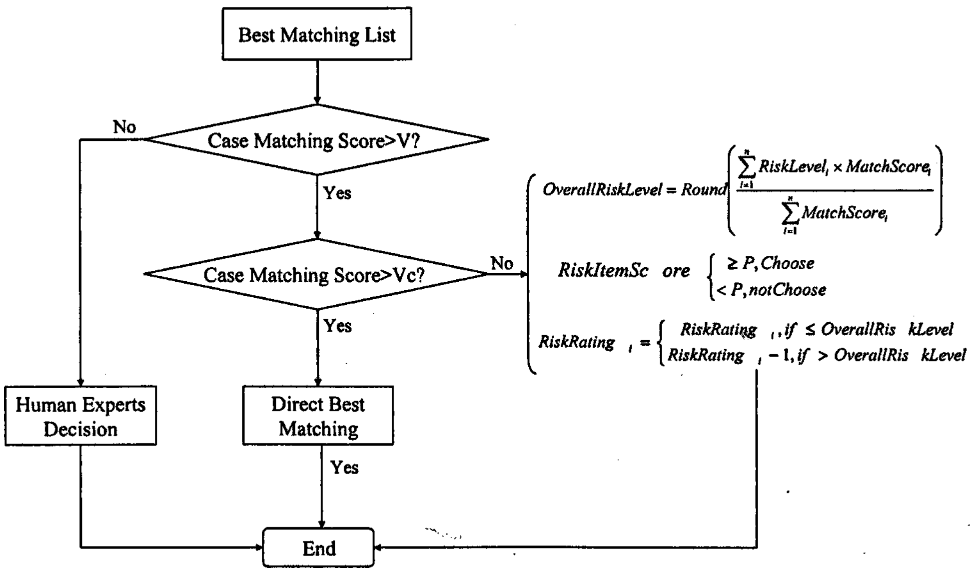

Figure 8.9 Best matching adaptation process

CBRisk again provides two options for the user, either the user goes through the adaptation process according to the flow chart (see Figure 8.7), or the system will reference a matching control parameter (V) to carry out the above process automatically, as illustrated in Equation 8.12 below.

$$MatchingScore\begin{cases} >= V, BestMatching \\ < V, HumanExperts \end{cases} \qquad \text{Equation 8.12}$$

where, *V* is the Matching control parameter.

If the case matching scores more than *V*, CBRisk will use the best matching case for adaptation. Otherwise, project team's experts will make their own decision for case adaptation. The final decision, along with the case description, will remain in the case base as a new piece of knowledge for future use.

Control Parameter V

The value V is decided according to the case matching weight. To use the findings of a matching result, there should at least 50% common features. Suppose the remaining 50% features are either different or absent, then the value V can be calculated using Equation 8.13 below.

$$V = 50\% \sum_{i=1}^{n} MatchContibuteScore_i - 25\% \sum_{i=1}^{n} MismatchingPenaltySore_i - 25\% \sum_{i=1}^{n} AbsentPenaltyScore_i$$
$$\Rightarrow V = \frac{1}{2} \sum_{i=1}^{n} \left(MatchContibuteScore_i - \frac{1}{2} \left(MismacthPenaltyScore_i + AbsentPenaltyScore_i \right) \right)$$

Equation 8.13

where *i* is the attribute, n is the number of matching attributes.

To use the value of the current case base in Equation 8.13, the present V value for CBRisk can be worked out as 51.5. Consequently, only when the case matching score is over 52, can be best matching adaptation can be recommended. Otherwise, the adaptation result from best matching is not accurate enough.

By using value V, the best matching cases are chosen for a new round of statistic calculation based on Equation 8.6-8.11 and a new risk list, risk rating and overall risk level are worked out.

Best Matching Parameter Vc

To use individual best matching cases to develop adaptation results, the author recommends a more conservative parameter Vc, with at least 75% features in common. In this case, similar with Equation 8.13, the Vc can be calculated using Equation 8.14 below.

$$Vc = 75\% \sum_{i=1}^{n} MatchContibuteScore_i - 25\% \left(\sum_{i=1}^{n} \frac{1}{2} (MismatchingPenaltySore_i - \sum_{i=1}^{n} AbsentPenaltyScore_i) \right.$$
$$\Rightarrow Vc = \frac{3}{4} \sum_{i=1}^{n} MatchContibuteScore_i - \frac{1}{4} \sum_{i=1}^{n} \left(\frac{1}{2} \left(MismacthPenaltyScore_i + AbsentPenaltyScore_i \right) \right)$$

Equation 8.14

where *i* is the attribute, n is the number of matching attributes.

According to Equation 8.14, the Vc value can be worked out as 100.75 for the current CBRisk model. When the matching score is over 101, the best matching adaptation is more confident. If there are any matching cases with matching scores of over 101, the risk identification result can be used as the adaptation result. If there is more than one case with matching scores of over 101, as theoretically the higher the better, the model will output the solution of the highest matching project. However, users can review both cases and change the solution if they prefer other ones.

8.6.3 Adaptation with Expert Opinion

Adaptation with expert opinion is used, when both the results from statistic adaptation or best matching adaptation are not satisfactory, especially when statistic matching was failed and there is no good match case. In this case, experts are asked to input the risk items and relevant risk level within their knowledge. However, as discussed in the literature review, this kind of adaptation is usually restricted by human behaviour in terms of subjective judgement, intuition and their relevant knowledge and experience. Consequently a careful verification should be carried out before the new case, adapted by experts, is saved in the case base. Otherwise, further decisions based on inaccuracies in the case base will lead to a vicious circle and damage the CBRisk model.

The final decision, along with the case description, will remain in the case-base as a new piece of knowledge for the future use.

8.7 SUMMARY

This chapter outlines the process of constructing the intelligent risk identification model: CBRisk. Firstly, the architecture of CBRisk is explained. Criteria formulization is demonstrated and databases developed, with an explanation of the process of data mapping. This chapter especially concentrates on the CBR application module. Case indexing to build the structure for the instances of the case base and case representation construct these instances to build the case base. The method of calculating attribute scores and case matching scores to decide the similarity of each case is explained in the content of case retrieval. Because ART*Enterprise does not integrate with adaptation functionality, a systematic adaptation process is developed. It incorporates three adaptation stages: statistic, best matching and human expert adaptation, to provide a flexible output. The graphic user interface module, including the input and output of the model and a variety of functionality screens, will be detailed in Chapter 10.

CHAPTER 9

VERIFICATION AND VALIDATION

9.0 INTRODUCTION

This chapter introduces the process of verification and validation (V&V), highlighting its desirability and outlining methods for ensuring a sufficient level of confidence. The process of verification and validation as applied to this thesis is discussed following its application from the derivation of the prototype model. The methods adopted are integrated with the development of the model ensuring that verification and validation is continuous from conception to completion. The model is verified from the collected data, case completeness to case representation. After the system is verified, model validation is carried out at the micro level. Further validation content also covers the robustness of the system and user friendliness. The various stages involved in the design and development of the CBRisk system are independently verified and validated by experts and project risk professionals.

9.1 WHAT IS VERIFICATION AND VALIDATION?

Verifying and validating play important roles in the development and implementation of a case-based reasoning system. Verification tests whether the system is built right and validation tests if the right system was built (Watson, 1997, O'Leary, 1993, Ng and Smith, 1996). Thus verification examines issues such as ensuring that the knowledge in the system is represented correctly, while validation examines procedures to ensure the system makes correct decisions.

9.1.1 Verification

Verification is defined by Davis (1992) as the process of ensuring that the conceptual model design has been transformed into a computer model with sufficient accuracy. Carson (1986) defines validation as the process of ensuring that the model is sufficiently accurate for the purpose at hand. There are various other definitions such as Shi (2002) who quotes Zeigler (1984) as stating that verification is a process to assure the simulation model is properly realised, whilst validation is a process to assess the degree to which the simulation model's input-output relations map onto those of the system. In simple terms verification is checking that the model performs as intended, whilst validation ensures that the model built is an accurate representation of the system under study.

In terms of verification of CBR models, the author prefers to use the definition provided by Adrion, *et al.* (1982) as the demonstration of the consistency, completeness and correctness of the system. This definition is supplemented to include redundancy to provide greater specificity to the notion of completeness. Verification's dependence upon software indicates that the specific nature of case-based systems needs to be elicited to perform verification. Because verification is system based, that is, for case-based systems verification is concerned with exploiting the software representations of, for example, cases and relationships between cases to establish tests for consistency, completeness, correctness, and redundancy. Consistency in case-based systems refers to parallel implementation of parallel structures whether those structures are words or relations between cases, such as trees. Completeness is concerned with the possibility that knowledge or cases are omitted. Correctness refers to determination of whether or not there are any ascertainable errors in the knowledge. Redundancy addresses the issue of duplication of knowledge, for example, duplicate versions of the same case. Verification also can be dependent upon the domain. By exploiting knowledge of the domain, the system developer can verify the knowledge in a case-based system (O'Leary, 1993).

9.1.2 Validation

According to Ng and Smith (1998) validation consists of demonstrating and confirmating that the CBRisk has been designed in a reasonable and logical manner and represents a relative solution. Shi (2002) notes that the complexity involved with modelling and in experimenting with the model greatly increases the chances of getting invalid results that do not typify the system being modelled. Pidd (1998) characterises one of these errors as a type zero error where the modeller asks the wrong questions so that the model does the totally wrong things or the model does not operate in the manner in which it is intended. In order to minimise this, V&V must be implemented to ensure, as much as is possible, that the simulation model is free of such errors and does actually characterise the real system being simulated.

Adrion *et al.* (1982) indicate that validation is the determination of the correctness of the final program or software produced from a development project with respect to the user needs and requirements. In many projects of system development, the needs and requirements can be established as a priority. However, case-based systems are used in situations where the problem is not well structured enough to develop a rule-based system. Thus, for case-based systems there are likely to be few situations where specifications can be elicited a priority. As a result, validation may have to employ different bases of comparison rather than requirements. The process of validation for case-based systems may be similar to the generation of the needs and requirements for other artificial intelligence (AI) systems. For example, validation may take the form of comparing an expert to the system for different test cases. Alternatively, there may be other approaches that could exploit the unique characteristics of case-based systems to generate comparison bases. These approaches could be cost beneficial by limiting the use of experts in validation processes. The development of alternative models of comparison is a primary focus of this article.

9.1.3 The Role of Verification and Validation

Robinson (1997) concludes that whilst verification and validation should be rigorously applied to models it is not possible for the process to arrive at absolute validity. Resinovie *et al.* (1997) concur with this: the process cannot be assumed to result in the perfect model as the perfect model would be the real system itself. Verification and validation therefore sets out to prove that a model is, in fact, incorrect but by showing that the model is not incorrect under different circumstances, there is increased confidence in the model and its results. The more the verification and validation tests are unable to show that the model "fails" or is incorrect, then the greater the extent to which confidence can be attached to the model.

If the system is not verified then there may be errors in the case representations. If the system is not validated, then it may not make quality decisions (O'Leary, 1993).

The processes for verification and validation are often classified as either a white-box or black-box process (Pidd, 1998). The white-box processes refer to validating internal components. Black-box processes are the opposite, "blacking out" the internal components, and verifying and validating the model as a whole. Black-box validation is the only process that requires a completed model. Verification established that the components of, and the model as a whole, function as is intended and sufficiently represent real world elements. Validation is a black-box process at macro level, determining that the model represents the simulated system. Sargent's (1996) summary indicates that the verification and validation procedure is tied with the development of the model and Nayani and Mollaghesemi (1998) concede that integration of verification and validation with the model development is crucial.

9.2 WHAT TO EVELUATE FOR CBRISK?

9.2.1 Some Unique Features of CBR System

This section draws out some of the unique aspects of case-based systems and uses them as a basis for eliciting different approaches to, and concerns, of verification and validation. Based on the comparison in completion in Chapter 4of CBR and other AI techniques, , the unique features of the CBR system can be summarized as follows.

- According to Ashley and Rissland (1988), case-based reasoning is used to capture expertise in domains where rules are ill-defined, incomplete or inconsistent. It is difficult to establish expectation or standards of behaviour in a case-based system (O'Leary, 1993).
- Unlike other knowledge bases that are static, case-based memory is dynamic as previous problems and their solutions became part of their experience (O'Leary, 1993, Waston, 1997). When a CBR system makes a mistake, the error will usually result in applying an ineffective solution to the current problem. This error may be due to a small case library (insufficient representation of the problem domain), a faulty retrieval mechanism (either does not retrieve the

most similar case or cannot discern that a case is too dissimilar for adaptation to be effective), or a faulty adaptation mechanism (Avelino *et al.,* 1998).

- Because cases typically are represented using frames rather than rules, this indicates that verification approaches for case-based structures could exploit the frame structure.

- Generally, the larger the case library, the more robust it is. Although case-based systems are often designed to create new solutions, typically the quality of the solutions is dependent, in large measure, upon the case base. The quality of the recommendations of a case based system is dependent upon the quality and quantity of the cases in the case base. In general, a case-based system will be able to generate a better recommendation if it has a larger rather than smaller case base (e.g., Ruby and Kibler, 1988; Goodman, 1989; Gaines, 1991).

- Similarity measurement, retrieval and indexing problems are the fundamental issues in a CBR system. It is likely that with a larger base of experience the situations faced by the system will have been seen before and thus a case will be available to assist in the solution of the problem. Even if no identical situation has been seen before, the system is likely to have solutions that are so-called 'near miss' situations (O'Leary, 1993). If the cases are highly correlated, then the system will be limited in the diversity of solutions it can generate. Thus, Bradtke and L-ehnert (1988) find that the most dramatic factor influencing the effectiveness of a case base is the number of unique problem states underlying the case base encoding. Some of these issues can be addressed using approaches such as the direct analysis of the cases by experts, comparison of expert solutions to the system, and investigation of the system reasoning (choice of cases) by experts.

- Case adaptation and case library updating reflect the learning ability of case-based systems.

Case-based systems are different from other types of system, thus a unique methodology for the verification and validation of CBR is desirable in order to analyze the capabilities and behaviour of these systems, and ensure that they are correct when turned over to the eventual users. Our work focuses in developing a systematic approach to sufficiently evaluate the existing case-based risk identification model.

9.2.2 Verification and Validation for CBR system

CBR is a simple technique with a lot of intuitive appeal but also with a cognitive basis. Verifying and validating are accomplished by comparing what is expected with what is present. In verification, the basis for these comparisons is the knowledge representation and the knowledge stored in those representations. Typically the case knowledge is represented using, for example, frames. Thus, these comparisons may include, for example, the structure of the cases (e.g., number of slots and number of slots filled with meaningful information), the structure of the interaction of the cases (e.g., frames typically use tree structures), and statistically based expectations (e.g., distribution of various case parameters) (O'Leary, 1993). This chapter uses statistical and similarity measures to examine the quality and diversity of the case base.

Learning mechanisms can impact on the performance of a case-based system and thus on validation results (O'Leary, 1993). As a result of those learning processes, typically, case-based systems add cases to the case base as the system learns. Therefore, an important issue in validating case-based systems is the impact on the future system performance of adding cases from current experience to the case base for future use. Critical questions include, does comparative performance change based upon the order in which cases are added to the case base? In validation, often human experts are the comparative basis. However, experts can be expensive or unavailable. One of the focuses of this chapter is the generation of alternative sources for comparison. Mathematical programming and statistics are developed as sources of comparative solutions. Examination of the cases in the case base is a critical concern of the quality of the decisions and, thus, to validation.

9.2.3 Verification and Validation Framework for CBRisk

According to Kleijnen (1995), when embarking on the verification and validation of the actual model it is necessary to carry out the verification before validation. This is in agreement with the simple definitions given in section 9.1 that indicate that the components of a CBRisk model and the model as a whole would have to be shown to be functioning correctly before the model could be tested to show that it addressed the purpose for which it was created. Other studies reviewed for this thesis have also adopted this approach to verification and validation (Nayani and Mollaghasemi 1998, Ng and Smith 1998).

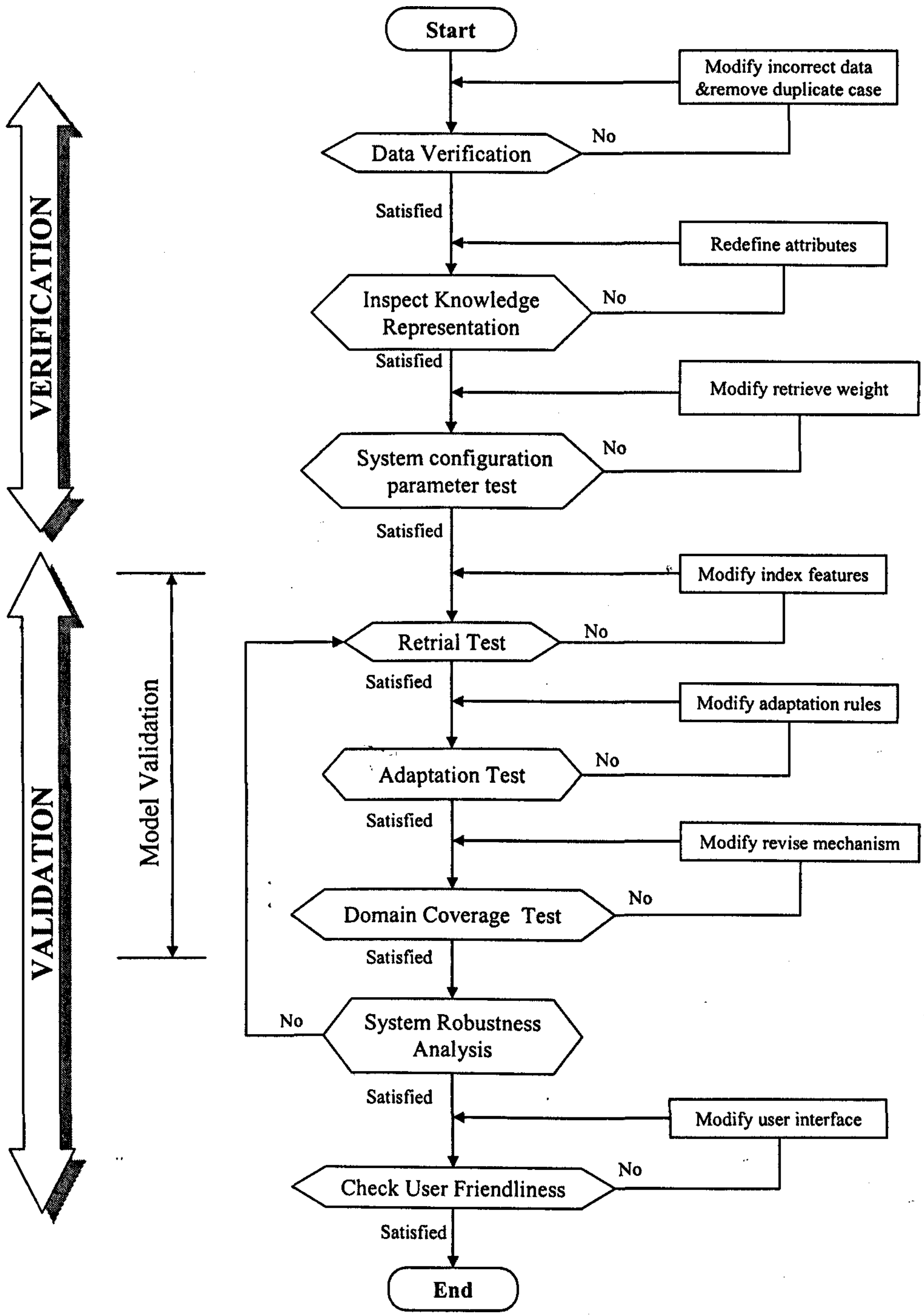

Figure 9.1 Verification &Validation Process of CBRisk

Figure 9.1 demonstrates the systematic processes in carrying out verification and validation. The verification is firstly carried out by three processes: Data verification: verification of the knowledge stored in the case base and inspection of representativeness; verification of the frame to acquire knowledge, and system configuration parameters test; and verification of the retrieval principles. Following verification, the validation is performed also through three processes: model validation focused on examining the internal CBR model from case retrieve, case adaptation and domain coverage; system robustness analysis; and user friendliness. If any of these processes do not pass the test, the relevant modification will be made until a satisfactory result is reached.

9.3 VERIFICATION

Verification tests if the system was built right. The verification could be divided into two aspects: verification by the system designer and verification by the domain experts. Verification by the system designer based on statistical analysis should then be compared with the domain experts' opinion. Based on the V&V processes in Figure 9.1, the following section explains how verification in each individual process is carried out.

9.3.1 Data verification

A model of a project is only as good as the data that is input, so if the input data is inaccurate, the model will not accurately reflect the project (Smith, 1998). Collected raw data is used to construct every piece of knowledge in the CBRisk model. As discussed in section 9.2.1, the quality of the case base can influence the case retrieval process, as well as model learning ability. As a result, data verification should be carried out at the beginning. Data verification includes two aspects: correctness and redundancy.

Correctness

Errors in consistency can occur for many reasons. For example, mistakes were made by misspelling or using different names for the same object or activity, and wrong numbers of value.

The correctness of a case based is measured by how often the case that is retrieved is the case in the case base that answers the query most effectively. Due to the heuristic nature of case bases, this is a difficult criterion to measure (Racine and Yang, 1996).

To check the correctness, the attributes were divided into two groups: those with numerical values and those with textual information (see Table 9.1).

Table 9.1 Attribute classification and analysis methods

Attribute	Type	Analysis	Attribute	Type	Analysis
% of Transportation	N	Box Plot	Region	T	Frequency
Grace Period	N	Box Plot	Country	T	Frequency
Year	N	Box Plot	Program of Targeted Intervention	T	Frequency
Year to Maturity	N	Box Plot	Proposed Terms	T	Frequency
Project Approval Year	N	Box Plot	Product Line	T	Frequency
Implementation Period	N	Box Plot	Lending Instrument	T	Frequency
Land Acquisition	N	Box Plot	Economic Evaluation Method	T	Frequency
Resettlement of People	N	Box Plot	Borrower	T	Frequency
Grant Amount	N	Box Plot	Front End Fee of Bank Loan	T	Frequency
FP_Goverment	N	Box Plot	Service Charge	T	Frequency
FP_other	N	Box Plot	Environment Category	T	Frequency
IBRD Commitment	N	Box Plot	Commitment Fee	T	Frequency
IDA commitment	N	Box Plot			
Total Project Cost	N	Box Plot			
Financial_NPV	N	Box Plot	Approval Date	T	No
Financial_FIRR	N	Box Plot	Effective Date	T	No
Economic_NPV	N	Box Plot	Implementation Agency	T	No
Economic_EIRR	N	Box Plot	Target Thematic Outcomes	T	No

Attributes with numerical values are presented on a box plot analysis, through this analysis significant errors of input can be found. For instance, Figure 9.2 shows the Box Plot result of FIRR, most of the values are within 1, but there is an individual value significantly outside the normal range. When comparing the data base and original document, an error was found that mis-input 8.3 as 8.3%. After modifying the data base, the new Box Plot is illustrated in Figure 9.3, which is more reasonable. This analysis was applied in all numerical attributes and further modifications were made.

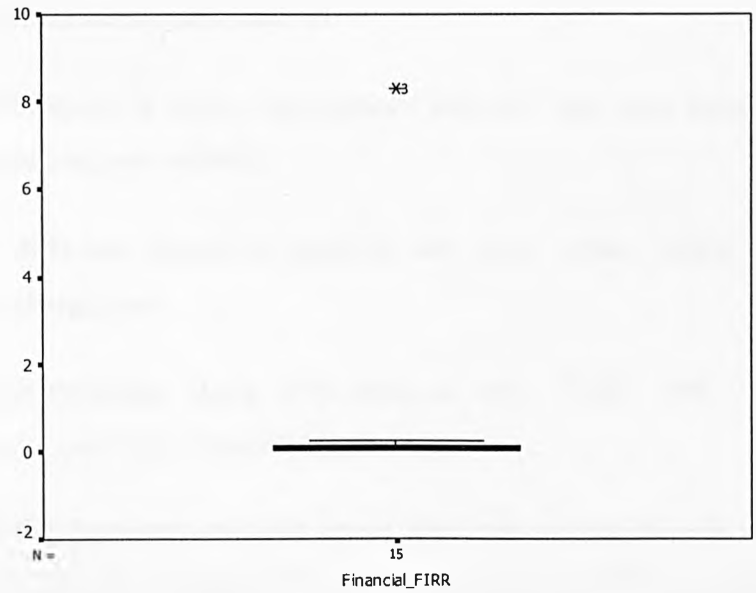

Figure 9.2 Box plot of FIRR before checking for data correctness

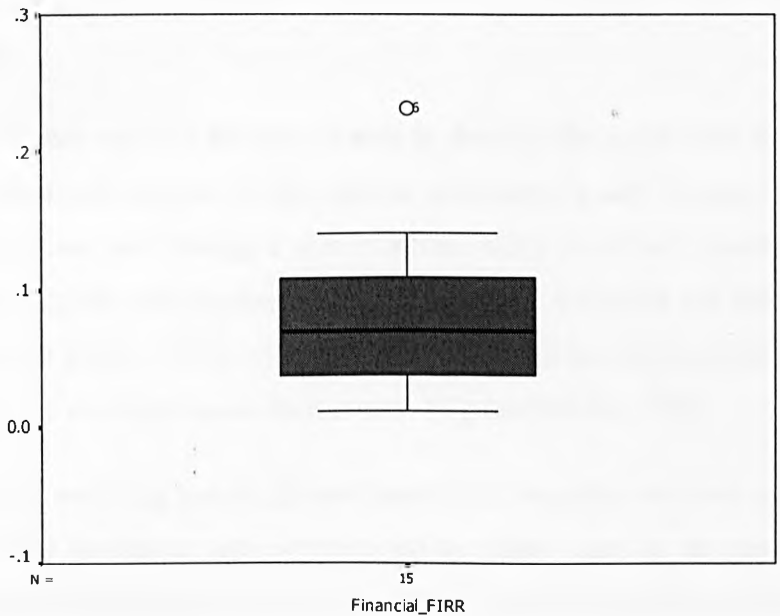

Figure 9.3 Box plot of FIRR after checking for data correctness

For attributes with non-numerical value, a frequency analysis is carried out for each individual value, except the four attributes that do not usually share any common value and do not contribute to case retrieved. They are approval date, effective date, implementation agency, and target thematic outcomes.

Through the frequency analysis the errors of misspelling and mis-input are obviously highlighted. Errors were made, such as:

- misuse of space or coma. "cost benefit analysis" and "cost-benefit analysis:" in economic analysis method;
- use of different format to represent the same value. "n/a%" and "N/A" of service charge; and
- using abbreviation along with original text. "VSL" and "Variable-spread Currency Loan" for proposed terms.

According to the frequency analysis report, the input entries of each error's attributes are unified. In order to eliminate the same error occurring in later applications, a drop down list filed with all possible values of relevant attribute was used for each non-numerical data entry.

Redundancy

Redundancy errors occur if the user is able to develop the same case more than once. This would normally happen in the system maintenance and validation stage, when starting a new case and finding a new solution which is saved it more than once, or only changed slightly and another match and solution is carried out and then saved in the case base (O'Leary, 1993). A case is redundant if it succeeds in the same situation and has the same conclusions as another case (Ng and Smith, 1998).

Due to the ever evolving nature of case bases, it is important to have a mechanism to determine if the incoming case is subsumed by other cases in the case base or if it subsumes existing cases in the case base. If two or more cases in a case base are very similar and are retrieved for the same set of queries, it is unnecessary to keep them all in the case base as it may degrade the efficiency of the case reasoner (O'Leary, 1993).

Although redundancy does not necessarily cause logical problems (Ng and Smith, 1998) or harm decision making (Waston, 1997), it might affect the efficiency of the system and cause administrative problems. Redundant cases can cause difficulties, especially in maintenance situations. One version of the case could be revised while another is not. This could cause confusion and possibly errors if the unrevised case was used by the system.

The possibilities of redundant cases can be mitigated if the user is required to establish a list of the specific cases prior to actually establishing the data within each case. As new cases are added, they would be added to the list of feasible cases. Each case would then be referenced by its specific name. Thus, a single version of any given case would be permitted.

Redundancy errors also occur if the user is able to enter the same attribute information into two or more attribute spaces for a given case. There are two basic situations: two attribute instances in the same attribute slots and the same two attribute instances in different slots in the same case. In the first instances, maintenance of one entry but not the other may cause ambiguity as to the proper contents. This could be eliminated if each attribute slot has space for only one entry. In the second situation, the attribute instance may be an inappropriate occurrence for that attribute instance. This could be eliminated if the attribute instances could only come from eligible lists, as discussed earlier (O'Leary, 1993).

9.3.2 Knowledge Representation Test

To acquire knowledge for the CBRisk model, the raw data should be allocated to a knowledge representation frame. The frame is made up of several attributes to describe a project. The inspection of representativeness aims to evaluate whether the knowledge acquired in the case base is satisfied. To achieve this, verification in this sector consists of two components: indices inspection and case coverage tests.

Indices Inspection

The purpose of the indices test is to verify if the attribute can efficiently represent a project and to check the completeness of each attribute.

During the data verification in section 9.3.1, the attributes with missing value have been marked as 'Nil'. At this stage the attributes with 'Nil' value are examined. When the missing value is more than 20%, this means the attributes are not always available. These attributes are then picked for further analysis. Table 9.2 lists 14 attributes that have high missing values: 8 numerical attributes and 6 textual attributes.

Table 9.2 List of attributes with incomplete entry value

Attribute	Missing	Percentage	Attribute	Missing	Percentage
Grace Period	24	40.00%	Finance Source	32	53.33%
Year to Maturity	24	40.00%	Commitment Fee	36	60.00%
Financial NPV	15	25.00%	Front End Fee for Loan	40	66.67%
Financial FIRR	15	25.00%	Service Charge	52	86.67%
Economical EIRR	27	45.00%	Proposed Terms	31	51.67%
Economical NPV	27	45.00%	Evaluation Methods	32	53.33%
Land Acquisition	24	40.00%			
Resettlement People	25	41.67%			

When the missing value is greater than 50%, the attributes are not often used to describe a project. As a result, these attributes might be unimportant attributes, which will be put into a temporary removal list. However, because the case base only has 60 cases so far for cases covering 11 countries, it is hard to make decisions based only on the results from the current data set. As the number of cases increases, the results may well change. As a result the decision of whether or not to remove the attribute from the case retrieval process is made by consulting opinions from domain experts (see section 9.3.3).

Case Coverage Test

Acquiring representative cases and case distribution are very closely linked. Case distribution refers to the coverage of cases across features. This section will consider two aspects of case distribution: feature shift and feature range.

Feature shift demands that cases be clustered more closely around the point of feature shift, particularly if relatively simple structural adaptation techniques are used (Waston, 1997). Case distribution should reflect the features of the solution space, not just the historical precedents of the problem space.

Many similarity functions, particularly those for assessing similarity between numerical features, consider the range of the feature (Waston, 1997). Consequently, during model development it is very important to consider the range of features in the case-base, their mean value, and their standard deviation. If there is one case with an abnormally low or high value, it can distort the accuracy of the similarity measure. Hence, developers should identify isolated cases and consider whether they are required in the case-base. If they are required, they should try and obtain more cases to provide a more even case distribution.

It is better to have an even distribution of cases across the problem area (Ian, 1997). The following lists appropriate approaches to execute a case coverage test:

- For numerical features, their standard deviations were checked. Deviations should be as low as possible. The mean of the feature value should also be near the middle of the feature range. If the standard deviation is high, this means the system has outlying cases and should be identified. More cases should be obtained for a with few cases features are then to improve the coverage for that feature. If the mean is toward one end of the range, this means too many cases at one end and not enough at the other. It will be neccessary to remove some cases in the heavy end, or increase more in the light end.

- The frequency of occurrence of each symbolic feature was checked. If the distribution is very uneven, there will be outlying cases which must be removed or more cases must to be obtained to improve the coverage for the feature.

- For both attributes, there will be a feature shift in the case-base and particular attention will be given to case coverage around the feature shift.

9.3.3 System Configuration Parameter Test

System configuration parameters are in charge of case retrieval, which is the core process in a CBRisk model. These parameters includes: the size of the case library; thresholds for quality match; and the matching weights for each attribute. Many CBR tools allow developers to adjust parameters that will affect the retrieval techniques. A system configuration parameter test is achieved through three processes: expert opinion, the distribution analysis, and the correlation analysis.

Expert Opinion

The weighting given to a feature will affect the degree by which it influences the closeness of a match during retrieval. Consequently, it is important that developers consider how feature weightings are assigned. Very often this seems to be a relatively ad hoc process, with judging by domain experts that one feature is more important than another.

Three groups of experts with different knowledge strength were involved in providing advice in the development of CBRisk. They are experts in: project risk management; artificial intelligence systems' development; and from the World Bank Transportation Sector.

Professor Hanna from the University of Wisconsin-Madison gave recommendations in terms of selection of attributes and possible resources for improving case adaptation. In addition, according to the recent research of the application of CBR in forecasting, the expenditure of residential buildings by Professor Russell and Yi of University of Wisconsin-Madison, the size of the case base was carefully decided.

A questionnaire (see Appendix C) seeking feedback in terms of attribute selection and weighting was sent to experts in the transportation sector of the World Bank. After comparison of the experts' advice, further modification of attribute weight was made and the decision of whether or not in the remove the attribute in removal list mentioned in section 9.3.2 was decided.

Distribution Analysis

A more formal approach to evaluate the retrieval weight is by using statistical techniques to analyze the case data. The distribution analysis of each attribute value is performed.

As demonstrated in Chapter 8.4, when handling case retrieval, there are three types of weight contributing to case matching. Based on the previous analysis in Table 9.3, if the missing value is more than 30%, a value which is not often available, so the absent penalty score is changed to 0. Similarly, for individual attributes, if the missing value is more than 50%, the relevant mismatch score should also be reduced to a certain level.

The distribution of the above attributes is then carried out. Based on this analysis, Table 9.4 explores the range of each attribute and how much the median data can represent the attribute, and how close the median and mean are. The '% Interval' represents the percentage of 95% confidence of medium interval against the whole range of the attribute, where the 'MM Interval' illustrates the percentage of the distance between mean and medium against the range of attribute. The small value of 'Interval%' illustrates that attribute values are centralizing, vice versa. The lager the 'MM Interval%' value is, the bigger the distance between mean and medium will be.

Table 9.3 Data distribution of attributes with incomplete entry value

Attribute	Mean	Median	Range	95% Interval	Interval%	MM Interval %
Grace Period	5.42	5	5	0	0.00%	-8.40%
Year to Maturity	21.46	20	25	0	0.00%	-5.84%
Implement Period	58.53	60	60	0	0.00%	2.45%
IDA commitment	34.03	0	300	0	0.00%	-11.34%
Land Acquisition	16247942	6866667	2.21E+08	8809929	3.99%	-4.25%
% of Transportation	89.80%	94%	0.45	0.0407	9.04%	9.33%
FP_Government	163.59	15.45	1555.5	166.47	10.70%	-9.52%
Total Project Cost	390.49	312.81	1740.6	186.45	10.71%	-4.46%
Grant Amount	194.07	195	608.4	77.9	12.80%	0.15%
Economic_NPV	657.23	488.03	2550	409.05	16.04%	-6.64%
IBRD Commitment	160.05	150	620	116.63	18.81%	-1.62%
Resettlement People	20990	12693	78278	20865	26.65%	-10.60%
Economic_EIRR	26.27%	23%	0.285	0.082	28.77%	-11.47%
Financial_NPV	239	70.4	972.01	306.57	31.54%	-17.35%
Financial_FIRR	7.84%	7%	0.222	0.078	35.14%	-3.78%

In Table 9.3, the bottom four attributes: resettlement people, EIRR, FNPV and FIRR have large 'Interval%' value and fairly large 'MM Interval%' value. These attributes contribute more to differentiation between cases than any others. Consequently, the relevant retrieval rate of these attributes is increased to 30%.

Correlation Analysis

Before analysing the correlation among the attributes it is necessary to understand the data set for each of the attributes. Because the Pearson correlation assumes that the single data is normal distribution, it is necessary to carry out a Q-Q plot for each attribute to check if it is normal distribution. Figures 9.4 and 9.5 provide an example. Figure 9.4 shows the Q-Q plot of Financial FIRR, as we can see the sample data distributed along each side of the Q-Q line with quite similar trends, so we can say FIRR is normal distribution. However, in Figure 9.5, the data set does not fit very well with the Q-Q line, so we conclude FNPV is not normal distribution.

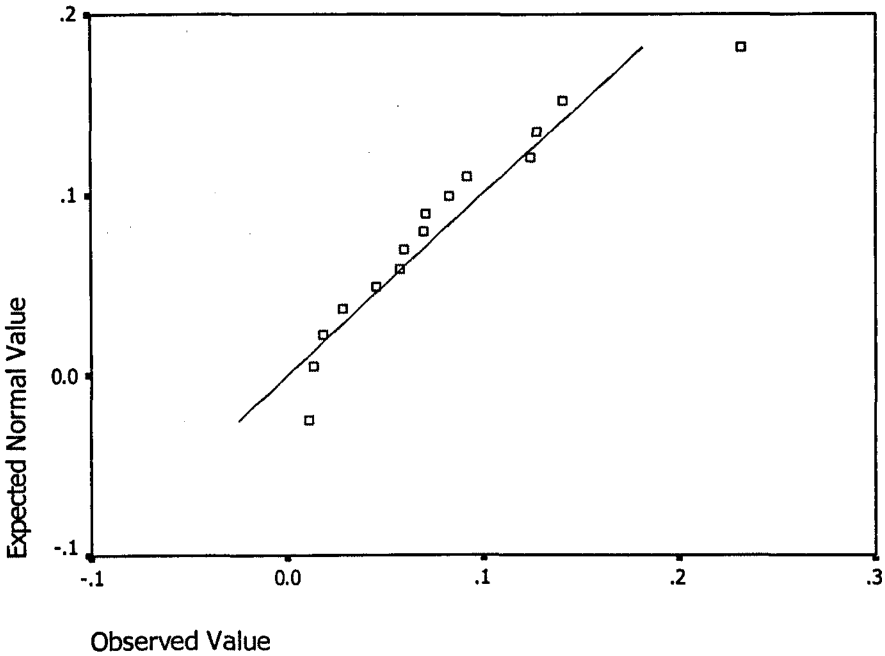

Figure 9.4 Normal Q-Q plot of Financial_FIRR

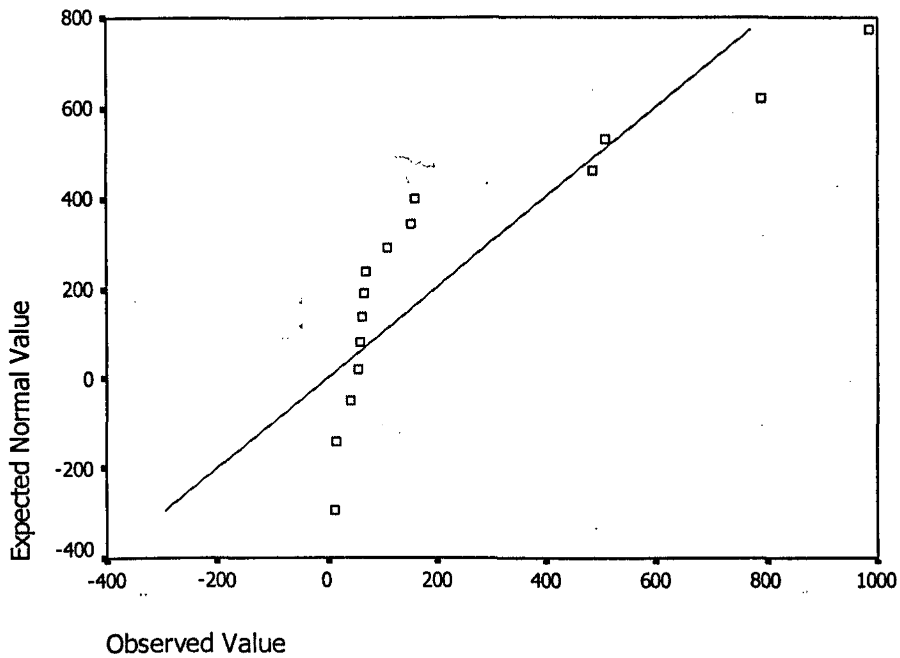

Figure 9.5 Normal Q-Q plot of Financial_NPV Commitment

Such analysis is performed for each attribute and references the Skewness and Kurtosis factors, which illustrate which the distance from normal distribution. In general, a

Skewness/Kurtosis value greater than one indicates a distribution that differs significantly from normal, symmetric distribution. Based on these analysis result, the attributes close to normal distribution are identified. They are: Grace Period, Year of Maturity, Finance Source, Implementation Period, IBRD commitment, IDA Commitment, Government Commitment, FIRR, ENPV, EIRR, Land Acquisition, Resettlement People and Year.

The Bivariate Correlation analysis among the selected attributes is then carried out. The result is listed in Appendix D. For each pair of attribute there are three correlation analysis parameters listed in the table: Pearson correlation coefficients, signification value, and number of cases with non-missing values. The Pearson correlation coefficient is a measure of linear association between two variables. The values of the correlation coefficient range from -1 to 1. The sign of the correlation coefficient indicates the direction of the relationship: positive or negative. The absolute value of the correlation coefficient indicates the strength, with larger absolute values indicating stronger relationships. The significance level (p value) is the probability of obtaining results as extreme as the one observed. If the significance level is less than 0.05, then the correlation is significant and the two variables are linearly related.

By reviewing correlation parameters for a pair of attributes, the correlation among attributes was found. Appendix E provides a table illustrating the correlation among attributes and the relevant level. The three matching weights of each attribute are then further modified, based on the correlation coefficients.

In order to improve case retrieval accuracy, the duplicated attribute contribution value should be removed. To perform this, the relevant attribute weights should be deducted according the correlation level between two correlated attributes. The calculation method is demonstrated below.

$$TotalDuplicatedValue = (Attribute_1 + Attribute_2) \times Pearson \qquad \text{Equation 9.1}$$

$$removalValue = \frac{TotalDuplicatedValue}{2} \qquad \text{Equation 9.2}$$

$$Attribute_1 removalValue = \frac{Attribute_1}{(Attribute_1 + Attribute_2)} \times removalValue \qquad \text{Equation 9.3}$$

$$Attribute_1 ModifiedValue = Round(Attribute_1 - AttributeremovalValue) \quad \text{Equation 9.4}$$

To integrate all the above equation (equation 1 to equation 4), the final modified score can be calculated through the following equation:

$$ModifiedValue_{ij} = Round(OriginalValue_{ij} \times (1 - \frac{Pearson}{2})) \quad \text{Equation 9.5}$$

where i=1, 2, ...n; n is the number of attribute

j=Match Contribution, Mismatch Penalty, Absent Penalty

ModifiedValue represents the new weight for relevant attribute and match score

OriginalValue represents the old weight for relevant attribute and match score

For instance, Table 9.4 lists the match contribution score, mismatch penalty score, absent penalty score and correlation parameters of 'Land Acquisition' and 'Total Project Cost'. As we can see, the correlation value between 'Land Acquisition' and 'Total Project Cost' is 0.518. According to Equation 9.5, the modified weight for each attribute match weights listed in Table 9.4 below.

Table 9.4 An example of modification of attribute weight based on correlation analysis

Status	Attributes	Match Contribution	Mismatch Penalty	Absent Penalty	Pearson Correlation
Original	Land Acquisition	5	2	0	0.518
	Total Project Cost	8	4	3	
Modified	Land Acquisition	4	1	0	
	Total Project Cost	6	3	2	

After the above verification processes the model can perform in a more reliable and confident level. It is ready to carry out system validation.

9.4 VALIDATION

Validation for CBRisk consists of three components: validation for the case-based reasoning model, system robustness analysis, and validation of the system friendliness.

9.4.1 Validation Model

The CBR process includes three basic processes: case retrieval, case adaptation and case library update. These reflect the learning ability of the case-based system. Two criteria will be used to evaluate the validation of the system: result acceptability criteria and system validity criteria. The model validation is carried out through three tests: retrieval test, adaptation test and domain coverage test.

Retrieval test

In a CBR system, if exactly the same search is performed twice, the same source case should be retrieved with the same accuracy. A set of searches are repeated several times to check for consistency. Inconsistency can be very disturbing to users who expect computers to give the same answer to the same question every time.

The retrieval test ensures that the comparison and retrieval function are correctly carried out. The author intends to use all cases in the case base to test the retrieval result and inspect the output. If the top retrieved case is the same case in the case base as was used as the test case, the test is designated as 'successful' and the case is marked to the successful case group. The proportion of the successful and failed tests will be computed. If the rate is greater than the system validity criteria (SVC), the CBR system can be considered to be retrieval valid. Figure 9.6 illustrates the process of this case retrieval test. To pass the case retrieval test, the SVC should be over 90%. All the cases in CBRisk were marked as retrieval success, which means the retrieval success rate is 100%, which is more than SVC. So the CBRisk passed the case retrieval test.

In many circumstances the retrieval time of a case based reasoning system is a critical evaluation parameter. However, in CBRisk, matched cases can be listed out immediately after sending the case match command. The retrieval time is usually less than one second. As a result the retrieval time is not included for consideration in this section.

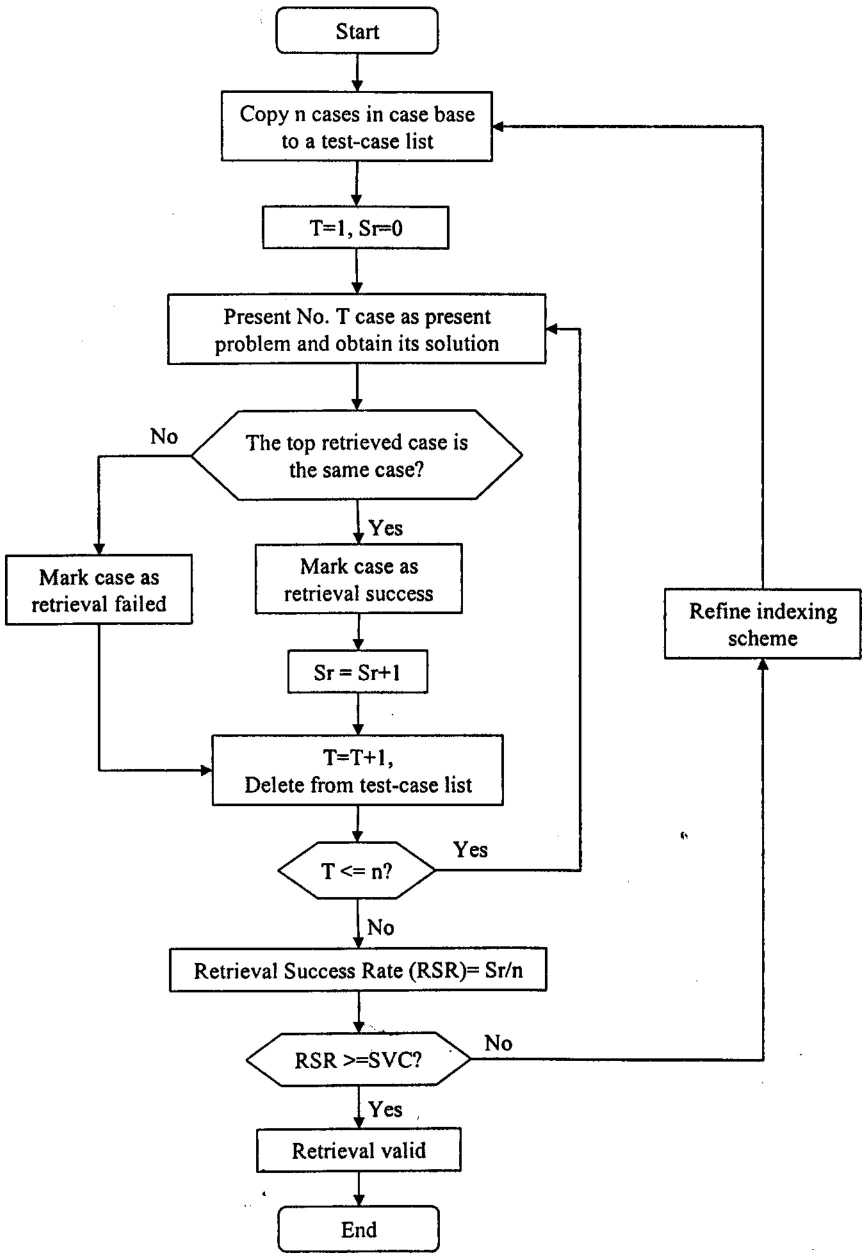

Figure 9.6 Retrieval test flow chart

Domain coverage test

Domain coverage test addresses the issue of how well the system is able to assimilate new cases added to the case library. The cases in case based are classified in groups based on sub-sector, such as highway, rail way, urban transportation etc. The first group is made up of the original case base. Then choose one of the other groups is chosen as the test group; the project description of these cases is input one by one; inspect the solution generated by the system is inspected and compared with the solution by the test cases.

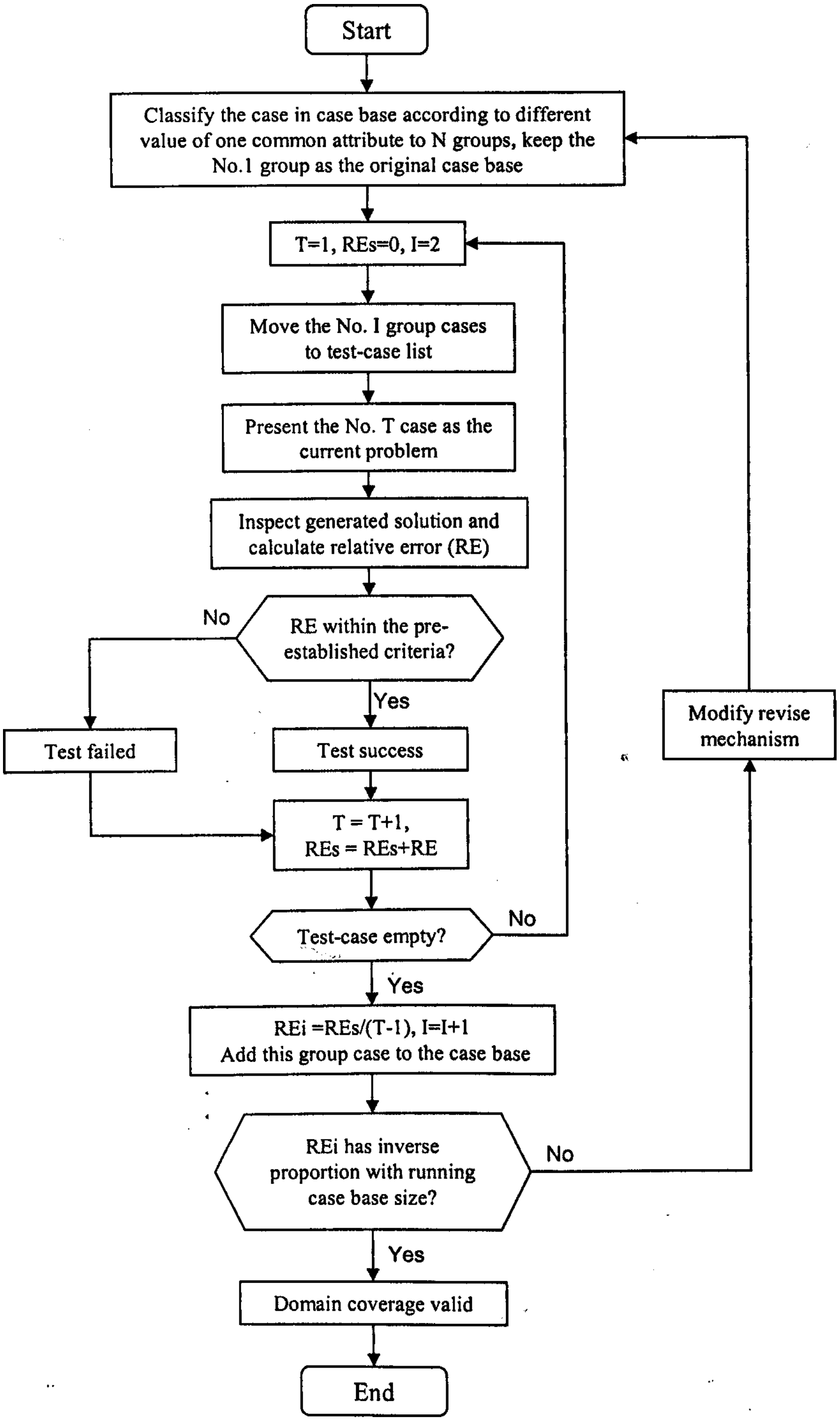

Figure 9.7 Domain coverage test flow chart

The author than calculates the relative error. If the relative error alls within the pre-established criteria, the test is noted as 'Successful'; otherwise the test is noted as

'Failed'. The average relative error of the group cases is computed, and other groups are tested in the same way. Also the test can be classified by other features, for instance the project amount, contact type, financial type etc. to continue the domain coverage test. Figure 9.7 presents the process of this case retrieval test.

According to the data analysis in Chapter 7.5, there are 60 projects from 16 countries, while 10 of them have no more than 2 projects. Furthermore, these projects might be from different sectors, such as highway, railway, or urban-transportation. As a result, the domain coverage test is not successful in some circumstances. At this moment it is not possible to add more suitable cases to improve the domain coverage as the documents for the previous year for World Bank projects are not available. However, the development of IT techniques and the use of electronic documents is gradually increasing within many institutions and companies, as it is within the World Bank. Project information easier to collect will be in future years. The author believes the domain coverage can be improved. Here, the author aims to demonstrate the methodology for carrying out the testing.

Adaptation test

An adaptation test ensures that the adaptation is properly made from valid retrieved cases. It reflects the learning ability of the CBR model and how accurate the system identified risk is. It involves removing the test case from the case-base first, inputting it as a new case, inspecting the solution generated by the system and comparing it to the solution of the test case. Its relative error is then calculated. If it lies within the result acceptability criteria, the test are designated as 'Successful', otherwise the test is designated as 'Failed'. Next input the test case back to the case base and move on to the next case. Repeat this cycle for each case in the case base and compute the proportion of the successful and filed tests as the 'case adaptation criteria' (CAC). If adaptation success rate (ASR) >= system valid criteria (SVC), the adaptation phase is considered to be valid. Figure 9.8 presents the process of this case retrieval test.

Considering the limitation of domain coverage for CBRisk right now, the case adaptation criteria is decided by comparing the relative error with the case matching score. Equation 9.6 below shows the underlining relationships and rules.

$$CaseAdaptationCriteria = \frac{CaseMatchingScore}{relativeError * 100} \qquad \text{Equation 9.6}$$

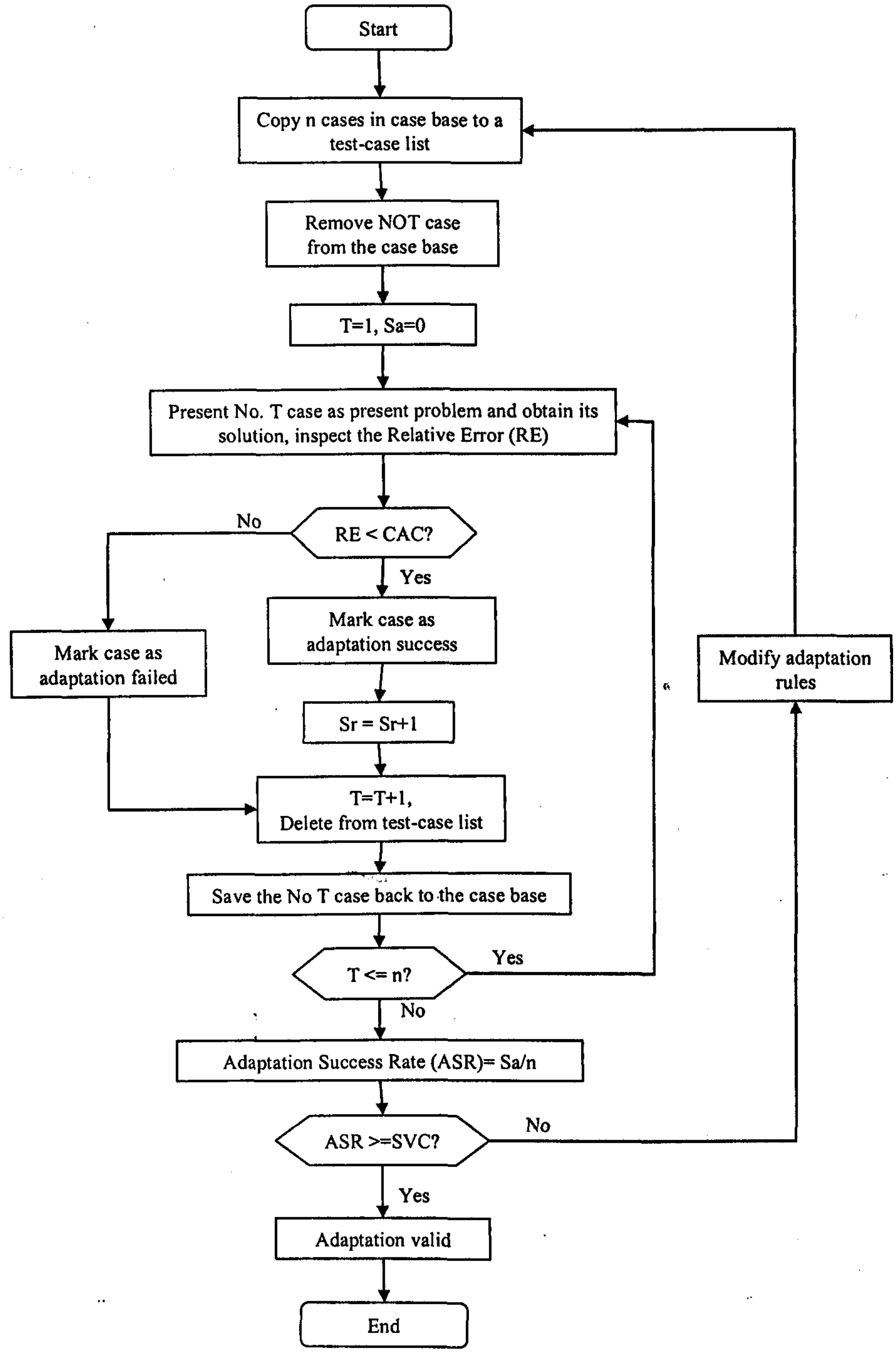

Figure 9.8 Adaptation test flow chart

The case adaptation includes two components: the adaptation of the overall risk level and the risk list. The adaptation of the overall risk level achieves 100% success. Using the risk list adaptation in CBRisk, 57 of 60 projects passed the case adaptation test. The adaptation success rate is 95%, which is more than valid criteria for the system. So CBRisk passed the adaptation test. Later on, as cases in the case base increase, the domain coverage can be improved. Once the model can easily pass the domain coverage test, it is more reasonable to set the case adaptation criteria with another value, for example 80%.

9.4.2 System Robustness Analysis

CBR is a problem solving technique which works by searching through a database of previously-solved problems for one or more cases whose identifying features closely resemble the current problem. When found, the solution employed in the historical cases are retrieved and applied to the current problem. However, if the historical case most closely resembling the current problem is not sufficiently similar, then the CBR system undertakes a process of modifying the corresponding historical solution wherever possible, in order to match the current problem better. This modification is referred to as adaptation, and is a function of the magnitude of the differences between the current problem with the new solution can be pose facto appended to the case library to increase its robustness.

During the robustness analysis phrase, the system is tested under different alternatives for the factors to assess the generality of the result. The model can be used to analyze the sensitivity of the case indices. Sensitivity analysis can identify which features affect the outcome and proportionally what their impact is. The sensitive analysis for CBRisk includes two aspects: the sensitiving of case retrieval and the sensitivity of the case adaptation/system learning. As both statistic analysis adaptation and best matching adaptation are based on the result of the case retrieval, if the matched cases and match score have the reasonable robustness, then the adaptation is also robust. Consequently, the sensitivity of the indices is first analysed.

The selection of indices could affect the accuracy of retrieval especially when the indexing features are sensitive to changes (Ng and Smith, 1998). Kriegsman and Barletta (1993) have suggested that the indices could influence the number of case retrievals. Ng and Smith (1998) states indices might effect the similarity score and hence the priorities of the retrieved cases. As a result, a sensitive analysis was carried out to identify the sensitivity of the indexing features.

A scenario project is created for sensitivity analysis purposes (see Appendix F). The analysis was carried out by gradually changing the best guess index of one feature to a range of ±2 while keeping the indices of other features unchanged. This enables the author to focus on the effects of change of one feature at one time. The adjusted indices were used for case retrieval. The similarity of scores of the retrieved cases was recorded. To enable the effect of change to be compared, variations in scores were transformed into percentage variations of the original scores. The same procedures were then applied to the other indexing features (Ng and Smith, 1998).

The selection of attributes used for sensitive analysis is based on three principles: 1) the match contribution score should no less than 2, otherwise the -1, -2 test is not possible; 2) the 'Nil' value for each individual attribute should not be more than 50%, as if many missing values exist, the analysis can not represent the real problem; 3) for correlated attributes, only some in each pair are chose. Finally 14 attributes (see Figure 9.8) were used to carry out 56 round matching to check the indices' sensitivity.

The result shows that the top 5 matched cases remain the same rank in all 56 round tests. Only 2% cases not remain the same rank and overall 11 cases in total were listed in the top 10. Table 9.5 provides the absolute maximum changing range of 14 attributes at each level for the top 5 cases. It illustrates, that for all the attributes, when individual match weight changes in the [-1, 1] level, the total match score changed no more than 3 and on the [-2,2] level, no more than 4. This illustrates that the system is very robust in terms of matching and ranking.

Table 9.5 Absolute sensitivity for top 5 cases

Rank	Matched Case	-2	-1	0	1	2
1	P058845	3	1	0	2	4
2	P003614	4	3	0	2	2
3	P069852	2	1	0	1	2
4	P041890	2	1	0	2	2
5	P058844	3	1	0	2	3

For each individual attribute, according to the raw data of sensitivity analysis, further statistical analysis is carried out.

Figure 9.9 provide the statistic analysis result of the overall influence of each individual attribute weight changed at -2, -1, 1, 2 level. Because the sensitive analysis considers the influence of weight change to the overall match score and rank for each project, the result is comparable for the same case in each round. As a result, only results for the top 5 matched projects are considered.

			Statistic	Std. Error				Statistic	Std. Error
SA_2	Mean		-.57	.209	SA1	Mean		.76	.129
	95% Confidence Interval for Mean	Lower Bound	-.99			95% Confidence Interval for Mean	Lower Bound	.50	
		Upper Bound	-.15				Upper Bound	1.02	
	5% Trimmed Mean		-.53			5% Trimmed Mean		.83	
	Median		-.50			Median		1.00	
	Variance		3.060			Variance		1.172	
	Std. Deviation		1.749			Std. Deviation		1.083	
	Minimum		-4			Minimum		-2	
	Maximum		2			Maximum		2	
	Range		6			Range		4	
	Interquartile Range		3.00			Interquartile Range		2.00	
	Skewness		-.070	.287		Skewness		-.554	.287
	Kurtosis		-1.359	.566		Kurtosis		-.115	.566
SA_1	Mean		-.14	.126	SA2	Mean		.86	.164
	95% Confidence Interval for Mean	Lower Bound	-.39			95% Confidence Interval for Mean	Lower Bound	.53	
		Upper Bound	.11				Upper Bound	1.19	
	5% Trimmed Mean		-.06			5% Trimmed Mean		.84	
	Median		.00			Median		.00	
	Variance		1.110			Variance		1.892	
	Std. Deviation		1.053			Std. Deviation		1.376	
	Minimum		-3			Minimum		-2	
	Maximum		1			Maximum		4	
	Range		4			Range		6	
	Interquartile Range		2.00			Interquartile Range		2.00	
	Skewness		-.625	.287		Skewness		.196	.287
	Kurtosis		.159	.566		Kurtosis		-.629	.566

Figure 9.9 Influence of change of individual features in sensitive level

As illustrated in Figure 9.9, the absolute values of Skewness are all less than 1, hence the influence in the final match score in each sensitivity level is normal distribution. The range of 95% confidence interval for mean is all less than 50% of the change of the attribute's weight and less than 1% of the match value, which means the system is robustness in calculating the similarity.

Figure 9.10 provides a comparison box plot of each sensitive level. From this we can see that the system is more stable in the [-1, 1] level than the [-2, 2] level as the influence distribution is more centralize and the median value is in the middle of the 95% confidence interval.

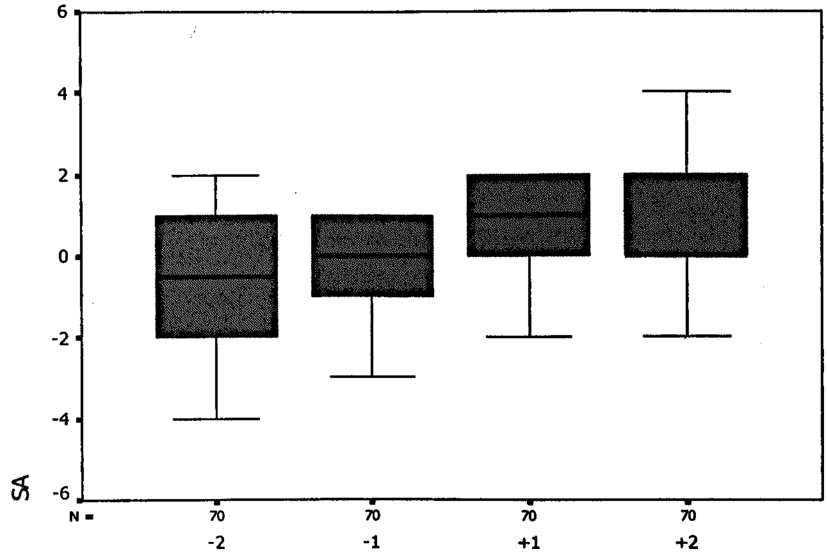

Figure 9.10 Box plot of top 5 cases sensitivity in relevant level

To find how the change of individual retrieval weights influence the case matching, the box plot for each attribute is drawn, as shown in Figure 9.11. Implementation duration, country, sub sector, environment category, grace period, and year of maturity have less influence, as the mean is close to 0 and the 95% confidence range is within 1. Target outcome and ENPV have medium influence. Total project cost, approval year, % of transportation, EIRR, FNPV and FIRR have strong influences.

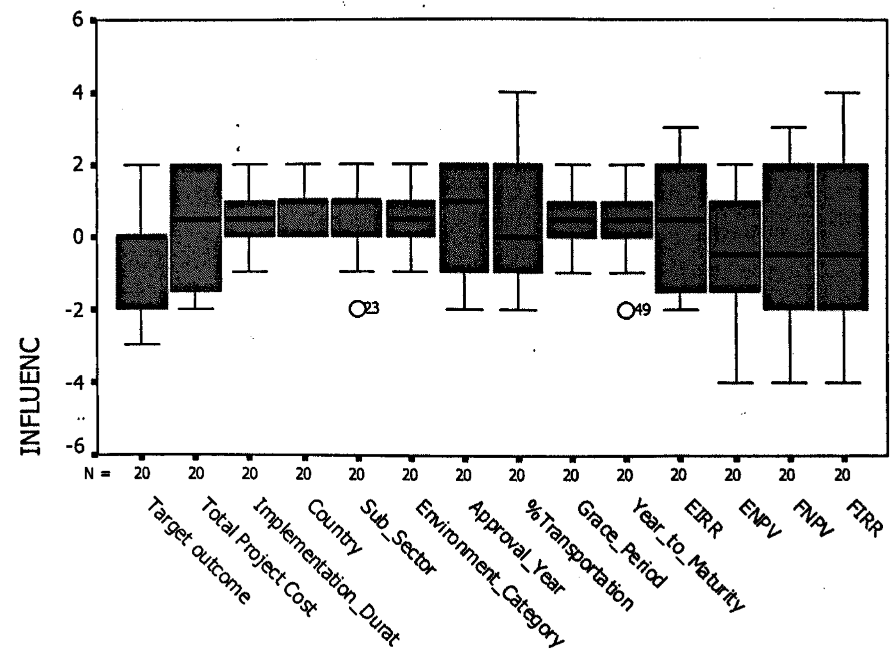

Figure 9.11 Box plot of sensitivity of individual attribute

To find out how much influence is different with each feature and whether the influence is positive or negative, the spider diagram of each attribute is shown in Figure 9.12.

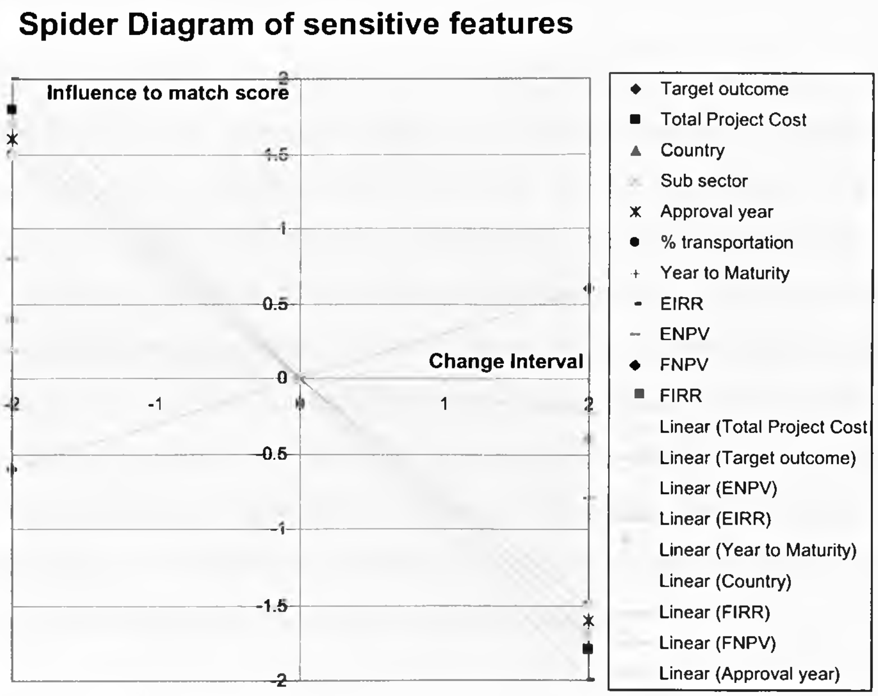

Figure 9.12 Spider diagram for individual sensitive attribute

Figure 9.12 illustrates how the change of 'Target outcome will provide a positive contribution to the case match score, and the change made to all other attributes will reduce the match score. The sensitivity of each attribute is clearly displayed in Figure 9.12. EIRR is the most sensitive attribute, whilst country is the most insensitive one. Figure 9.12 also illustrates in [-2, 2] change level, the most sensitive attribute only reduce 2 score for the over all match score. In real practice this usually does not make a major change to the case ranking, or to the adaptation result. Thus, conclusions can be drawn that the system is reasonably robust.

9.4.3 User Friendliness

O'Keefe *et al* (1987) have suggested that the participants of face validation can be the project team members, potential system users, or people who are knowledgeable about the application domain. Brandon and Basden (1998) and Tuthil (1990) have suggested that it would be advantageous to have experts who were not involved in the knowledge acquisition process to participate in the validation. Because of the difficulty of having the expert coming to Leeds and actually using the system, the validation group is mostly from academic faculties.

As the purpose of CBRisk is to provide an easily accessible, easily understood and flexible tool to help with decision making in risk identification, as described in the previous chapters, the complete model will act as an experienced expert, who accumulates his experience through the practice and reasoning based on the previous experience he has. As a result, most of the participants involved in the user friendliness validation were university students who had knowledge of project risk management. A simple training was provided at the beginning and an explanation of the CBRisk model was given. Any questions were answered and they completed a questionnaire at the end. 18 questionnaires were completed for the user friendless test; the result shows the user interface achieved reasonable satisfaction in terms of ease of understanding and follows up, flexible approach, and satisfactory functionality.

9.5 SUMMARY

This chapter describes the development of a systematic verification and validation process for CBRisk. Verification was underlined first. The collected raw data, data structure and data distribution were evaluated in "data verification"; the knowledge acquisition was examined in "inspect knowledge representation"; and the role of each feature in the system's overall behaviour was assessed in "system configuration parameter test". All the above processes were completed successfully, and hence the model was shown to be verified.

Next, the validation was performed using three tests. (1) Model validation focused on examining the internal CBR model from case retrieval, case adaptation and domain coverage. Validation inspects the result obtained by the system; the system robustness analysis investigates the reliability and sensitivity of the system; and the user interface checks the user friendliness of the system. The case retrieval test was fully successful; the domain coverage test was partially successful shown the restriction of data availability; the case adaptation test was perfectly successful for overall risk level and adaptation of the risk list was successful if the influence of poor domain coverage in some circumstances is ignored by comparison with the case matching level. (2) System robustness analyzes the reliability and sensitivity of the system. Through a range of sensitivity analysis, it was apparent system that can provide a stable operation had been built. How to draw a conclusion, based on the analysis result, was also explained. (3) The user interface checks the user friendliness of the system. The feedback shows that the user interface achieved a highly satisfactory result in terms of ease of understanding, ease of using and flexibility. The results of the validation were promising, which demonstrate that a right prototype CBRisk model for risk identification based on the World Bank transportation projects was built.

CHAPTER 10

CASE STUDY & REVIEW OF CBRISK

10.0 INTRODUCTION

The purpose of this chapter is to examine and to demonstrate the application of the developed model to a real project. This chapter uses CBRisk in the China Fujian second highway project. It demonstrates how the required data can be gathered from a real project and how to use the developed model to identify risks for a certain project. The real data of this project is used in CBRisk from representing the project through data entry, retrieval of the case to find the matching list, reuse of the identification result of matched cases and revision of this result to work out a new solution and finally review of the solution with real practice. The latter section of this chapter analyses the CBRisk outputs in different stages against the original project appraisal document and demonstrates the case refined by comparing it with the real project performance.

10.1 PROJECT INTRODUCTION

The project used in this case study is "P51705: China Fujinan Second Highway Project". This project is selected for several reasons:

- this project has been completed, so the author can evaluate the model is performance against the initial risk identified in the World Bank Project Appraisal Documents before the project started;
- some project managers involved in this project were able to provide suggestions; and
- the author was familiar with the background situation of the project and the underlying policies.

10.1.1 Project Background

Chinese economic development since the opening of the economy in the late 1970s has resulted in a 9.5% average annual rate of economic growth. Key facets of this growth were rapidly increasing domestic and foreign trade as well as increasing personal mobility and consumption of energy. These, in turn, have caused the demand for transport infrastructure and higher quality transport services to surge. Supply capacity, however, was constrained and recognized as one of the most serious bottlenecks to future economic growth and efficiency. In particular, the use of vehicles was growing very rapidly. The road network in China ranks among the sparsest in the world relative to geographic area and population and very few intercity expressways were in existence (World Bank, 1998).

The Government of China (GOC), through its Ministry of Communication (MOC), has embarked on a major program of expressway construction, comprising some 35,000 km, at an estimated cost of some US$150 billion, and leading to the development of the National Trunk Highway System (NTHS), as shown in Figure 10.1.

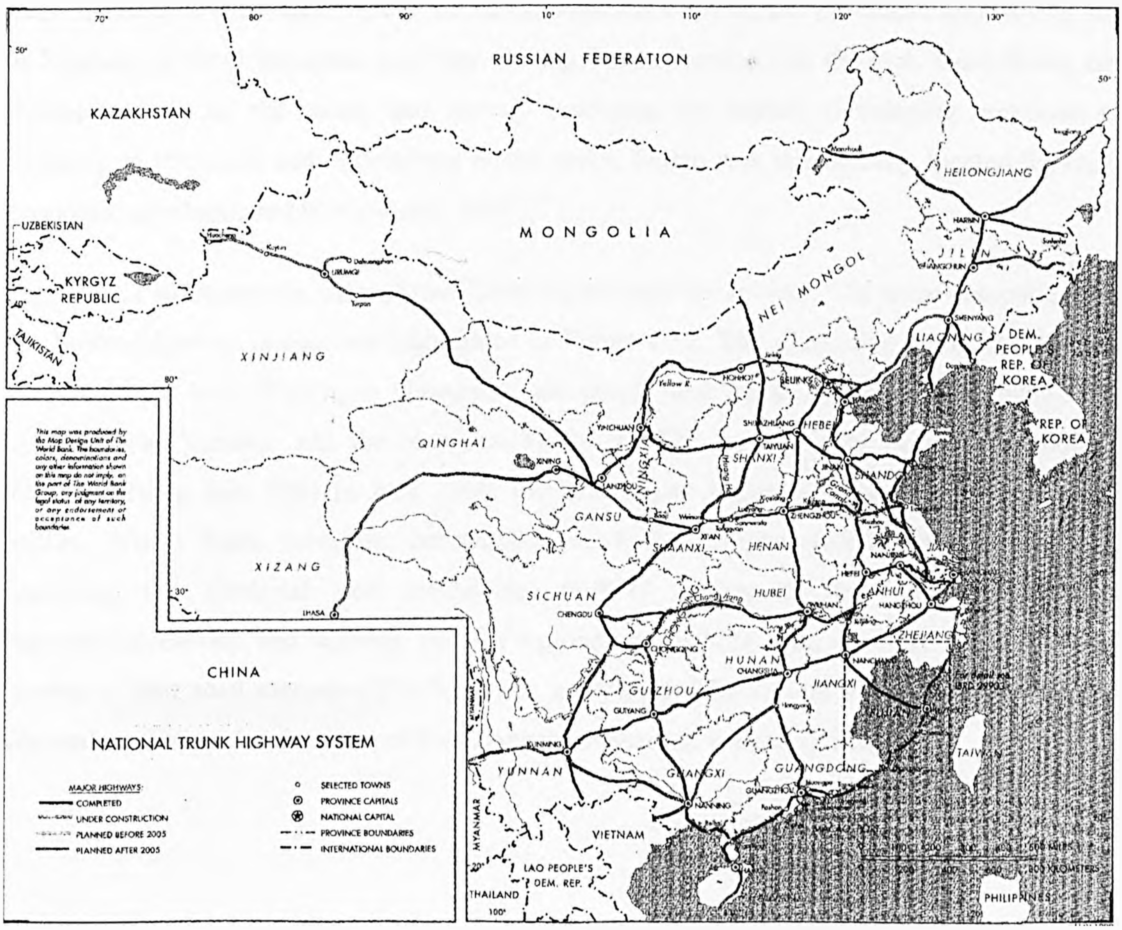

Figure 10.1 Map of China national trunk highway system

This system consists of 12 inter provincial trunk expressways, spanning China and connecting some 100 major cities. The coastal expressway between Tongjiang at the Russian border and Sanya on Hainan Island was among the four highest priority corridors in this program. It crosses the provinces of Heilongjiang, Jilin, Liaoning, Shandong, Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong and Hainan over a distance of approximately 5,200 km. It was designed as a high-grade high-performance facility, at four lanes - dual carriageway standard; fully access controlled, and will be operated throughout as a toll highway.

Fujian Province, with an area of about 121,000 km^2, is situated along the southeast coast of China. It has a population of about 33 million inhabitants. Eighty percent of the land area is classified as mountainous while flat and rolling areas with the centres of economic activity are located along the province's 3,300 km eastern coastline (see Figure 10.2). The location of this project is squared. This geographical configuration presents a natural barrier to the flow of communication and has retarded the development of north-south and east-west transport linkages between Fujian and its neighbouring provinces. Proximity to Taiwan (China) and Hong Kong, however, coupled with preferential government policy, have led to the establishment of successful special economic zones at Fuzhou, the province's capital city and at Xiamen, its most important port city. Facing Taiwan straight on the east, Hong Kong and Southeast Asia on the south, and directly bordering the rapidly developing provinces of Zhejiang in the north and Guangdong in the south, Fujian was strategically located for rapid economic development (World Bank, 1998).

Figure 10.2 illustrates the map of the China Fujian highway project. The three components of the Fujian highway project are highlighted in Figure 10.2. The green line marks the 154 km long highway from Fuzhou to Quanzhou; the purple line marks the 82 km highway from Quanzhou to Xiamen; and the blue line marks the 132 km highway from Zhangzhou to Zhao'an. From July 1994 to June 2000, the first Fujian highway project, with a US$ 124 million World Bank loan, the construction of 81.9 km Quanzhou-Xiamen expressway, including the electrical and mechanical (E&M) equipment for tolling monitoring, telecommunications and lighting and the upgrading of three road sections (58.3 km) and paving of four road sections (200.7km) was completed. The second Fujian highway project focused on further development of the Zhangzhou-Zhao'an express (ZZE).

Figure 10.2 Map of China second Fujian highway project

10.1.2 Project Description

Project Objectives

The aim of the proposed Second Fujian Highway Project (FH2) was to increase economic activity along the Fuzhou-Shenzhen high-priority transport corridor through relieving congestion, facilitating trade and mobility and helping increase efficiency and traffic safety in Southern Fujian. Shenzhen is one of the rapidly developing cities along the east coast. It is just one hour by train from Hong Kong. Since 1980, it was well developed as a communication window between China and the world, especially the counties in Southeast Asia. Figure 10.2 provides a map of Fujian province where the project was based, and its neighbour Guandong province to which Shenzhen belongs. It illustrates the 3 highways along the east coast: Fuzhou-Quanzhou, Quanzhou-Xiamen and Zhangzhou-Zhao'an. The project ended at the border city Zhao'an and the further travel can be made using the existing express between Zhao'an to Shenzhen and on to Hong Kong, as shown in brown line.

The project was considered to have the following objectives (World Bank, 1999):

- to relieve traffic congestion and improve the mobility and integration of inter provincial trade, commerce and tourism between Fujian and Guangdong Provinces by assisting in the development of a key section of China's National Trunk Highway

System (NTHS) in the principal coastal corridor;

- to strengthen highway institutional capacity at the Fujian Provincial Transport Department (FPTD) and related sector institutions, through provision of training, technical assistance and equipment, in planning, design, construction, operation and maintenance of Fujian's highway network;
- to develop and sustain a policy dialogue in the two key policy areas of the commercialization and corporation of a provincial toll road authority and of province-wide highway maintenance management; and
- to improve the safety of road transport.

These objectives did not change until the end of the project. These objectives were achieved gradually along with the project implementation. When project finished, all the above objectives were accomplished.

Benefits and Target Population

The major quantifiable benefits of the project were the reduction in vehicle operating costs and time costs of road users. The expressway provided a shorter route with improved surface conditions and riding comfort, improved geometry, reduced traffic interference, and better safety facilities than the existing roads. The total population in the expressway corridor between Fujian and Guangdong in a direct sense was estimated at about 10 million (based on 1997 data) living in Southern Fujian (World Bank, 1999).

A major expected indirect benefit of a long-term nature was the development of new fruit growing areas mainly, for export purposes to Guangdong, Hong Kong and Southeast-Asian countries. When it was completed and open to traffic, southern Fujian obtained substantially improved access to and from the existing high quality port and airport facilities of Shantou, an important economic centre in Guangdong Province and the major entry and exit point for trade, commerce and tourism in south eastern China.

Project Component

The investment components of the Second Fujian Project (shown in Table 10.1) included the construction of: (i) Zhangzhou-Zhao'an Expressway (ZZE) (132.5 km) in Fujian; (ii) an interconnecting road program (about 13 km); and (iii) other roads (about 42 km). These physical components were strategic investments contributing to the early completion of an inter provincial coastal expressway of about 5,200 km between the Heilongjiang border with the Russian Federation in the North and Hainan Province in the South. The Zhangzhou-Zhao'an expressway was the only section not yet financed linking Fuzhou with Guangzhou and Hong-Kong.

Table 10.1 Project Components (World Bank, 1999)

Component	*Category*	*Cost Intl Contingencies (US$M)*	*% of Total*	*Bank Financing (US$M)*	*% of Bank Financing*
Component 1. Construction of Expressway and Interconnecting Roads	Physical-Institutional				
1.1 Construction of ZZE		493.3	82.8	179.4	36
1.2 Construction of Interconnecting roads		11.3	1.9	2.4	21
1.3 Construction of other roads		35.0	5.9	6.7	19
1.4 Equipment		4.8	0.8	3.7	77
1.5 Construction supervision services		16.6	2.8	4.8	29
Component 2. Institutional Strengthening and Capacity building	Institutional-Policy	1.7	0.3	1.0	100
2.1 Studies/technical assistance in Highway Maintenance Management, Toll Rate; and Road Safety					
2.2 Staff training programs, covering all aspects of highway planning, design, construction, operation, finance and maintenance					
Component 3. Front-end Fee		2.0	0.3	2.0	100
Component 4. Land Acquisition and Resettlement	Implementation on support	30.9	5.2	0.0	0
	Total	**595.6**	**100**	**200**	**34**

Key Policy and Institutional Reforms

The institutional and policy components for the project comprised: (1) staff training and equipment to improve province-wide highway financing, management, and operation; (2) road safety program; (3) studies on highway maintenance management and toll rates; and (4) construction supervision services for the expressway section, its interconnecting roads, and other priority roads.

Implementation Period:

The project implementation was from October 1999 to December 2004. It took 5 years and 3 months.

Executing Agencies:

Under the Fujian Provincial Transport Department (FPTD), the Fujian Provincial Expressway Construction Directorate (FPECD) has the overall responsibility for project preparation and implementation. The Zhangzhou Municipal Expressway Construction Directorate (ZMECD) was the executing agency for Zhangzhou-Zhao'an expressway and the interconnecting road program under the Project. The Zhangzhou Municipal Transport Department was responsible for the other road subprojects. The TA studies and training components and the procurement of equipment were executed by FPECD (World Bank, 1999).

10.1.3 Project Analysis

Project Cost and Financing

The estimated total cost of the project was US$597.4 million with components of US$200 million from IBRD, US$200 million from central government, and US$291.4 million from provincial government. Financial NPV is 4021 million RMB at 7.0 % discount rate with the overall FIRR of 12.5% (World Bank, 1999).

The financial evaluation focuses on the Zhangzhou - Zhao'an expressway which was the main component of the project and was estimated to be financial income with profit making and repayment capabilities. Tolls were charged as a function of vehicle type and size and distance travelled at the same level and structure as the tolls that was collected on Quanzhou-Xiamen expressway. The World Bank financed around 32 percent of the investment and 25 percent will be financed locally. The expressway financial rate of return was 12.5 percent with a net present value of 4,021 million RMB Yuan at 7.0 percent discount rate with an investment recovery period of 16 years. (World Bank, 1999).

Economic

The undertaken economic analysis method for this project was the World Bank's Cost-Benefit Analysis. The economic NPV is Y 4,036 million RMB at 12% discount rate with the overall EIRR of 19.6% (World Bank, 1999).

The estimated economic internal rate of return (EIRR) was 19.4 percent for the entire Zhangzhou - Zhao'an expressway corridor. The EIRR of the entire project civil works (98 percent of total project costs), including the interconnecting link roads and the other roads program was 19.6 percent. The sensitivity analysis indicated that if construction costs are increased by 20 percent and user benefits are reduced by 20 percent, the project EIRR was

15.5 percent, which indicated that even under this pessimistic scenario the project was justified. The switching values analysis indicated that construction costs have to be multiplied by a factor of 2.0 or user benefits have to be multiplied by a factor of 0.5 to yield a project EIRR of 12.0 percent. The risk analysis indicated that the most likely project EIRR was 16.2 percent, the worst case scenario EIRR was 12.6 percent, the best case scenario EIRR was 19.7 percent, and there was a 7.7 percent probability that the EIRR was less than 12 percent. Therefore, the project was economically well justified (World Bank, 1999).

Technical

The technical parameters and estimated costs for construction of the expressway were established by detailed feasibility, design and engineering studies undertaken by the Fujian Provincial Design Institute and were reviewed by Japanese engineering consultants and checked against actual unit costs for the ongoing Fujian Provincial Highway Project. The cost estimate of the project reflected April 1999 prices and was based on the latest available engineering studies and prevailing unit rates for civil works. The cost estimate included physical and price contingencies. Physical contingencies have been calculated at 10% of the base cost of civil works and construction supervision. Price contingencies were calculated for foreign costs, using US$ annual escalation factors of 2.9% in 2000 and 2.5% during 2001-2004. Price contingencies for local costs were based on local annual escalation of 5.0% in 1999 and 5.5% during 2000-2004 (World Bank, 1999).

Social

The project required land acquisition and resettlement. Resettlement Action Plans (RAPs) were in accordance with Bank's Operational Directive 4.30 and have been found satisfactory. Efforts were made during project preparation to minimize land acquisition and resettlement through modifications in highway alignments and designs (World Bank, 1999).

Table 10.2 Land acquisition and resettlement impacts of the highway components

Highway Sections	Land Acquisition (KM^2)	Affected HHS/PAPs		Affected Enterprise	
		HHs	PAPs	Enterprises	Works
Main Expressway (ZZE) component and interconnecting roads	10249	7268	32744	20	236
Chiling-Huxi Highway	181	334	1440		
Zhangzhou Hua'an Road	111	106	461	1	22
Baishui Large Bridge	39	67	248		
Total	**10580**	**7775**	**34893**	**21**	**258**

The project required land acquisition for the expressway, linking roads and other roads which caused readjustment of agricultural land in affected villages and displacement of affected people. A summary of project impact related to land acquisition and resettlement is given in Table 10.2, while more detailed information is contained in. In Table 10.2, the land acquisition area is transferred from Chinese area unit "Mu" to standard KM^2. HHs represents house holds and PAPs represents project affected persons.

Environmental Assessment

Environmental Assessment is one of the Bank's ten Safeguard Policies. The Bank undertakes environmental screening of each proposed project to determine the appropriate extent and type of Environmental Assessment (EA) needed. The Bank classifies the proposed project into one of four categories, depending on the type, location, sensitivity, and scale of the project, as well as the nature and magnitude of its potential environmental impacts.

This project was a category "A" from an environmental point of view. A Category A project is likely to have significant adverse environmental impacts that are sensitive, diverse, or unprecedented. These impacts may affect an area broader than the sites or facilities subject to physical works. The EA for a Category A project examines the project's potential negative and positive environmental impacts, compares them with those of feasible alternatives (including the "without project" scenario), and recommends any measures needed to prevent, minimize, mitigate, or compensate for adverse impacts and improve environmental performance. For a Category "A" project, the borrower is responsible for preparing a report, normally an Environmental Impact Assessment.

Environmental impact assessments (EIA) and Environmental Action Plans (EAP) have been carried out for the main roads included under the project. Major environmental impacts included noise and air pollution during construction and operation; alteration of hydrological regimes, soil erosion and impacts on local ecology during construction; impacts on cultural relics; and transportation of construction materials during construction. The revised final EIAs, EAPs and EA Summaries were submitted in March 1999 and were found to be satisfactory to the Bank (World Bank, 1999).

Primary beneficiaries and other affected groups

The major beneficiaries were the communities along the Zhangzhou-Zhao'an corridor as well as intercity traffic between cities in the corridor and beyond. Benefits resulting from improving the existing highway condition would accrue to the population throughout the province. FPTD and its related institutions would also benefit through the institutional strengthening and training contained in the project (World Bank, 1999).

10.2 MODEL APPLICATION

The following section uses the project information to demonstrate how to use the CBRisk to identify risks and the functionalities included in this model.

Before starting a new application it is necessary to load all the cases to the case base. In the Menu bar of CBRisk screen, as shown in Figure 10.3 below, chose "CBR" and then the sub-menu option "Load Project Data", all the previous project data saved in the data base is loaded to the case base pf CBRisk.

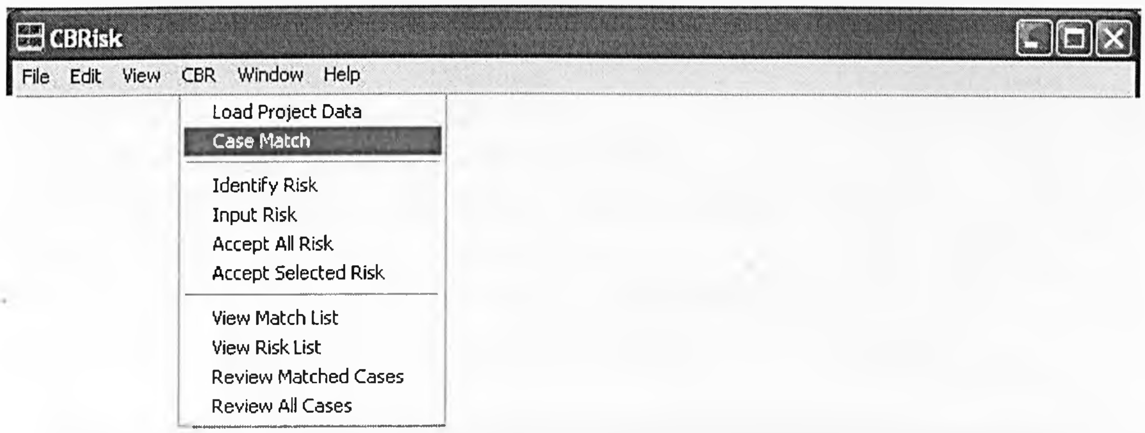

Figure 10.3 The menu bar of CBRisk and CBR options

After finishing loading cases to the case base, CBRisk is ready to start a new application.

10.2.1 Data Gathering

According to the project information explained in section 10.1, the relevant project features to fit in the CBRisk problem representation attributes (see Table 7.2) are extracted and listed in Table 10.3 below.

The information described in Table 10.3 is then input into the CBRisk data entry form, as shown in Figure 10.4. In this form, on the right hand side are the functionality buttons. This form is connected with an external MS Access database, project description table through Art*Enterprise Data Integrator. "Clear" is used to reset the form to start a new input. "Create" is pressed to input information to a new presented-case which is saved in the temporary system space. "Delete" is used to delete the most recently presented case the user created. "Export" can export a new created case to the external Access database. "Print" allows the user to print a hard copy of current screen information. "Query" provides an alternative database matching, which can be used to compare with the CBR matching result.

Table 10.3 Project features of Fujian second highway project

Project ID	P051705	Project Name	Fujian Second Highway Project
Finance_Source	Loan	Government_Committee	395.6
Region	East Asia & Pacific	IBRD_Commitment	200
Country	China	IDA_Commitment	0
Grace_Period	5	Total_Project_Cost	595.6
Year_to_Maturity	20	Implementation_Duration	63
Project_Status	Active	Front_End_Fee_of_Bank_Loan	0.01
% of Transportation	100%	Land_Acquisition	10581333 km^2
WB_Amount	200	Resettlement_People	34893
Main_Loan_Credit	45020	Environment_Category	A
Commitment_Fee	0.75%	Borrower	Ministry of finance
Financial_NPV	486.22	Sub_Sector	Roads and highways
Financial_FIRR	0.125	Lending_Instrument	Specific Investment Loan
Economical_NPV	488.03	Economic_Evaluation_Method	Cost-Benefit Analysis
Economical_EIRR	0.196	Proposed_Terms	Single currency, specify US$, LIBOR-based
Approval_Year	1999	Program_of_Targeted_Intervention	No
Effective_Date	10/1/1999	Implementing_Agency	Ministry of Communication

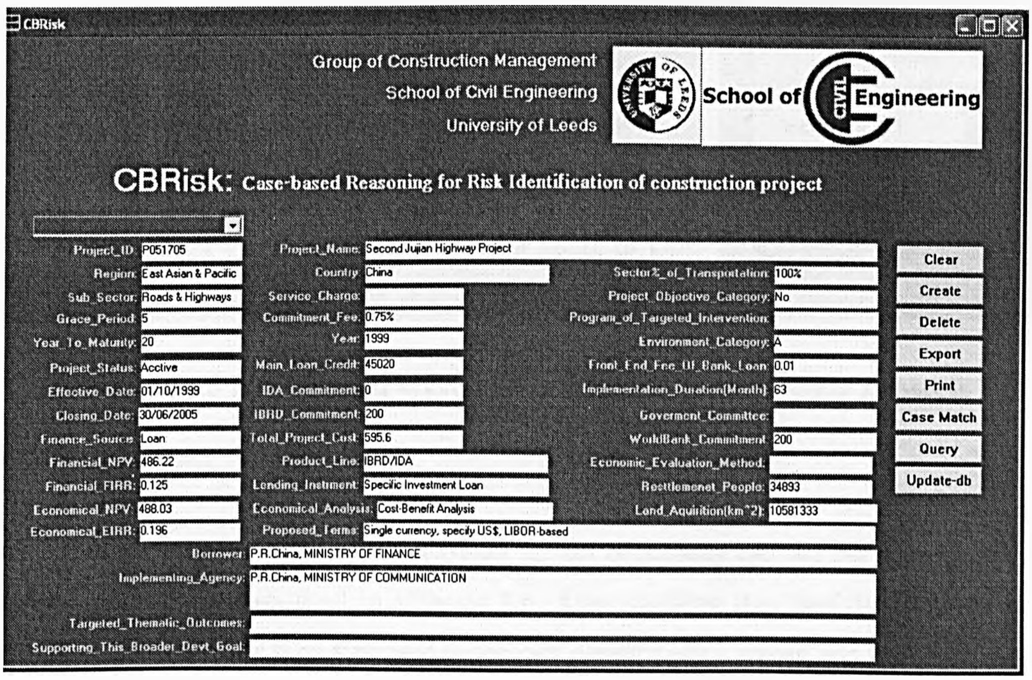

Figure 10.4 Input project descriptions in data entry screen

When inputting of project basic information is finished, by pressing "Case Match" the system firstly generates an instance to declare the input information as a presented-case. The system then analyses each inputted attribute value against the cases stored in the case base based on the case retrieval philosophies explained in section 8.5.

10.2.2 Finding Similar Projects

After pressing "Match Case", the matched cases along with relevant match score were shown in the case match screen (see in Figure 10.5). User can review the project details of a particular matched case by highlighting the relevant case and clicking 'View project Details' on the right hand side. This matching result can be exported as external textual file by clicking 'Export Match Result' button. If the user wants to modify the project description, he can go back to the data entry screen by clicking the 'Back' button. When the inputted information and match result are satisfactory, by clicking 'Identify Risks' button, the case adaptation scheme explained in section 8.6 is used to process the information of the projects to risk identification.

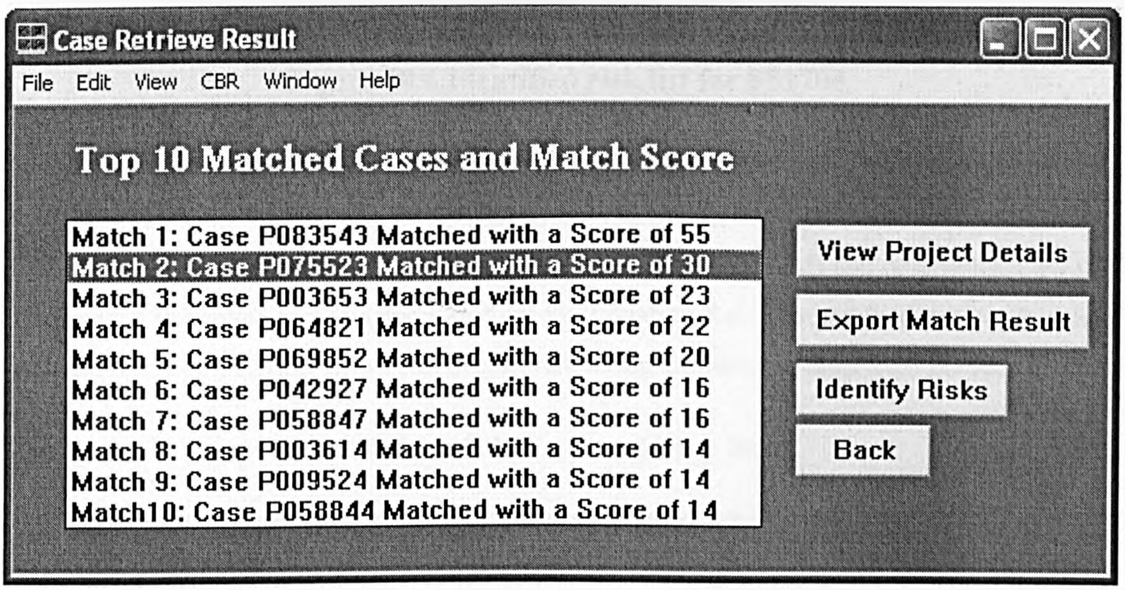

Figure 10.5 Case match screen shows the matched similar projects

10.2.3 Identify Risks

The process of identifying risks in a CBRisk model is actually carried out through the case adaptation procedure introduced in Chapter 8.6. After clicking the "identify risk" button in Figure 10.5, the system is then switched to the risk identification screen and the statistic based adaptation result is firstly shown as Figure 10.6.

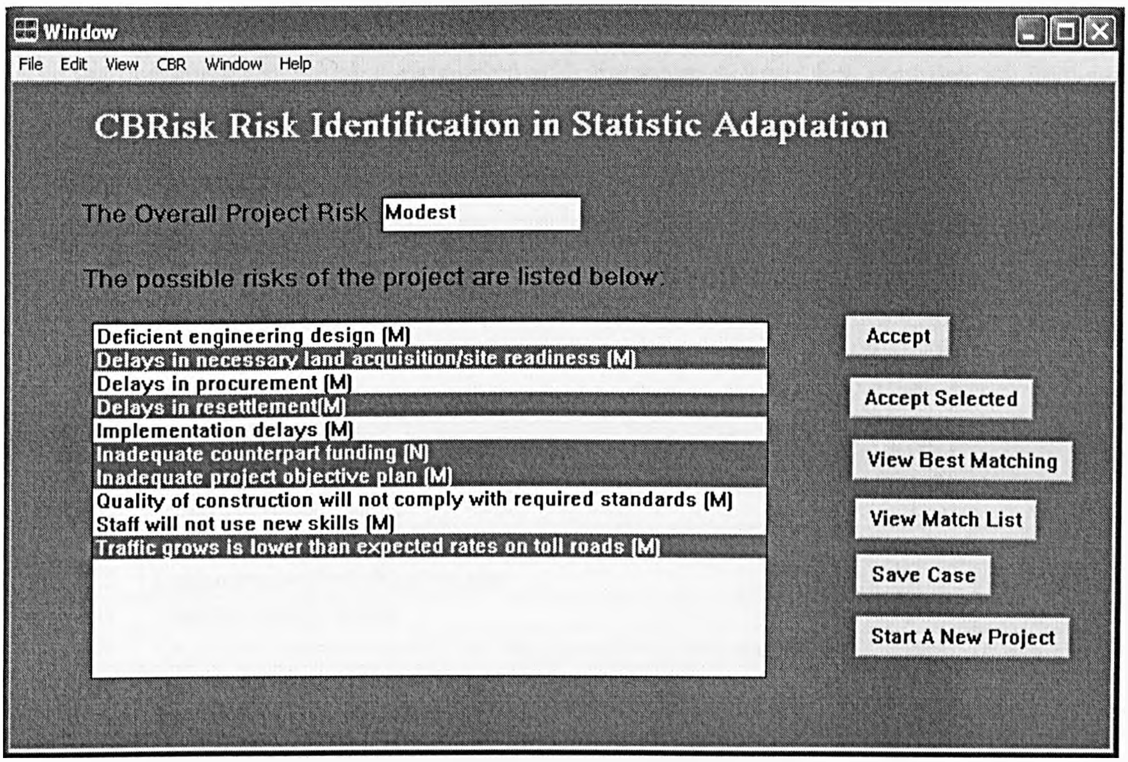

Figure 10.6 Identified risk list for P51705

In Figure 10.6, on top of the screen shows the overall risk level for P51705 as modest. As explained in Table 8.2, the overall risk level and risk rating for each individual risk item are classified into four categories: High (H), Substantial (S), Modest (M) and Negligible/Low (N). The overall risk level is calculated based on the Equation 8.6. The matched project list along with the relevant matching score and overall risk rating are listed in table 10.4 below.

Table 10.4 Case matching list

Project	Name	Match Score	Risk Level	Risk Rate
P083543	Road Maintenance Program (Phase 2)	55	N	1
P075523	Second Infrastructure Asset Management Project	30	M	2
P003653	China Container Transport Project	23	M	2
P064821	Road Maintenance Project	22	M	2
P069852	Wuhan Urban Transport Project	20	M	2
P042927	Mekong Transport and Flood Protection Project	16	M	2
P058847	Xinjiang Highway Project (03)	16	M	2
P003614	Guangzhou City Center Transport Project	14	N	1
P009524	Dhaka Urban Transport	14	M	2
P058844	Third Henan Provincial Highway Project	14	M	2

The identified risk items along with relevant risk rating for each item is listed underneath the overall risk rating. This risk list is generated based on the method introduced in Chapter 8.6.1.

Within the CBRisk, each risk item score is calculated based on Equation 8.7. If the score is large than parameter P, which is calculated with Equations 8.8 and 8.9, then the risk item can be selected. The risk rating of the selected score is then computed based on Equations 8.10 and adjusted with Equation 8.11. The finally risk list is then shown as in Table 10.5 below. The risk code and risk rating in this table are translated into the description text based on the risk log, and then displayed in Figure 10.6.

Table 10.5 Risk identification based on statistic adaptation for project P051705

Risk Code	Risk Item	Risk Rating	Risk Level
101	Delays in procurement	1	M
146	Delays in necessary land acquisition / site readiness	1	M
225	Inadequate project objective plan	2	M
163	Implementation delays	2	M
140	Quality of construction will not comply with required standards	2	N
111	Inadequate counterpart funding	1	N
158	Staff will not use new skills	2	N
183	Traffic grows is lower than expected rates on toll roads	1	N
147	Delays in resettlement	1	N
143	Deficient engineering design	1	N

According to the case adaptation procedure shown in Figure 10.9, this primary result is viewed by the user. If the risk items and relevant risk rating are satisfactory, the user can simply click "Accept All" and then, the risk identification process is finished. If the user does not agree with some of the risk items/risk rating, he can highlight the accepted risk items as shown in Figure 10.6, such as "Delays in land acquisition and resettlement (N)" etc, and then click "Accept Selected". In this case, the system will only save the chosen information as an identified risk result for current project. The selected risk items are stored in a temporary memory in the system. The user can add some more items that they identified later on, based on their knowledge and experience.

However, if the user is not satisfied with this risk list, he can alternatively review the solution of the best match cases by click "View Best Match". As described in section 8.6, to carry out the best matching adaptation for CBRisk, the case matching score should be more than the control parameter V=52. Even through the matching score is less than the direct best matching parameter V_c=102, because there is only one case P083543 that reaches this criterion, the risks from the "Road Maintenance Program-Phase II" project in Vietnam were used as best the matching results, as shown in Figure 10.7 below.

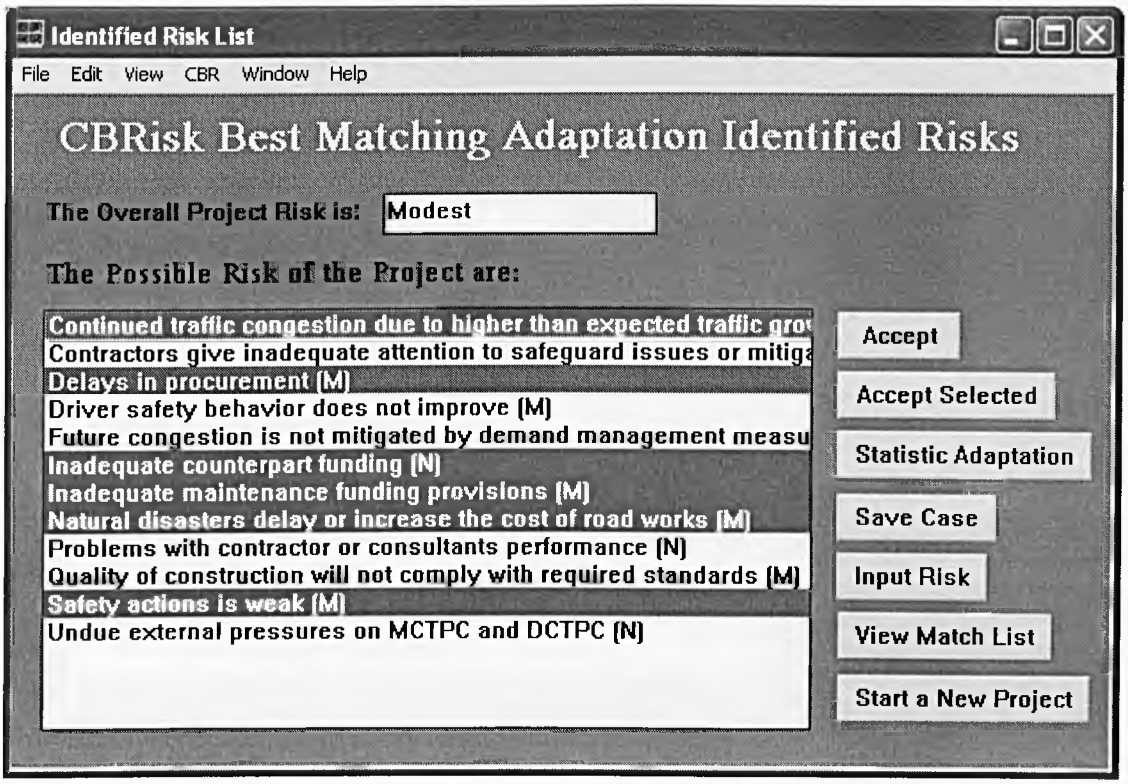

Figure 10.7 Best matching adaptation result based on P083543

The user can use the entire content of this risk list as the identified risk for the current project. Alternatively, he can choose satisfactory items as part of a solution for the current project, and store them in a temporary memory to use in an expert based adaptation, as shown in Figure 10.8. However, if the selected risk items are insufficient, the user can input the further risk items and relevant risk rating identified based on his own knowledge, as shown in Figure 10.8. At this stage, if the user saved any selected risks in Figure 10.6 or 10.7, the system would automatically bring them up into the further risk input screen and the user could finally decide whether he would like to use these risk items or not. The user can also reference the statistic adaptation result, best matching result, and solutions from other similar projects.

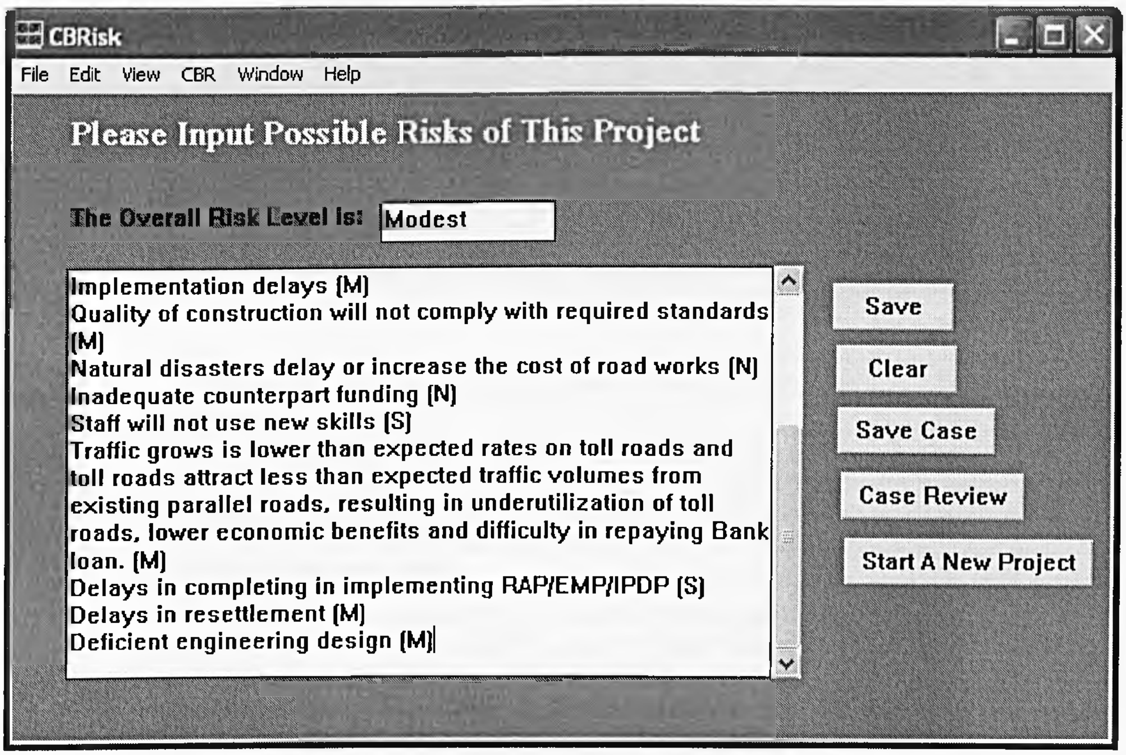

Figure 10.8 Input user identified risks for P51705

As the original risk identification document for this project is available, the author input the risk list based on the World Bank documents (see Figure 10.8) and added the risk of natural disaster which will be explained further in section 10.3. After users input their chosen risk items and risk rates, this complete risk list, along with the project description in Figure 10.4, is added to a temporary space in CBRisk. As explained in Chapter 8.6.1, before the statistic based adaptation is carried out, the risk list for each project should be coded with the relevant risk code from the risk log generated in Chapter 7. Thus, the user input risk list is firstly reviewed by the system administrator to evaluate whether each case can be added to the case base. If he is satisfied, the administrator will allocate the relevant risk code for each risk item and risk level, as shown in Table 10.6 below.

Table 10.6 Risk coding for Project P051705

Risk Code	Risk Item	Risk Level	Risk Rating
101	Delays in procurement	M	2
162	Problems with contractor or consultants' performance	M	2
113	Inadequate maintenance funding provisions	S	3
146	Delays in necessary land acquisition / site readiness	M	2
225	Inadequate project objective plan, training program irrelevant to needs	M	2
114	Construction materials and human resources not available on time	N	1
163	Implementation delays	M	2
140	Quality of construction will not comply with required standards	M	2
210	Natural disasters delay or increase the cost of road works	N	1
111	Inadequate counterpart funding	N	1
158	Staff will not use new skills	S	3
183	Traffic growth is lower than expected rates on toll roads and toll roads attract less than expected traffic volumes from existing parallel roads, resulting in underutilization of toll roads, lower economic benefits and difficulty in repaying Bank loan.	M	2
148	Delays in completing implementation of RAP/EMP/IPDP	S	3
147	Delays in resettlement	M	2
143	Deficient engineering design	M	2

The coded risk list along with case description in Table 10.3, can then be retained in the case base as a piece of new knowledge for future use. By understanding that once a project is saved in the case-base, it can be used as a piece of new knowledge to retrieve problems for other projects. It is important that the saved case is correct.

10.2.4 Review Previous Projects

In Figures 10.8 and 10.5 the user will be able to review the matched cases by clicking "Case Review". The user is then led to the case reviewing screen. By highlighting the relevant case, the details of the relevant case can be reviewed as shown in Figure 10.6.

In Figure 10.9, the top left side is a drop down box which lists the top ten matched cases. When choosing a relevant project, the detailed information for this project is shown on the screen. The user can simply click "Next" or "Previous" to review all other matched projects. If the user would like to get more information on a project, by clicking "Online Resource" the user is led to the relevant website that can download all documents available for the project. Different from reviewing the match screen, the reviewing screen displays the risk items in their original information before being coded.

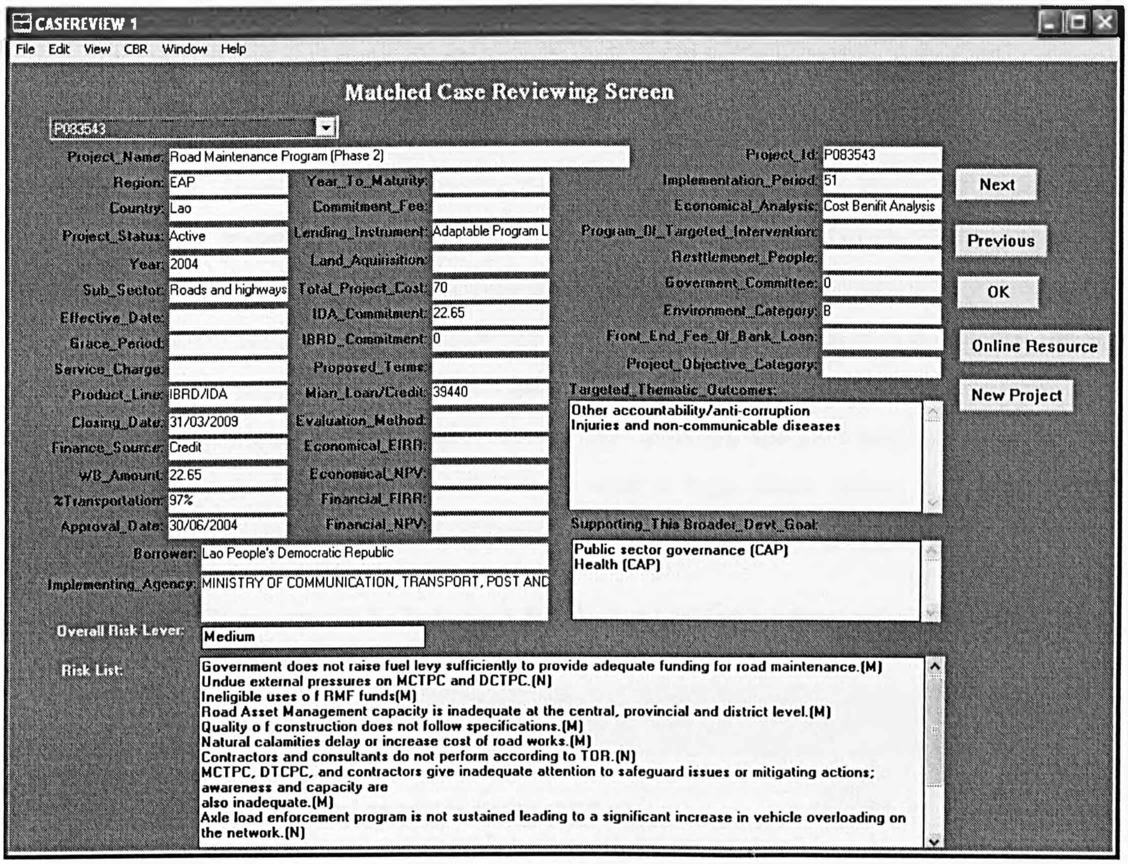

Figure 10.9 Reviewing matched case screen: P083543

By reviewing the detail of matched cases, the user can get some idea of the kind of risks associated with a similar project. This can provide the user with decision support in choosing appropriate risk items, especially when case matching is poor and the expert based adaptation is used.

Project P083543 is the best matching case for "Second Fujian highway project". As the matching score is only 55, which is in the middle matching level, the risk items that can be borrowed from its risk list are those around 50%. The matching score can give the user an approximate idea of how much risk information should be referenced from each matching project.

10.3 DISCUSSION

10.3.1 Model Performance

As demonstrated in Section 8.6, the solutions coming out of CBRisk include a statistic adaptation result and best matching adaptation result. As a result, the evaluation of model performance of this case study focuses on these two outputs.

Statistic Adaptation result

To compare the identified risks through CBRisk and the original risk list in the World Bank project appraisal document, the author can estimate how well the CBRisk works. Table 10.7 lists the common risks identified from both the World Bank 'Project Appraisal Document' and the CBRisk model for project P51705. In the World Bank project appraisal document, the risk items listed do not have sequences, which mean if the risk rating is ignored; all risk items have the same priority. However, in CBRisk as the risk items are selected based on the risk item score, which are listed sequentially the items with a high score means will have a high probability risk. The risk items listed in Table 10.7 are in a descending sequence.

Table 10.7 Common risks between PAD and statistic adaptation for P051705

Risk Item	CBRisk Rating	Original Rating
Delays in resettlement	M	N
Delays in necessary land acquisition/site readiness	M	N
Quality of construction will not comply with required standards	M	M
Inadequate counterpart funding	N	N
Staff will not use new skills	M	M
Traffic growth is lower than expected rates on toll roads	M	N
Deficient engineering design	M	N
Delays in procurement	M	M
Implementation delays	M	M
Inadequate project objective plan	M	M

Table 10.7 highlights the risk items which have a different risk rating from the original Project Appraisal Document and CBRisk statistic adaptation output. It illustrates that the risk rating of overall individual risk items from CBRisk are slightly higher than the rating from PAD.

As explained in Chapter 8.6.1, the risk rating reflects on the impact of risk. It is calculated based on the cases that contained similar risk items and is adjusted by comparison with project "overall risk level". However, the risk rating that is less than the overall risk level is considered not to be in the range of adjustment. Therefore, for some of the risk items the risk rating is slightly higher than the real situation. The risk item score is used to calculate the probability of a risk item against other risk items in matched cases. Risk rating is focused on the impact of the risk items. It is calculated according to the cases sharing the same risk item. As a result, it is reasonable that the risk rating of individual risk items might be slightly higher than expected. As the overall risk level is more accurate as it is calculated based on all the matching cases, if the risk item has a higher risk rating than the overall risk level, it is suggested that the overall risk level is used to make appropriate adjustment.

Secondly, the matched case list as shown in Figure 10.5 has the highest matching score of 55 and the rest of the scores are all less than 35, which illustrates the case matching is below the average matching level for the second Fujian highway project. In other words, the matched cases are not very similar to the current project. As a result, the case adaptation result risk identification that was carried out based on the matching cases, was not expected to be perfect.

Table 10.8 Uncommon risks between PAD and statistic adaptation for P051705

Source	Risk Item	Risk Rating
WB PAD	Human resources not available on time	N
	Training program irrelevant to needs and objectives	M
	Selection of incompetent contractors	M
	Policy and Institutional Reform will not receive large ownership and will be unsuccessful	M
CBRisk	Natural events delay or increase the cost of road works	N
	Problems with contractor or consultants performance	M

Table 10.8 lists the different risk items identified by the World Bank project appraisal document and CBRisk. Among them, the "selection of incompetent contractors" in PAD and the "problems with contractor or consultants' performance" has correlation, because incompetent contractors more likely have performance problems later on.

CBRisk identified an additional risk item: "Natural events delay or increase the cost of road works". Figure 10.7 illustrates the location where the project is implemented. As the highway express is built along the east coast, typhoons and storms occur quite often in these areas. When this happens, damaged facilities and site works need to be repaired, which will require further input resources, time and labour. As a result, implementation might be delayed and the cost increased.

Other risk items listed in PAD are not identified by CBRisk. Because the similarity of the matched score is not high, accuracy of the identification can be impacted. In addition, through data collection and the generation of a risk log, the author noticed that projects from the same country are likely to use similar risk descriptions and they usually share some common risks. Moreover, the generation and classification of risk items is important. If classifying risk items at a more detailed level, fewer projects will share the same risk code. This can reduce the probability of risk items, and affect the risk item score. As a result some risks might be ignored because of the low risk item score calculated in the case adaptation stage. On the other hand, if the risk classification is very brief, the user cannot get enough detail of what sort of risks accompany the project, and is not able to achieve the initial role of risk identification explained in Chapter 2.

Best Matching Result

Similarly, Table 10.9 lists the common risk items for the World Bank project appraisal document and the best matching adaptation from the CBRisk. The risk items with different risk ratings are highlighted. The overall risk level is modest, as is the original identification. Six out of a total of twelve risk items are common. Identification of risk items through best matching has a 50% accuracy. Among these items, 67% of the risk ratings are correct.

Table 10.9 Common risks between PAD and best matching project: P083543

Risk Item	Original Rating	Best Match Rating
Lower than expected growth rates of traffic volumes	N	N
Inadequate counterpart funding	N	N
Delays in preparation and implementation	M	N
Selection of incompetent contractors	M	M
Delays in procurement	N	M
Inadequate quality of construction	M	M
Overall Risk Level	**M**	**M**

Because there is a significant difference between other uncommon risk items, the author does not provide further analysis. However, the author noticed that the best matching result include two items that were also included in the statistical adaptation result: "Natural events delay or increase the cost of road works" with the risk rating of Modest, and "Problems with contractor or consultants' performance" with the risk rating of Modest. These two risk items may catch the attention of project manager and may be considered for inclusion in the final risk identification list later on.

10.3.2 Case Refinement

As illustrated in Figure 6.9, the final stage of CBRisk is refining the cases stored in the case base. The identified risks for each project are stored in the system. As the risk base and risk log are developed based on the project appraisal document, the risks identification information used in the case adaptation is generated before the project actually starts. However, the real project performance might not be similar as expected in the project appraisal document. At the project appraisal stage, because there are a number of project features containing a variety of uncertainties, the risk identification result at the appraisal stage might not fit the real project exactly. Consequently, subsequently modifying the identified risks according to the real project performance can provide a more accurate adaptation in risk identification. Meanwhile, case refinement provides a view of the accuracy of the initial appraisal compared with the real project performance in an attempt to give the user idea if the accruals that can be expected from CBRisk's risk identification, when it works successfully.

Experts involved in the implementation of this project were invited to evaluate the risk identification result through a questionnaire (see Appendix G) for both statistic based adaptation result and the best matching adaptation results as displayed in Figures 10.6 and 10.7. They are a representative of the A2 contract for the Zhangzhao expressway, and the project manager in a construction bureau of Guandong province. They agreed with most of the identified risks, and added other their suggestions on risk rating.

- For "Inadequate quality of construction", the risk rating is N.

Because the "Second Fujian Highway Project" is a World Bank loan project, many parties, including the government, are concerned about it from the design stage to its construction. The highly qualified designers and contractors are trying their best to complete the project according to relevant standards. The managers of the design and contractor companies will be fined for inappropriate administration if standards are below the specified quality. As a result there is less possibility that the quality of construction is inadequate, and the relevant risk rating should be changed to Negligible/Low.

- For "Selection of incompetent contractors", the risk rating is N.

Because Zhangzhao Expressway is a big and notable project, incompetent contractors cannot be included in the shortlist. The companies authorized by the contractor are all top-ranking professional companies with high reputations in highway projects. The tendering committee chose the contractor carefully by considering companies' overall qualifications, techniques, facilitates and experience in similar projects. Consequently, the risk rating of "selection of incompetent contractors" should be amended to Negligible/Low as well.

Comparing previous risk identification with real performance is the way to refine the case base. Based on the above analysis, the new risk list for the Fujian second highway project is generated and input into the CBRisk model, as shown in Figure 10.7. This includes the change of the risk ratings for "Inadequate quality of construction" and "Selection of incompetent contractors" from Modest to Negligible/Low. Two more risk items are added to the risk list. They are "Natural disasters' delay or increase in the cost of road works" and "Problems with the contractor or consultant's performance", both with risk ratings of Modest.

As time goes by more projects are completed and the information of real performance can be used to refine cases in the case base. When more cases are refined, the accuracy of the solution for existing cases is enhanced. Consequently, risk identification based on the existing project risk list can be improved and a better system performance can be achieved.

10.4 SUMMARY

In this case study the second Fujian highway project is used to demonstrate how the CBRisk model works. This project was selected because the experts involved in the project were able to provide further evaluation and the author is familiar with the background of the project. Furthermore, in using CBRisk, the author found that during application this project represented some issues, for instance, the overall risk rating of individual risk items is slightly higher and the classification of risk items might influence the accuracy of CBRisk. Because the matching score of this project is in the middle, the best matching result should be compared with the matching score. From this case study, the author found that compared with the finial risk lists, the statistic adaptation achieved about 80% accuracy and about 50% of the best matching adaptation items are correct, which is almost equivalent to the matching level. Compared with the real project performance, the original risk identification from the World Bank project appraisal document is about 90% accurate. This illustrates that the CBRisk model is unlikely to achieve more than 90% identification accuracy for a project and the user should not expect absolute accuracy of risk identification from CBRisk at this stage. The risk rating of individual risk items in the statistical adaptation is slightly lower than the original PAD; but fairly close to best matching adaptation. The author expects that, as well as improving the domain coverage, the matching score can also be improved; the accuracy of the CBRisk identification can be enhanced. It will be necessary to collect more cases to be retained in the case base. In addition, when more cases are closed, the case base can be better refined, which can also enhance the accuracy level of CBRisk.

CHAPTER 11

CONCLUSION

11.0 REVIEW

This final chapter brings together the main findings of the research. The context of the thesis is summarized and the contribution to knowledge is discussed along with ideas for future research.

The literature review revealed that risk management is an important process in construction project management as it provides in support for decision making under conditions of uncertainty by reducing the possibility of delay, overrun and failure. In other words, there is a greater chance of a project not meeting its objectives if risk management is not used. Attempts have already been made to introduce better risk management schemes, approaches, or systems in construction firms. Despite a variety of risk management and identification methods being developed, these approaches have not resulted in the expected benefits to the construction industry. To improve risk management, risk identification should firstly be improved. Because of the complex nature of construction projects and the characteristics of project risk identification, an intelligent knowledge based model that can mimic the human reasoning process is required to support decision making and improve risk identification.

11.1 CONCLUSION

The conclusions drawn from this thesis are relevant to the process of risk identification and within the limitations of this research. Here the key conclusions of this study are outlined.

11.1.1 Objective 1: Investigating Problems in Current Project Risk Management

Risks can only be measured, controlled, transferred and managed once they have been identified. Risk identification is the process of systematically deciding which risks are associated with a construction project. Risk identification is fundamental in project risk management as it underlies the risks needed to be taken forward for further analysis, the risk analysis method to be used, and the scope of risk analysis. Consequently, the accuracy of risk identification can directly influence the accuracy of risk analysis, so as to the later on decision making. In other words, if the risk identification is not accurate, risk analysis will not be accurate or sufficient, and the decisions made based on the risk analysis will not be robust. Therefore, to improve risk management, risk identification should first be improved. Risk identification is usually based on information from a variety of projects and the experts' knowledge. However, information might not be well documented and the expert knowledge may be lost because of job changes or retirement. Hence, the limitations of current risk identification methods result from a lack of information/knowledge and communication among project participants.

11.1.2 Objective 2: Knowledge Management for Risk Identification

This thesis demonstrates the importance of information collection and sharing if the risk identification process is to be improved. Risk identification is a process which is largely qualitative in its approach. To improve qualitative risk analysis, information and knowledge management are key issues. Risk identification processes a range of raw data and information about the project and builds up knowledge based on them. The author found that current risk identification methods are largely based on expert knowledge and therefore identification is influenced by human behaviour such as intuition, subjective judgement and the restriction of human ability to handle complex reasoning. Well known risk identification methods, such as questionnaires/checklists, brainstorming, examination of historical documents, and Delphi are all based on previous knowledge or experts' experience. Hence, knowledge sharing and transfer are important for risk identification. Psychological research has found that people make future decisions based on the three most recent decisions made and there are about nine decision attributes that a person can effectively encompass each time. This illustrates that humans have a limited information processing capability. This limited information-processing capability means that they cannot directly deal with complex problems even though the information may be available in some form.

The author highlighted the importance of managing knowledge and information in relation to project risks for risk identification; and the benefit gained through knowledge management, including promoting the application of risk management, improving communication among project participants and enhancing knowledge sharing.

11.1.3 Objective 3: Choosing Appropriate Methods and Tools for Improving Risk Identification

Overcoming the limitations of human reasoning is another core issue if risk identification is to be improved. Findings showed that modelling a judge's decision-making process resulted in more accurate outcomes than the judge's own decisions. As a result, an expert's bias and availability will directly influence the result of risk identification and therefore influence the final risk analysis output. The configuration and style of these features would mean that computer based AI models are applicable.

Compared with traditional risk identification methods, artificial intelligence techniques have advantages in terms of reasoning, data processing and analysis. ANN has the advantage of self-learning, self-organizing and real time operation, but ANN is like a black box that cannot provide explanations for decisions and a knowledge-based system explains its decisions by referencing to rules which the user may not fully understand or accept. Case based reasoning stands out from other AI system because:

- CBR is particularly applicable to highly dynamic or poorly-understood domains, or where expert knowledge is difficult to divine or encode, such as project risk.
- It has the advantage in handling non-numerical information. Moreover, it can reason even when some project information is missing. The output solution is available as a textual description, which is easy to understand.
- The reasoning process is retrievable. As a result, if the system output is not satisfactory, the user can review the matched cases and obtain more reliable

information to support decision making.

- Case-based reasoning learns from previous experience and solves similar problems by referencing earlier solutions. Therefore, the case-based risk identification system can act as a risk expert.
- It accumulates experience from solving problems and can provide detailed information in a timely manner.

Case-based reasoning was compared with a variety of other artificial intelligent techniques and it was concluded that CBR was the most suitable approach to adopt for risk identification. Consequently, a case based reasoning system that mimics the human reasoning process was adapted as an appropriate approach. A case-based reasoning model for project risk identification is particularly beneficial in knowledge sharing and transfer, impartial reasoning and decision making support.

After consideration of the problem and data available the author chose Mindbox's Art*Enterprise as the tool to develop the prototype model. ART*Enterprise has very good overall functionality and flexibility when compared with the other CBR packages such as data integrator. It allows easy database access, comprehensive user interface building, supports case-based reasoning, rule-based reasoning procedural programming and object-oriented programming.

11.1.4 Objective 4: Developing an Approach for Risk Identification

A prototype model, CBRisk, was developed by the author and tested with project data from the World Bank transportation projects. It is the first time that case-based reasoning has been used for project risk identification.

First, the requirements specification was described. Based on this, the conceptual framework of CBRisk was designed. Subsequently, the World Bank project data used to build up the model was collected and analysed. The structure of the data set was constructed for different functionality modules. The CBRisk application model was then developed through case indexing, case representation, case retrieval and case adaptation. Particular attention was paid to the most important two processes: case retrieval to find similar projects and case adaptation to generate risk identification results for new projects. In case retrieval, formulae used to calculate attribute score or case score were explained. Because ART*Enterprise does not provide case adaptation, a systematic adaptation procedure to provide flexible solutions was developed by the author. The underlying principles for each process and the way to estimate relevant parameters were explained in detail.

Case adaptation allows the computer to learn from previous knowledge and produce solutions based on this knowledge. The adaptation methods in CBRisk have three options: a statistical based adaptation scheme, best matching cases, and human expert's decision. In the statistical adaptation, risk rating in terms of impact and opportunity were both considered. To ensure the system can provide a more accurate output, rules to adjust the risk rating to a reasonable level were generated. The adaptation process provides a robust choice that allows the CBRisk model to capture more accurate solutions, again demonstrating its role in decision support and encouraging the involvement of the human experts. This process is useful for maintaining and improving the quality of the case-base, which in the long term can further improve the system's performance.

CBR as a paradigm to simulate the process of human problem solving is particularly useful in coping with complex information. CBRisk demonstrates for the first time that it is possible to utilise advanced programming techniques to identify project risks in a robust and objective manner. The CBRisk model makes knowledge and expertise in risk identification more readily available and should also improve communication between project participants. Risk identification and estimating the accuracy of, and confidence in, investment decisions may also be improved if this approach is adopted. This is because CBR model can reason in the same way as human beings but in an impartial, objective manner. CBRisk is a prototype that implements CBR in modelling risk identification for construction projects. It does not try to take over the role of experts or force them to accept the output of the system; rather it facilitates human experts in their decision making.

For the first time a coherent and integrated databank of information from a number of previous projects were brought together, leading to the robust identification of risk sources and a preliminary assessment of their potential impact. Given the range and variety of the World Bank projects, the decision support system developed can be of benefit to other clients engaged in transportation work.

The primary contribution to knowledge of this research is to demonstrate that a case-based approach enhances the knowledge sharing and transfer of project risk identification and provides support to decision making at project appraisal stage. The case-based reasoning risk identification model is not only useful for coping with complex information and knowledge in relation to project risks, but also in reducing the subjective influence from human behaviour and by providing impartial decision support. This model provides a platform for referencing previous knowledge in textual format for the user. This approach is easy to understand and accept. Besides saving the time, expense, and effort of the project manager, this approach intends to exploit the advantages of knowledge sharing, increasing confidence and efficiency in investment decisions, and enhancing communication among project participants. This should bring about a change to risk identification and promote the application of risk management to the construction industry.

11.1.5 Objective 5: Verification and Validation of the Model

A verification and validation (V&V) approach is a crucial tool to support building a successful system. To ensure that the model was built appropriately and was able to achieve the expected functionality, a systematic verification and validation processes was developed.

Verification determines if the model has been built correctly. The verification process included three components: data verification which verified the knowledge stored in the case base in terms of correctness and redundancy; inspect representative which verified the frame for acquiring knowledge with indices inspection and a case coverage test; and a system configuration parameters test which verified the retrieval principles according to experts' opinion, distribution analysis and correlation analysis. All these assessed the role of each feature of a case in the system's overall behaviour. During verification, the model was verified step by step from collected data, knowledge representation and system configuration parameters. In each step, a detailed explanation of which data/value should be considered and which statistical analysis method should be carried out was provided, along with which indicators should be used to guide the improvement, and how to make amendments based on relevant indicators. A calculation formula was used to adjust the correlation effect of attributes matching weights based on the Pearson parameter. After verification inappropriate information was revised, and the model was ready to produce proper output for decision support.

Validation determines if the right model was built. The validation was performed using three tests. Model validation focused on examining the internal CBR model from case retrieval, case adaptation and domain coverage. Validation inspects the result obtained by the system; the system's robustness analysis investigates its reliability and sensitivity of the system; and the user interface checks the user friendliness of the system. The case retrieval test was fully successful; the domain coverage test was partially successful given the restriction of data availability; the case adaptation test was perfect for overall risk level and the adaptation of risk list was successful if the influence of poor domain coverage in some circumstances is ignored by comparing it with the case matching level. System robustness analysed the reliability and sensitivity of the system. Through a range of sensitivity analysis, it was concluded that the system can provide a stable operation. How to draw collusion based on the analysis result was also explained. The user interface checks the user friendliness of the system. The results show that the user interface is highly satisfactory in terms of ease of understanding, ease of use and flexibility. The results of the validation demonstrated that the right prototype model for project risk identification has been built.

Further, because of the unique features of CBR systems against other AI approaches, the systematic verification and validation processes developed in this thesis can be referenced and transferred to V&V for other CBR systems.

11.1.6 Objective 6: Applying the Model

The author demonstrated how to use the CBRisk for the second Fujian highway project and the overall functionality of this model. By comparing the risk identification in the original World Bank document with the output of CBRisk, the performance of CBRisk was discussed. Based on the actual performance of the real project, how to achieve case refinement was then demonstrated.

The author expects that, along with improvements in domain coverage, the matching score can also be improved and the accuracy of the CBRisk identification can be enhanced. This requires the collection of further cases and their retention in the case base. In addition, when more cases are closed, the case base can be better refined. This might result in enhancing the accuracy level of CBRisk.

11.2 RESEARCH LIMITATIONS

Because of the complex nature of construction projects, risk identification is difficult at project appraisal stage, as it involves a great deal of information and uncertainty. No method should be thought of as a panacea, but only as a tool for better understanding and better application. The limitations of the research lie in the following issues.

11.2.1 Limitations of Case-based Reasoning

No modelling approach is perfect, and CBR is no exception. As case-based reasoning provides solutions based on previous similar cases, if the domain coverage is insufficient in amount and scope, or if a case with special circumstances is introduced, a case-based reasoning system might use previous experience without criticizing or validating it in the context of the new situation and may not be able to identify and thus consider the most appropriate cases when solving the present problem with case-based reasoning. Relying on previous experience without validation may result in inefficient or incorrect solutions being recommended, causing an increase in problem-solving time or errors that may have negative effects on the process of learning. These are potential weaknesses of case-based reasoning systems. They should serve as a warning to decision-makers that they should not shirk their duty of using their own minds and experiences to criticize and evaluate potential solutions made by a case-based reasoner.

CBRisk is a decision support system that produces information to be considered when people are trying to make decisions about difficult risk identification problems. It is not a decision-maker itself. Because of a variety of problems with data collection, the case coverage of CBRisk is not fully satisfied at this moment in time. Hence, the accuracy of risk identification might be influenced. The domain coverage can be improved when more satisfactory project information is collected. In another words, if a firm wants to build up a CBRisk system for itself with its existing risk register database, the concerns of domain coverage would be reduced.

To carry out accurate risk identification, it is important to have sufficient cases. However, it does not imply that the more cases retained in case base, the better performance the model because, when a certain category of cases are retained in large numbers, this may influence the system output. This can be compared to the peer pressure from powerful people in the brainstorming method. As a result, when increasing the number of cases, it should always be borne in mind that the distribution of the attributes should be balanced, and the domain coverage test should be carried out when more cases are added.

11.2.2 Limitations of ART*Enterprise

The lack of support for developing the CBR matching and retrieval mechanisms in a programming language and the limited time scale for this research, encouraged the author to utilise a CBR development tool. Several successful leading CBR software packages were evaluated. ART*Enterprise was selected for the development of CBRisk. However, during the process of system development, the author identified certain limitations of the ART*Enterprise development shell.

One major disadvantage of ART*Enterprise development shell was that unlike ReMind, it did not provide case adaptation support. As a result an important functionality, to learn from previous knowledge in the CBR cycle, must be carried out by further programming by the system developer. In terms of integrating an external database to the ART*Enterprise application, ART*Enterprise does not provide a robust performance. When connected to an Excel worksheet it should highlight all the data area, and name this area, to ensure that all data can be transferred to the AE application. When connected to Access, the system also appears unstable when connecting different tables. Errors are sometimes expectedly displayed. Another problem was discovered when performing case matching: the system is not sufficiently stable. When recalling same commands using the same system, some unreasonable errors might emerge. This only can be resolved by exiting and restarting the system. The author expects the later version of ART*Enterprise to address these problems.

Other identified limitations of ART*Enterprise are that the shell does not support dynamic data exchange (DDE) and Dynamic Link Library (DLL). The problems were aggravated as there was no debugging facility available in ART*Enterprise. ART*Enterprise was thus more suitable for initial prototyping, and the fully functional version of CBRisk should be developed using a programming language, such as Borland C++, and Java. Moreover, when developing the model with these programming languages, the case retrieval approach can be further improved by using Generic Algorithms (GA).

11.2.3 Limitations of CBRisk

When this research was undertaken, the development of a CBR system for other construction projects problems was a relatively unexplored area.

The first limitation of CBRisk comes from the knowledge used to develop it. The projects used to construct a knowledge base were restricted to World Bank transportation projects in Asian and Pacific areas. It is impossible, in one thesis, to consider all the attributes of each project. A complete set of project documents usually varies from 300 to over 1000 pages, and the project appraisal usually contains over 100 pages. It is a great deal of work to read through the project documents, pick out relevant key features and build up a knowledge base. Because of restrictions on the data source, only 60 projects were considered and only 45 attributes were selected for the case retrieval calculation.

When exploring the World Bank project database, the author found that since 1997 a standard project appraisal format was used and risk analysis was one of the sections. However, a project appraisal document was not available for all the projects. In some Projects, risk assessment was not carried out during the project appraisal, or had not been well recorded in the project appraisal document. Consequently, only 60 projects were included in the case base.

The projects were based in 11 different countries and the projects in each country were from different sub-sectors, such as highways, railways, urban-transportation etc. Hence, the case matching might be poor in some circumstances due to the insufficient domain coverage. Because the cases from China and India from a higher percentage of the case base, and the projects in these two countries are mainly in the highways and railways sectors, the system currently operates better when coping with cases from China and India. However, with the accumulation of more cases in each domain section, case matching can be improved as well as the model adaptation result. Furthermore, since 2004 many risk assessment in the appraisal document provided long descriptions of the analysis, which made the risk coding more difficult and could influence the model adaptation results.

The second limitation of CBRisk comes from the lack of involvement of relevant experts. Because of the specialization of the data source, there are a limited number of people who are familiar with the World Bank transportation projects. There were even fewer experts who agreed to help with the system evaluation. System validation was mainly based on self evaluation of the system by reusing similar project data sets and statistical analysis. Furthermore, there is no detailed standard to apply to risk analysis in World Bank project appraisal stage. In many circumstance the same risk item is described in various forms in an existing risk analysis section. Risk analysis was carried out in various levels. For the same type of projects, very detailed risk descriptions were provided, whilst for others only a very brief risk list was provided. This may influence the generation of the risk log and allocate relevant risk code for each project, so as to influence the risk adaptation result and finally the system output: identified risks. If more experts could be involved from the outset, from designing the risk register, defining the knowledge representation attributes to finally evaluating the system, it would be a great benefit in developing a better risk identification model within a certain firm or organization.

11.3 RECOMMENDATIONS FOR FUTURE WORK

The author has identified four principle areas of research that would complement and carry forward the research work in this thesis.

It is worth considering how such a system can be deployed in practice. The system would have to be integrated within existing software for major functions, such as Delphi or Java, to provide a better interface, flexible output report and exchange of data in the model. In this situation, careful consideration has to be given to the model development tool. The important factors to consider are the stability, robustness and seamless interoperability with the existing system. The ART*Enterprise development environment used in the present work is good for rapid application development and prototyping but is not appropriate for developing a large-scale operational system. It is more likely that a general purpose programming language and a CBR programming-language-based library would be used. This allows the use of other libraries that support different KBS techniques such as rule-based reasoning, fuzz logic and neural networks to be used for handling relevant aspects of the system. In some circumstances, a combined reasoning is more powerful in handling complex or unusual situations. For instance, if the match score of matching cases were all very low, the results obtained from case-based reasoning would usually not be good enough and some alternative reasoning method such as rule-based reasoning could be used to provide alternative solutions to support decision making.

Further work is being conducted to extend the problem domain to cover other sectors in the World Bank projects. such as generating and verifying a more appropriate risk log for all the World Bank projects; selecting proper project description attributes for all projects and increasing the case base to cover enough cases in each domain sector. The extension of the model into a construction domain rather than transportation is a long-term further development. This would require the development of a knowledge base taken from other sectors, such as infrastructure. This would be a major undertaking, but would almost certainly be necessary for eventual practical use and general availability.

This research develops a generic approach in constructing information and knowledge of World Bank transportation projects. As is apparent from risk identification, the functionality of risk response by using case-based reasoning can also be developed based on methods demonstrated in this thesis. As demonstrated in the literature review, qualitative risk analysis is more important and quantitative risk analysis is used to support decision making in qualitative analysis. The role of risk management for a general project manager usually means the identification of the potential risks and an approach to mitigate them in order to save the project from danger and achieve the target in terms of time, budget and quality. Thus, a case-based risk management system should provide risk identification as well as a risk response strategy. The further work is to integrate risk response to the existing system.

This research has produced some direct applications that can be applied in the construction industry. Construction firms should consider developing and employing their own risk base for improving risk identification. The knowledge developed by this research can be transferred to other firms especially for firms that already have an established risk register system. For these firms the project description attributes are carefully selected and recorded in a standard format based on a risk log. The further work can simply follow the same processes explained in this thesis to integrate the database and case base, define the objects and connections, carry out the adaptation and V&V process, and then built up the whole case-based risk identification system.

Further research would be valuable to attempt a combined reasoning system. The case-based reasoning technique is good for the risk analysis process because it is useful for tasks that are experience-intensive, that lead to inconsistent outcomes, that have incomplete rules to apply, and for which it is hard to acquire domain experience. In the real construction industry, there are many situations where quantitative and detailed information to evaluate uncertainty is not available. Because of the inductive nature of CBR, it is well suited for integration with other reasoning paradigms grounded on more general knowledge such as rule-based, which reasons in a deductive way. The system combining rule- and case-based reasoning technologies will bring a better and more appropriate use of intelligence to a given process or problem. It may learn from previous experience and can continue the learning process as it is used and developed. A rule based system will allow the identification of risks relating to projects from a wide variety of backgrounds where similar problems are faced. Depending on the type of problem, a rule-based system can narrow down the problem domain, and a case-based reasoning system can compare the input data against its case base to find analogous situations. The power of such system can be harnessed to help human experts more quickly solve a complex risk contingency.

REFERENCE

Aamodt, A. (1990). "A computational model of knowledge-intensive problem solving and learning." Current trends in knowledge acquisition, **Wielinga, B., et al.**, Eds., IOS Press, Amsterdam.

Aamodt, A., and Palaz, E. (1994). "Case-based reasoning: foundational issues, methodological variations, and system approaches." *AI communications*, 7(1), 39-59.

Accountants (The Canadian Institute of Chartered Accountants), (2004). "Risk Alert Newsletter." http://www.cica.ca/index.cfm/ci_id/7862/la_id/1.htm.

Acorn, T., and Walden, S. (1992) "Support management cultivated reasoning technology for Compaq customer service." The Proceedings of *In Innovative Applications of Artificial Intelligence 4,* Menlo Park, California, US.

Adrion, W., Branstad, M., and Cherniavsky, J. (1982). "Validation, verification, and testing of computer software." *ACM Computing Surveys*, 14(2), 159-192.

Akintoye, A.S., and MacLeod, M.J. (1997). "Risk analysis and management in construction." *International Journal of Project Management*, 15(1), 31-38.

Al-Bahar, J.F., and Crandall, K. C. (1990). "Systematic Risk Management Approach for Construction Projects." *Journal of Construction Engineering and Management*, 16(3), 533-546.

Al-Harbi, K.M.A. (2001). "Application of the AHP in project management." *International Journal of Project Management*, 19(1), 19-27.

Al-Tabtabai, H., and Alex, A.P. (2000). "Modelling the cost of political risk in international construction projects." *Project Management Journal*, 31(3), 4-13.

Alterman, R. (1989) "panel discussion on case representation." The Proceedings of *DARPA Case-based Reasoning workshop*, San Francisco, USA.

Althoff, K.D., Auriol, E., Barletta, R., and Manago, M. (1995). "A Review of Industrial Case-Based Reasoning Tools." *AI Perspectives Report, AI Intelligence*.

Anonymous (1993). "The Bayesian approach." *Association for Computing Machinery, Communications of the ACM*, 3(11), 74-75.

Anumba, C.J., Ugwu, O.O., Newnham, L., and Thorpe, A. (2002). "Collaborative design of structures using intelligent agents." *Automation in Construction*, 11, 89-103.

Anumba, C.J., Ren, A., Thorpe, A., Ugwu, O.O., and Newnham, L. (2003). "Negotiation within a multi-agent system for the collaborative design of light industrial buildings." *Advances in Engineering Software*, 34, 389-401.

Artto, K.A. (1997). "Fifteen years of project risk management applications- where are we going?" *Managing risks in projects*, **Kahkonen, K. and Artto, K.A.** Eds., E&FN Spon, London.

Ashley, K.D., and Rissland, E.L. (1988) "Waiting on Weighting: A Symbolic Least Commitment Approach." The Proceedings of *AAAI-88, American Association for Artificial Intelligence.*

Association for Project Management (APM) (2004). *Project risk analysis and management guide*, High Wycombe.

Audley, R.J. (1967). *Decision making*, British Broadcasting Corporation, London.

Baker, S., Ponniah, D., and Smith, S. (1999). "Risk response techniques employed currently for major projects." *Construction Management and Economics*, 17(2), 205-213.

Balci, O. (1994). "Validation, Verification, and testing techniques through the life cycle of a simulation study." *Annals of Operations Research*, 53, 121-173.

Baloi, D., and Price, A. (2003). "Modelling global risk factors affecting construction cost performance." *International Journal of Project Management*, 21(4), 261-269.

Bareiss, E.R. (1989). *Exemplar-based knowledge acquisition: a unified approach to concept representation, classification, and learning*, Academic Press, Boston.

Barletta, R., and Mark, W. (1988) "Explanation-based indexing of cases." The Proceedings of *DARPA Case-based Reasoning workshop*, San Francisco.

Barletta, R. (1991). "An Introduction to case-based reasoning." *AI Expert*, 6(8).

Barry, L.J. (1995). "Assessing risk systematically." *Risk Management*, 42, 12-17.

Bateson, G. (1988). *Mind and Nature: A Necessary Unity*, Bantam.

Bellazzi, R., Montani, S., Protinale, L., and Riva, A. (1999) "Integrating rule-based and case-based decision making in diabetic patient management." The Proceedings of *The third*

International Conference on Case-Based Reasoning, ICCBR-99, Seeon Monastery, Germany, 27-30.

Berg, B.L. (2001). *Qualitative research methods for the social sciences*, Allyn & Bacon, Needham Heights, MA.

Bergmann, R., and Wilke, W. (1996). "On the Role of Abstraction in Case-Based Reasoning." Advances in Case-Based Reasoning: Lecture Notes in Artificial Intelligence 1168, Smith, I., and Faltings, B., ed., Springer-Verlag, Berlin.

Birnbaum, L., and Collins, G. (1989) "Reminding and Engineering Design Themes: A Case Study in Indexing Vocabulary." The Proceedings of *DARPA Case-Based Reasoning Workshop*, San Francisco, USA.

Blackler, F. (1995). "Knowledge, knowledge work and organizations: an overview and interpretation." *Organization Studies*, 16(6), 16-36.

Boisot, M. (1998). *Knowledge assets, securing competitive advantage in the information economy*, Oxford University Press, Oxford, UK.

Borner, K. (1995). "Modules for design support," University of Freiberg, Germany.

Botten, N., Kusiak, A., and Raz, T. (1989). "Knowledge bases: integration, verification and partitioning." *European Journal of Operations Research*, 42, 111-128.

Boussabaine, A.H., and Kaka, A.P. (1998). "A neural networks approach for cost flow forecasting." *Construction Management and Economics*, 16(4), 471-479.

Bradtke, S., and L-ehnert, W. (1988) "Some experiments with case-based search." The Proceedings of *Case-based Reasoning Workshop*, San Mateo, CA.

Bratko, I. (1986). *PROLOG programming for Artificial Intelligence*, Addison-Wesley.

Brightware (1999). *ART*enterprise manuals*, Brightware Inc., Novato.

British Standard Institute (1996). *Guide to project management*, Milton Keynes.

British Standard Institute (1996). *BS 8444-3:1996, IEC 60300-3-9:1995 Risk management. Guide to risk analysis of technological systems*, BSI.

British Standard Institute (2000). *BS 6079-3 Project management. Guide to the management of business related project risk*, Milton Keynes.

British Standard Institute (2001). *BS IEC 62198:2001, Project risk management. Application guidelines*, BSI.

Brown, K.A., Willis, P.G., and Prussia, G.E. (2000). "Predicting safe employee behaviour in the steel industry: development and test of a sociotechnical model." *Journal of Operations Management*, 18(4), 445-465.

Campbell, D.T., and Fiske, D.W. (1959). "Convergent and discriminate validation by the multitrait-multimethod matrix." *Psychological Bulletin*, 56, 100-122.

Carbonell, J. (1983). "Learning by Analogy: Formulating and Generalizing Plans From Past Experience." *Machine Learning*, 2.

Carr, V., and Tah, J.H.M. (2001). "A fuzzy approach to construction project risk assessment and analysis: construction project risk management system." *Advances in Engineering Software*, 32, 847-857.

Carrillo, P., and Anumba, C. (2002). "Knowledge management in the AEC sector: an exploration of the mergers and acquisitions context." *Knowledge and process management*, 9(3), 149-161.

Carson, T. (1986). "Closing the gap between research and practice: Conversation as a mode of doing research." *Phenomenology and Pedagogy*, 4(2), 73-85.

Carson, J.S. (1986). "Convincing Users of Model's Validity is Challenging Aspect of Modelers's Job." *Industrial Engineering*, June, 74-85.

Carter, R.L., Doherty, N.A. and London. (1974). *Handbook of risk management*, Kluwer-Harrap Handbooks, London.

Carter, B., Hancock, T., Morin, J.M., and Ro, N. (1996). *Introducing Riskman: the European project management methodology*, London.

Caudill, M., and Butler, C. (1990). *Naturally intelligent systems*, MIT Press., Cambridge.

Champan, C., and Ward, S. (2002). *Managing project risk and uncertainty*, Wiley, Chichester.

Chan, A.P.C., Yung, E.H.K., Lam, P.T.I., Tam, C.M., and Cheung, S.O. (2001). "Application of Delphi method i selection of procurement systems for construction projects." *Construction Management and Economics*, 19(7), 699-718.

Chan, A.P.C., Yung, E.H.K., Lam, P.T.I., Tam, C.M., and Cheung, S.O. (2001). "Application of Delphi method in selection of procurement systems for construction projects." *Construction Management and Economics*, 19(7), 699-718.

Chapman, C.B., and Ward, S.C. (1997). *Project risk management: process, techniques and insights*, John Wiley & Sons, Chichester.

Chapman, R.J. (1998). "The effectiveness of working group risk identification and assessment techniques." *International Journal of Project Management*, 16(6), 333-343.

Checkland, P., and Howel, S. (1998). *Information, systems and information systems, making sense of the field*, Joth Wiley & Sons, UK.

Chen, D., and Hartman, F T (1998). "A neural network approach to risk assessment and contingency allocation." *AACE International Transactions*, 17, 1-6.

Choo, C.W. (1998). *The knowing organization*, Oxford University Press, Oxford, UK.

Choo, C.W., Detlor, B., and Turnbull, D. (2000). *Web work: information seeking and knowledge work on the world wide web*, Kluwer Academic Publishers, Dordrecht.

Clayton, C.R.I. (2001). *Managing geotechnical risk improving productivity in UK building and construction*, The Institution of Civil Engineering and Thomas Telford Ltd, London.

Cohen, L., and Manion, L. (1994). *Research methods in education*, Routledge, London.

Cragun, B.J., and Steudel, H.J. (1987). "A decision-table-based processor for checking completeness and consistency in rule-based expert systems." *Man-Machine Studies*, 26, 633-648.

CRM Manual (1987). *The construction risk management manual*, International Risk Management Institute, Dallas, USA.

Crossland, R., McMhahon, C.A., and Simms Williams, J.H. (1998). "Survey of current practices in managing design risk," University of Bristol.

CSCI (2004). "About RiskPAC." *http://www.csciweb.com/riskpac.htm*

Curet, O. (1995). *The application of case-based reasoning to assist accountants in identifying top management fraud: a study of the problem domain and the methodologies issues in the development, implementation and evaluation of a case-based learning and reasoning tool*, The

Institution of Electrical Engineers, London.

Currie, W.L. (1994). "The strategic management of large scale IT projects in the financial services sector." *New Technology, Work and Employment*, 9(1), 19-29.

Currie, W.L. (2003). "A knowledge-based risk assessment framework for evaluating web-enabled application outsourcing projects." *International Journal of Project Management*, 21, 207-217.

Davenport, T.D. (1997). *Information ecology, mastering the information and knowledge environment*, Oxford University Press, Oxford, UK.

Davenport, T.H., and Prusak, L. (1998). *Working knowledge. How organizations manage what they know*, Harvard business school press, Boston.

David, B.S. (1991) "Principles for case-based representation in a case-based aiding system for lesson planning." The Proceedings of *DARPA Case-based Reasoning workshop*, San Francisco,.

Davis, P.K. (1992). "Generalizing Concepts of Verification." Validation and Accreditation (VV&A) for Military Simulation, RAND, Santa Monica, CA.

Decisioneering, (2004). "Crystal Ball 2000 Standard Edition." http://www.decisioneering.com/crystal_ball/.

Definitive Software, (2004). "Definitive Scenario." http://www.definitivesoftware.com/default.aspx?tabid=30.

Diagnostic Strategies (2001). *Understanding and selecting CBR technology: taxonomy of case-based reasoning technologies.*

Dickson, G.C.A. (1991). *Risk Analysis, second editions*, Witherby, London.

Doyle, M., Hayes, C., Cunningham, P., and Smith, B. (1998). "CBR Net: Smart Technology Over a Network. TCD Technical Report TCD-CS-1998-07," Trinity College Dublin, available at http://www.cs.tcd.ie/publications/tech-reports/trindex.98.html.

Drucker, P. (1969). *The age of discontinuity: guidelines for our changing society*, Harper & Row, New York.

Dutton, D., and Maun, K. (1997). "Case-based reasoning for materials selection in design." *International Journal of Construction Information Technology*, 5(2), 103–111.

Dzeng, R.J., and Tommelein, I.D. (1994) "Case storage of planning knowledge for power plant construction." The Proceedings of *the First Congress Held in Conjunction with A/E/C Systems*.

Edmondson, A. (1999). "Psychological safety and learning behaviour in work teams." *Administrative Science Quarterly*, 44(2), 350-383.

Edum-Fotwe, F.T., Thorpe, A., and McCaffer, R. (2001). "Information procurement practices of key actors in construction supply chains." *European Journal of Purchasing & Supply Management*, 7, 155-164.

Edwards, L. (1995). *Practical risk management in the construction industry*, Thomas Telford, London.

Edwards, P.J., and Bowen, P.A. (1998). "Risk and risk management in construction: a review and future directions for research." *Engineering, Construction and Architectural Management*, 5(4), 339-349.

Egbu, C., Vines, M.P., and Tookey, J. (2003) "The role of knowledge management in e-procurement initiatives for construction organizations." The Proceedings of *the 19th Conference of Association of Researchers in Construction Management*, Brighton, UK, 661-670.

Egbu, C., Hari, S., and Kumar, B. (2003) "Knowledge management practice in small and medium enterprises in the construction industry." The Proceedings of *the 19th Conference of Association of Researchers in Construction Management*, Brighton, UK, 653-660.

Eisehardt, K.M. (1989). "Building theories from case study research." *Academy of Management Review*, 14(4), 532.

Englehardt, J.D. (1997). "Bayesian benefit-risk analysis for sustainable process design." *Journal of Environmental Engineering*, 123(1), 71-79.

Fagan, M., and Corley, S. (1998) "CBR for the reuse of corporate SQL knowledge." The Proceedings of *EWCBR-98*, Berlin, 382-391.

Fagan, M., and Bloor, K. (1999) "Case-based reasoning for candidate list extraction in a marketing domain." The Proceedings of *ICCBR-99*, Berlin, 426-437.

Faltings, B., Hua, K.F., Schmitt, G. and Shih, S.G. (1991) "Case-based representation of architectural design knowledge." The Proceedings of *DARPA Case-Based Reasoning Workshop*, 307–316.

Flanagan, R., and Norman, G. (1993). *Risk management and construction,* Blackwell, Oxford.

Fleming, N. (1996). "Coping with a Revolution: Will the Internet Change Learning? ", Lincoln University, Canterbury, New Zealand.

Flemming, U., Coyne, R., and Snyder, J. (1994) "Case-based design in the SEED system." The Proceedings of *Proceedings of the First Congress held in Conjunction with A/E/C Systems.*

Fong, S. W. (1987). "Risk management." *The Cost Engineering,* 25, 12-16.

Gaines, B. (1991). *The trade-off between knowledge and data in knowledge acquisition,* AAAI/MIT Press, Cambridge, MA.

Gebhardt, F., Voss, A., Grather, W., and Schmidt-Belz, B. (1997). *Reasoning with Complex Cases,* Kluwer Academic Publisher.

Gentner, D. (1983). "Structure-Mapping: A Theoretical Framework for Analogy." *Cognitive Science,* 7(2).

Gero, J.S. (1990). "Design prototypes: a knowledge representation schema for design." *AI Magazine,* 11(4), 26–36.

Goker, M., and Roth-Berghofer, T. (1999). "Development and utilization of a case-based help--desk support system in a corporate environment." Case-Based Reasoning Research and Development, **Althoff, K.D., Bergmann, R., and Branting, L.K.**, Eds., Springer Verlag, Berlin.

Goodman, M. (1989) "CBR in battle planning." The Proceedings of *DARPA Case-based Reasoning workshop,* San Francisco, USA.

Grinnell, J.R.M. (1997). *Social work research and evaluation: quantitative and qualitative approaches, 5th Edn,* Peacock Publishers, Itasca, IL.

Gupta, U. (1991). *Verification and validation of knowledge-based systems,* IEEE Computer Press, New York.

Haimes, Y.Y. (1998). *Risk modeling, assessment, and management,* Wiley, New York.

Hakim, C. (2000). *Research design: successful designs for social and economic research, 2nd Edn,* Routledge, London.

Hammond, K.J. (1987) "Explaining and repairing plans that fail." The Proceedings of *the*

International Joint Conference on Artificial Intelligence.

Hammond, K.J. (1989) "On functionally motivated vocabularies: an apologia." The Proceedings of *DARPA Case-based Reasoning workshop*, San Francisco, USA.

Hartman, F. (1997). "Proactive risk management- myth or reality?" *Managing risks in projects*, **Kahkonen, K., and Artto, K.A.**, Eds., E&FN Spon, London.

Hennessy, D., and Hinkle, D. (1992). "Applying case-based reasoning to autoclave loading." *IEEE Expert*, 7, 21-26.

Hertz, D.B., and Thomas, H. (1983). *Risk analysis and its applications*, John Wiley and Sons, New York.

Ho, S.S.M., and Pike, R.H. (1992). "Adoption of Probabilistic Risk Analysis in Capital Budgeting and Corporate Investment." *Journal of Business Finance and Accounting*, 19(3), 387-405.

Hufschmidt, M.M., James, D.E., Meister, A.D., Bower, B.T., and Dixon, J.A. (1983). *Environment, Natural Systems and Development: An Economic Valuation Guide*, Johns Hopkins University Press, Baltimore.

Inkpen, A. (1996). "Creating knowledge through collaboration." *California Management Review*, 39(1), 123-140.

Institution of Civil Engineers and the Faculty and Institute of Actuaries (1998). *RAMP: risk analysis and management for projects*, Thomas Telford, London.

Institution of Civil Engineers and the Faculty and Institute of Actuaries (2000). *RAMP: risk analysis and management for projects*, Thomas Telford, London.

Instrument and Country (2005). "Methodological notes on estimation of economic rates of return."

International Security Technology, (2004). "About CORA." http://www.ist-usa.com/aboutcora.htm.

Investopedia, (2005). "Commitment fee."
http://www.investopedia.com/terms/c/commitmentfee.asp.

Jick, T.D. (1979). "Mixing qualitative and quantitative methods: triangulation in action."

Administrative Science Quarterly, 24(4), 602-611.

Kangari, R. (1988). "Construction risk management." *Civil Engineering System*, 5, 114-120.

Katmar Software, (2004). "PRA: Project Risk and Contingency Analysis." http://www.katmarsoftware.com/pra.htm.

Kerlinger, F.N. (1986). *Foundations of behavioral research, 3rd Ends*, Harcourt Brace College Pubs, London.

Kermally, S. (2002). *Effective knowledge management: a best practice blueprint*, John Wiley & Sons, Chichester.

Kleijnen, J.P.C. (1995). "Verification and validation of simulation models." *European journal of Operational Research*, 82(1), 145-162.

Klein, M.R., and Methlie, L.B. (1995). *Knowledge-based Decision Support Systems with Application in Business, 2nd Edn*, John Wiley & Sons, New York.

Klerfors, D., (1998). "Artificial Neural Networks." http://hem.hj.se/~de96klda/NeuralNetworks.htm.

Knight, F.H. (1921). *Risk, uncertainty and profit*, Houghton Mifflin, Boston.

Kolodner, J.L. (1983). "Maintaining organization in dynamic Long-term memory." *Cognitive Science*, 7(4), 243-280.

Kolodner, J.L. (1992). "The role of experience in natural problem solving," Georgia Institute of Technology, Atlanta.

Kolodner, J.L. (1993). *Case-based Reasoning*, Morgan Kaufman, San Mateo.

Kolonder, J.L. (1983). "Reconstructive memory: a computer model." *Cognitive Science*, 7(4), 281-328.

Koton, P. (1989). "Using experience in learning and problem solving," Massachusetts Institute of Technology, Massachusetts.

Kumar, R. (1996). *Research methodology: a step-by-step guide for beginners*, SAGE, London.

Kumaraswamy, M.M., and Thorpe, A. (1997). "A computerized construction project management evaluation system." *Advances in Engineering Software*, 25, 197-206.

Lebowitz, M. (1987). "Experimental with incremental concept formation: UNIMEN." *Machine Learning*, 2(2), 103-138.

Leedy, P.D. (1989). *Practical research: planning and design*, Collier Macmillan, London.

Leunga, H.M., Chuaha, K.B., and Tummala, V.M.R. (1998). "A knowledge-based System for identifying potential project risks." *International Journal of Management Science*, 26(5), 623-638.

Lewis, L., et al., (1992). "A case-based reasoning solution to the problem of redundant engineering in large scale manufacturing." *International Journal of Expert Systems*, 4(2), 189-200.

MacroSystem, (2004). "Risk Analyzer." http://www.add-ins.com/analyzer/

Maher, M.L., and Balachandran, B. (1994). "Multimedia approach to case-based structural design." *The journal of computing in civil engineering*, 8(3), 359-376.

Maher, M.L., Balachandran, B., and Zhang, D.M. (1995). *Case-Based Reasoning in Design*, Lawrence Erlbaum Associates, USA.

Maher, M.L., and Garza, A.G. (1996). "Developing case-based reasoning for structural design." *IEEE Expert*, 42–52.

Maqsood, T., Finegan, A.D. and Walker, D.H.T. (2003) "A conceptual model for exploring knowledge canalization from sources of innovation in construction organizations: extending the role of knowledge management." The Proceedings of *The 19th Conference of Association of Researchers in Construction Management*, Brighton, U.K., 613-621.

Marcot, B. (1987). "Testing your knowledge-base." *AI-Expert*, August(1987), pp43-47.

Meredith, J. (1998). "Building operations management theory through case and field research." *Journal of Operations Management*, 16(4), 441-454.

Merna, T. (2000). "Risk management of an agricultural investment in a developing country utilising the CASPAR program." *International Journal of Project Management*, 18(5), 349-360.

Merna, T., and Njiru, C. (2002). *Financing infrastructure project*, Thomas Telford, London.

Miles, M.B., and Huberman, A.M. (1994). *Qualitative data analysis, 2nd Edn*, Sage, Thousand Oaks, CA.

Mitchell, T.M. (1997). *Machine Learning*, McGraw-Hill, New York.

Moore, C. J., Lehane, M.S., and Price, C.J. (1994) "Case-Based Reasoning for Decision Support in Engineering Design." The Proceedings of *the IEE Colloquium on Case-Based Reasoning: Prospects for Applications*, 4/1-4/4.

Moore, C. J., and Lehane, M.S. (1999). "Development of a Case Representation Strategy for a Bridge Design Case Base." *Engineering Structures*, 21(3), 219-231.

Morell, L. J. (1988) "Use of meta knowledge in the verification of knowledge-based systems." The Proceedings of *IEA-AIE*, 847-857.

Morgan, M.G. (1990). "Choosing and managing technology-induced risks." Readings in risk, Glickman T. S. and Gough, M., ed., Resources for the Future, Washington.

Mukherjee, A., and Deshpande, J.M. (1995). "Modeling Initial Design Process Using Artificial Neural Networks." *ASCE Journal of Computing in Civil Engineering*, 9(3), 194-200.

Musgrove, P.B., Davies, J., and Izzard, D. (1996) "A comparison of nearest neighbour, rule induction and neural networks for the recommendation of treatment at anticoagulant out-patient clinics." The Proceedings of *the second UK workshop on Case-based Reasoning*, UK: Salford University.

Mustafa, M.A., and Al-Bahar, J.F. (1991). "Project risk assessment using the analytic hierarchy process." *IEEE Transactions on Engineering Management*, 38(1), 46-52.

Navinchandra, D. (1988) "Case-based reasoning in CYCLOPS, a design problem solver." The Proceedings of *DARPA Case-based Reasoning workshop*, San Francisco.

Navinchandra, D. (1991). *Exploration and innovation in design: towards a computational model*, Spinger Verlag, New York.

Nayani, N., and Mollaghesemi, M. (1998) "Validation and verification of the simulation model of a photolithography process in semiconductor manufacturing." The Proceedings of *the 1998 Winter Simulation Conference*, 1017-1022.

Nazari, A. (1999). "*Decision support system for assessment of the viability of super project*," University of Leeds, Leeds.

Ng, S.T. (1996). "Case-based Reasoning decision support for contract prequalification," University of Manchester Institute of Science and Technology (UMIST), Manchester, England.

Ng, S.T., Smith, N.J., and Skitmore, R.M. (1998). "A case-based reasoning model for contractor prequalification." *International Journal of Construction Information Technology*, 6(1), 47-61.

Nonaka, I. (1991). "The knowledge-creating company." *Harvard Business Review*, 69(6), 96□104.

Nonaka, I., and Takeuchi, H. (1995). *The knowledge-creating company: how Japanese companies create the dynamics of innovation*, Oxford University Press, Oxford.

O'Leary, D.E. (1993). "Verification and validation of case-based systems", *Expert Systems with Applications: An International Journal*, 6(1), 57-66.

Palisade Corporation, (2004). "@Risk 4.5 for Excel." http://www.palisade-europe.com/html/risk.html.

Patterson, F.D., and Neailey, K. (2002). "A risk register database system to aid the management of project risk." *International Journal of Project Management*, 20, 365-374.

Pearce, M., Goel, A.K., Kolodner, J., Zimring, C., Sentosa, L., and Billington, R. (1992). "Case-based decision support: a case study in architectural design." *IEEE Expert*, 7(5), 14-20.

Pender, S. (2001). "Managing incomplete knowledge: Why risk management is not sufficient." *International Journal of Project Management*, 113(2), 79-87.

Perry, J.G., and Hayes, R. W. (1985). "Risk and its Management in Construction Projects." *Proceedings of Institute of Civil Engineers*, 78, 499-521.

Perry, J.G. (1986). "Dealing with risks in contract." *Building Technology and Management*, 1998(April), 23-26.

Pidd, M. (1992). *Computer simulation in management sciences,3rd Edn*, Wiley, Chichester.

Pidd, M. (1998). "Computer simulation for operational research in 1984." Developments in operational research, **Eglese R.W. and Rand, G. K.**, Eds., Perganon press, Oxford.

Planning & Markets (2005). "Economic and Financial Assessment."

Pouliquen, L.Y. (1970). "Risk Analysis in Project Appraisal." *World Bank Staff Occasional.*

Primavera, (2004). "Monte Carlo 3.0 for Primavera, project risk analysis software." http://www.primavera.com/products/p3_montecarlo.html.

Probst, G., Raub, S., and Romhardt, K. (1999). *Managing knowledge: building blocks for success*, John Wiley & Sons, Chichester.

Proha Group, (2004). "Futura Software." http://www.futurarisk.com/.

Project Management Institute (PMI) (2000). *A guide to the project management book of knowledge*, Project Management Institute, Pennsylvania, USA.

Quigley, E. J., and Debons, A. (1999) "Interrogative theory of information and knowledge." The Proceedings of *SIGCPR '99*, New Orleans, USA, 4-10.

Racine, K., and Yang, Q. (1996) "Maintaining Unstructured Case Base." The Proceedings of *International Conference of Case-based Reasoning (ICCBR 1997)*, 553-564.

Raftery, J. (1994). *Risk analysis in project management*, E&FN Spon, London.

Rajamoney, S.A., and Lee, H.Y. (1991) "Prototype-based reasoning: an integrated approach to solving large novel problems." The Proceedings of *Ninth National Conference on Artificial Intelligence: AAAI-91*, 34-39.

Riesbeck, C., and Schank, R. (1989). *Inside Case-based Reasoning*, Lawrence Erlbaum.

Risk+ for MS Project (2004).

Rivard, H., and Fenves, S.J. (2000). "SEED-Config: A case-based reasoning system for conceptual building design." *Artificial Intelligence for Engineering Design, Analysis, and Manufacturing*, 14, 415-430.

Robinson, S. (1997) "Simulation model verification and validation: increasing the users confidence." The Proceedings of *the 1997 Winter Simulation Conference.*

Roland, C., and Low, A. (1993). "Risk Analysis in Project Planning.- A simple spreadsheet application using Monte-Carlo Techniques." *Project Appraisal*, 8(3).

Rowe, W.D. (1988). *An anatomy of risks*, Robert E. Krieger Publishing Co., Malabar, USA.

Royal Society (1992). *Risk: analysis, perception and management, Report of Royal Society Group*, Royal Society, London.

Ruby, D., and Kibler, D. (1988) "Exploration of case-based problem solving." The Proceedings of *Case-based Reasoning Workshop*, San Mateo, CA, 345-356.

Sargent, R.G. (1996) "Subjective validation methods using graphical displays of data." The Proceedings of *the 1996 Winter Simulation Conference, IEEE.*

Sargent, R.G. (2000) "Verification, validation, and accreditation of simulation models." The Proceedings of *the Winter IEEE simulation conference*, IEEE, Piscataway, NJ, 50--59.

Scarbrough, H. (1999). "Knowledge as work: conflicts in the management of knowledge." *Technology Analysis and Strategic Management*, 11(1), 5-16.

Schank, R.C., and Abelson R.P. (1977). *Scripts, plans, goals, and understanding: an inquiry into human knowledge structures*, Erlbaum Associates, Hillsdale.

Schmitt, G.N. (1993). "Case-based reasoning in an integrated design and construction system." *International Journal of Construction Information Technology*, 1(3), 31–51.

Scott_morton, M.S. (1984). "The state of art of research." The information system research challenge, McFarlan, F. W., ed., Harvard Business School Press, 13-41.

Shank, R.C. (1982). *Dynamic memory: a theory of learning in computers and people*, Cambridge University Press.

Shi, J.J. (2002). "Three methods for verifying and validating the simulation of a construction operation." *Construction Management and Economics*, 6, p483-491.

Shih, S.G. (1991) "Case-based representation and adaptation in design." The Proceedings of *CAAD futures*, Braunschweig, 301.

Sianipar, P.R.M., and Adams, T.M. (1997). "Fault tree model of bridge element deterioration due to interaction." *Journal of Infrastructure System*, 3(3), 103-110.

Simister, S.J. (1994). "Usage and benefits of project risk analysis and management." *International Journal of Project Management*, 12, 5-8.

Simpson, R.L. (1985). "A computer model of case-based reasoning in problem solving: An investigation in the domain of dispute mediation," School of Information and Computer Science.

Smith, N. J. (1998). *Managing risk in construction project*, Blackwell Science, Oxford.

Smith, N. J. (2003). *Appraisal, risk and uncertainty*, Thomas Telford, London.

SoftBase, (2004). "BDSS (Bayesian Decision Support System) 2.0."

http://static.highbeam.com/s/softbase/november302001/bdssbayesiandecisionsupportsystem20/.

Soibelman, L., and Peña-Mora, F. (1999) "M-RAM: An Artificial Intelligence Distributed System to Support the Conceptual Phase of Structural Design." The Proceedings of *Civil-Comp99-The Seventh International Conference on Civil and Structural Engineering Computing*, Oxford, England, 13-15.

Sommerville, I. (1996). *Software Engineering*, Addison-Wesley, Workingham, England.

Spek, R. V. D., and Spijkervet, A. (1997). *Knowledge management: dealing intelligently with knowledge*, CIBIT, Utrecht.

Steelhammer, J., and Mazzoni, P. (2002). "*ART*Enterprise 10: a discussion of key capabilities*." Mindbox Inc., California.

Stenmark, D. (2002) "Information vs. knowledge, the role of intranets in knowledge management." The Proceedings of *the 35th Hawaii International Conference on System Sciences*, 1-10.

Stolpmann, M., and Wess, S. (1998). *Intelligent System for E-Commerce and Support*, Addison Wesley.

Stone, J.R., Blockley, D.I., and Pilsworth, B.W. (1990). "Managing risk in civil engineering by machine learning from failures." *IEE Colloquium on Machine Learning*, 1-4.

Strauss, A.L., and Corbin, J.M. (1998). *Basics of qualitative research: techniques and procedures for developing grounded theory, 2nd Edn*, Sage Publications, Thousand Oaks, CA.

Strewart, T. A. (1997). *Intellectual capital*, Nicolas Brearly.

Stroulia, E., Shankar, M., Goel, A., and Penberthy, L. (1992) "A Model-Based Approach to Blame Assignment in Design." The Proceedings of *the Second International Conference on AI in Design, AID'92*, Pittsburgh, 519-537.

Swan, J., Scarbrough, H., and Preston, J. (1999) "Knowledge management□the next fad to forget people?" The Proceedings of *European Conference on Information Systems (ECIS)*, Copenhagen, 23-25.

Syncopation Software (2004). "DPL Professional."
http://www.syncopationsoftware.com/products.html

Tah, J.H.M., Thorpe, A. and McCaffer, R. (1993). "Contractor project risks contingency allocation using linguistic approximation." *Computer System Engineering*, 4(2), 281-293.

Tah, J.H.M., Carr, V., and Howes, R. (1998). "An application of case-based reasoning to the planning of highway bridge construction." *Engineering, Construction and Architectural Management*, 5(4), 327–338.

Tah, J.H.M., Carr, V., and Howes, R. (1999). "Information modelling for case-based construction planning of highway bridge projects." *Advances in Engineering Software*, 30(7), 495-509.

Tecolote Research Inc. (2004). "Risk analysis." *http://www.tecolote.com/Services/RiskAnalysis.htm*

The Association of Project Management (2002). "Project risk analysis and management." *A guide by APM (compiled by Norris, C., Perry, J. and Simon, P.).*

The Institute of Internal Auditors, (2004). "Assessing Business Risk for Internal Auditors updated!" http://www.theiia.org/training/index.cfm?act=seminar.detail&semID=92.

Thompson, P., and Perry, J. (1992). *Engineering construction risk*, Thomas Telford, London.

Tiwana, A. (2000). *The knowledge management toolkit: practical techniques for building a knowledge management system*, Prentice Hall, London.

Tiwana, A. (2001). *The essential guide to knowledge management: e-business and CRM applications*, Prentice Hall, Upper Saddle River.

Toakley, A.R., and Ling, S.M.C. (1991) "Risk management and the building procurement process." The Proceedings of *Innovation and Economics in Building Conference*, Brisbane, Australia, 63-67.

Treasury, H. (2000). *Management of risk- a strategic overview*, HMSO, London.

Ugwu, O.O., Anumba, C.J., and Thorpe, A. (2004). "The development of cognitive models for constructability assessment in steel frame structures." *Advances in Engineering Software*, 35, 191-203.

Ugwu, O.O., Anumba, C.J., and Thorpe, A. (2005). "Ontological foundations for agent support in constructability assessment of steel structures- a case study." *Automation in Construction*, 14, 99-114.

Uher, T.E. (1993) "Risk management in the building industry." The Proceedings of *the 1993 Australian Institute of Project Management Conference*, Coolum.

Uher, T.E., and Toakley, A.R. (1999). "Risk management and the conceptual phase of project." *International Journal of Project approximation*, 17(3), 161-169.

Vollrath, I. (1998) "Reuse of Complex Electronic Designs: Requirements Analysis for a CBR Application." The Proceedings of *the 4th European Workshop, EWCBR-98, Advances in Case-Based Reasoning*, Dublin, Ireland, 136-147.

VTT (2004). "Temper System for project risk management." *http://cic.vtt.fi/eds/temper.htm*

Walliman, N. (2001). *Your research project, Step by step guide for the first-time researcher*, Sage Publications, London.

Ward, S.C., Chapman, C.B., and Curtis, B. (1991). "On the allocation of risk in construction project lifecycle." *International Journal of Project Management*, 9(3), 140-147.

Ward, S. (1999). "Assessing and Managing important risks." *International Journal of Project Management*, 17, 331-336.

Watson, I., and Marir, F. (1994). "Case-based Reasoning: a review." *The Knowledge Engineering Review*, 9(4), 335-381.

Watson, I. (1997). *Applying Case-Based Reasoning: techniques for Enterprise Systems*, Morgan Kaufmann.

Watson, I. (2001) "Knowledge Management and Case-Based Reasoning: A Perfect Match?" The Proceedings of *FLAIRS Conference*, http://www.cs.auckland.ac.nz/~ian/papers/cbr/cbr_perfect.pdf, 118-122.

Webb, A. (2003). *The project manager's guide to handling risk*, Aldershot: Gower.

Welcome Software (2004). "Open Plan Connect Overview."

Wiig, K.M. (1993). *Knowledge management foundations: thinking about thinking - how people and organizations create, represent, and use knowledge*, Schema Press, Arlington.

Williams, T.M. (1993). "Risk-management infrastructures." *International Journal of Project Management*, 11, 5-10.

Williams, T.M. (1994). "Using a risk register to integrate risk management in project definition." *International Journal of Project Management*, 12(1), 17-22.

World Bank (1998). *China-Fujian Second Highway Project Information Document.*

World Bank (1999). *China-Fujian second highway Project Appraisal Document.*

World Bank, (2005). http://www.worldbank.org.

World Bank, (2005). "Lending instrument."
http://siteresources.worldbank.org/PROJECTS/Resources/lendinginstrumentbrochure.pdf.

Yang, S., and Robertson, D. (1994) "A case-based reasoning system for regulatory information." The Proceedings of *the IEE Colloquium on Case-Based Reasoning: Prospects for Applications*, 3/1-3/3.

Yin, R.K. (1994). *Case study research: design and methods, 2nd Edn*, Sage Publication, Newbury Park, CA.

Zeigler, B.P. (1986). *Multifaceted modelling and discrete event simulation*, Academic Press, London.

Zoysa, S.D., and Russell, A.D. (2004) "Re-use of knowledge in risk management- finding a competitive advantage." The Proceedings of *the First International Conference of World of Construction Management*, Toronto, Canada, 387-400.

Appendix A: Project List in Case Base

Project ID	Project Name	Country
P078284	Emergency Transport Rehabilitation Project	Afghanistan
P009524	Dhaka Urban Transport	Bangladesh
P037294	Road Rehabilitation and Maintenance Project (03)	Bangladesh
P071435	Rural Transport Improvement Project	Bangladesh
P059481	Rural Access Project	Bhutan
P071207	Provincial and Rural Infrastructure Project (PRIP)	Cambodia
P050036	Anhui Provincial Highway Project	China
P003653	China Container Transport Project	China
P077137	Fourth Inland Waterways Project	China
P051705	Fujian Highway Project (02)	China
P058843	Guangxi Highway Project	China
P003614	Guangzhou City Center Transport Project	China
P081749	Hubei Shiman Highway Project	China
P070441	Hubei Xiaogan-Xiangfan Highway Project	China
P003619	Inland Waterways Project (02)	China
P070459	Inner Mongolia Highway Project	China
P041890	Liaoning Urban Transport Project	China
P041268	National Highway Project (04)	China
P058846	National Railway Project	China
P076714	Second Anhui Highway Project	China
P058845	Second Jiangxi Highway Project	China
P075602	Second National Railways Project (Zhe-Gan Line)	China
P056596	Shijiazhuang Urban Transport Project	China
P058844	Third Henan Provincial Highway Project	China
P045788	Tri - Provincial Highway Project	China
P045915	Urumqi Urban Transport Improvement Project	China
P069852	Wuhan Urban Transport Project	China
P058847	Xinjiang Highway Project (03)	China
P073776	Allahabad Bypass Project	India
P071244	Grand Trunk Road Improvement Project	India
P010566	Gujarat State highway Project	India
P070421	Karnataka State Highways Improvement Project	India
P072539	Kerala State Transport Project	India

P077856	Lucknow-Muzaffarpur National Highway Project	India
P069889	Mizoram State Roads Project	India
P050668	Mumbai Urban Transport Project	India
P077977	Rural Roads Project	India
P050649	Tamil Nadu Road Sector Project	India
P009972	Third National Highways Project	India
P067606	Uttar Pradesh State Roads Project	India
P040578	Eastern Indonesia Region Transport Project	Indonesia
P074290	Second Eastern Indonesia Region Transport Project	Indonesia
P003993	Sumatra Regional Roads	Indonesia
P042237	Provincial Infrastructure Project	Lao
P083543	Road Maintenance Program (Phase 2)	Lao
P056200	Transport Development Project	Mongolia
P045052	Road Maintenance and development	Nepal
P010556	Highways Rehabilitation	Pakistan
P004397	Road Maintenance and Rehabilitation Project	Papua New Guinea
P039019	First National Roads Improvement and Management Project	Philippines
P057731	Metro Manila Urban Transport Integration Project	Philippines
P075523	Second Infrastructure Asset Management Project	Samoa
P075173	Highways Management Project	Thailand
P042927	Mekong Transport and Flood Protection Project	Vietnam
P059663	Road Network Improvement Project	Vietnam
P004833	Urban Transport Improvement Project	Vietnam
P059864	VN-2nd Rural Transport Project	Vietnam
P009995	Andhra Pradesh State Highway Project	India
P064821	Road Maintenance Project	Lao
P052293	Infrastructure Asset Management Project	Samoa

Appendix B: Risk Log for CBRisk

Code	Risk	Category
101	Delays in procurement	Market
102	Delays in award of contracts	Market
103	Delays in disbursement	Market
104	Delays in contract implementation	Market
105	Procurement inadequate competition for works and services.	Market
106	Procurement capacity is inadequate	Market
107	Procurement procedures not applied transparently	Market
108	Procurement is not carried out in conformance with Bank Guidelines	Market
109	Insufficient accountability of the use of internal and external resources	Communication
110	Counterpart resources available in timely manner	Communication
111	Inadequate counterpart funding	Finance
112	Counterpart funding is not available in a timely manner	Finance
113	Inadequate maintenance funding provisions	Finance
114	Construction materials and human resources not available on time	Finance
115	Fiscal strain leading to an inability to provide project and counterpart funding, delaying implementation and raising costs	Finance
116	Funds are not managed effectively and transparently	Finance
117	Delays in submission of financial reports	Finance
118	Delays in transfer of funds	Finance
119	Credit is not available for local contractors	Finance
120	No separate bank accounts for different projects	Finance
121	Provincial governments allocate adequate funding for roads from block grants	Finance
122	Rural finance program does not provide funds for services and people in rural areas	Finance
123	Major archeological finds during construction could delay information and increase cost of the project	Finance
124	External and local funding is inadequate	Finance
125	Lack of capacity to maintain and operate the existing and new assets and plan new/major capital investments.	Finance
126	Possibility of inappropriate procurement and misuse of funds	Finance
127	Cost savings not passed on to users	Finance
128	Price control committees delay or do not approve proposed toll rates which could impact the ability to repay the Bank loan.	Finance
129	Not use maintenance funds allocated for physical works efficiency	Finance
130	Unit costs and level of latent defects are not borne out in tendering and execution for periodic maintenance works.	Finance
131	Effective coordination is not established among project agencies	Industry

132	Cooperation between the various stakeholders of the project does not continue	Industry
133	Inability to recruit competent works supervisors	Communication
134	Human resources are inadequate, both in number and quality	Communication
135	Staff resist change and organizational restructuring	Communication
136	Opportunities are opened to trainees to apply new skills in workplace	Communication
137	Inadequate supply of competent consultants, contractors	Communication
138	Trained staff or personnel assigned to project do not perform well.	Communication
139	Absence of implementation of measures to enhance human resource management in areas of transfers, skills specialization and personal accountability in rural road agencies.	Communication
140	Quality of construction will not comply with required standards	Construction
141	Newly reconstructed and improved roads are not maintained	Construction
142	Non-government stakeholders not having a say in decision making pertaining to road network maintenance	Construction
143	Deficient engineering design	Construction
144	Design of equipment is not as expected	Construction
145	Unrealistic cost estimates	Construction
146	Delays in necessary land acquisition / site readiness	Social
147	Delays in resettlement	Social
148	Delays in completing in implementing RAP/EMP/IPDP	Social
149	Delays in approval of project-related government policies prior to award of civil works contracts	Social
150	Interference of local people in acquisition of land could lead to delays in implementation and resettlement cost overrun.	Social
151	Legislation for major improvements to land registration and titling not authorized	Social
152	Mines and unexploded ordnance are not effectively cleared from the project areas.	Social
153	Strategic location become contested or subject to fighting	Social
154	Bank resource for supervision is inadequate	Communication
155	Unfamiliarity with the Bank procedures could delay implementation	Communication
156	Trained staff lack the opportunity to apply acquired skills in the workplace	Communication
157	Training program irrelevant to needs and objectives	Communication
158	Staff will not use new skills	Communication
159	Lack of experience in implementing	Communication
160	Low staff skills to implement the new management system	Communication
161	Inadequate information sharing and skill transfer	Communication
162	Problems with contractor or consultants performance	Construction
163	Implementation delays	Construction
164	Delays in construction	Construction
165	Sufficient electric power may not be available for electric traction.	Construction

166	Inadequate maintenance	Construction
167	Weak ownership and support of institutional strengthening action plan	Technology
167	Policy and Institutional Reform will not receive large ownership and will be unsuccessful	Technology
168	Problems with institutional capacities to execute their components in a timely manner	Technology
169	No timely implementation of institutional reforms	Technology
170	Reform action plans are not sufficiently realistic and applicable.	Technology
171	Does not continue to receive political support for reform	Technology
172	Does not continue to give priority to reform	Technology
173	Ineffective implementation of broader civil service reforms	Technology
174	Data generated by expanded environmental monitoring system is not used by municipal leaders to make more reformed decisions on mitigation measures	Technology
175	Traffic management system not enforced	Transportation
176	Future congestion is not mitigated by demand management measures	Transportation
177	Newly increased traffic capacity is not efficiently managed and maintained.	Transportation
178	Bus operators do not make enough use of communications and information systems and priority measures to improve services.	Transportation
179	Weak traffic regulation enforcement	Transportation
180	Policy guidelines for vessel monetization are not established	Transportation
181	Safety actions is weak	Transportation
182	Road user charges not implemented or structured to achieve substantial cost recovery	Transportation
183	Traffic grows at lower than expected rates on toll roads and toll roads attract less than expected traffic volumes from existing parallel roads, resulting in underutilization of toll roads, lower economic benefits and difficulty in repaying Bank loan.	Transportation
184	Corruption in procurement of civil works, goods and consulting services is controlled	Transportation
185	Continued traffic congestion due to higher than expected traffic growth.	Transportation
186	Inadequate priority continues to be given to the provision of transport infrastructure and services	Transportation
187	Public transport is not able to expand services to rural areas	Transportation
188	Inefficient operation of the traffic management system	Transportation
189	Inefficient Road Network Development	Transportation
190	No improvement in enforcement of axle-load control over network	Transportation
191	Unstable situation and competition in the transport sector	Transportation
192	Railway subsidy will be increased.	Transportation
193	Selection of incompetent contractors	Communication

194	Inadequate management capacity at all levels	Communication
195	Sustained commitment from GOl agencies to Project development objectives.	Communication
196	Reorganization not disruptive, instead taken as an opportunity to improve efficiency.	Communication
197	Road transport industry has sufficient incentive to perform more efficiently	Communication
198	Project processes lack transparency	Communication
199	Inadequate supervision and weak leadership/ownership	Communication
200	Poor law and order	Communication
201	Does not continued commitment to implement TA's final recommendations	Communication
202	Weak competition for tendered works and services	Communication
203	Vehicle emissions control strategy is not enforced	Communication
204	Environment Management lack of execution of continue policy	Communication
205	lack of local government support for proposed works	Political
206	Political pressures interfere with implementation and adversely affect the quality of construction	Political
207	Unstable internal political situation in the country	Political
208	Political influence continues to project finance	Political
209	Political will and internal leadership to sustain reforms	Political
210	Natural disasters delay or increase the cost of road works	Environment
211	Undue external pressures on MCTPC and DCTPC	Political
212	Vessel size conversion with increased size is not realized	Environment
213	Lack of stakeholder and private sector interest and support	Communication
214	Driver safety behavior does not improve	Transportation
215	contractors give inadequate attention to safeguard issues or mitigating actions; awareness and capacity are also inadequate.	Construction
216	Overall macroeconomic instability disturbs public spending on capital investment	Communication
217	Not implement some of the agreed policy prescriptions and enforce regulations, Government contracting policies not adjusted or improved	Legal
218	There is long ship lock waiting time.	Transportation
219	Revenue potential from power sales is impaired by delayed implementation of power sector restructuring	Financial
220	Irrelevant selection of black spots	Transportation
221	Ship operators fail to invest in large vessels.	Transportation
222	Inadequate measures to apply and enforce risk management plans	Management
223	Shock waves caused by high-speed trains in tunnels.	Environment
224	Business environment is not conducive to the survival of small contractors	Environment
225	Inadequate project objective plan	Management
226	Delayed reports	Management

Appendix C: Features Selection Questionnaire

1. Personal Information	
Name	
Email	
Occupation	
Company/Organization/Institute	
Country	
Experience in World Bank Projects	
If other, please specify	
In which Sector	
Total No. of projects involved in	

2. Attribute Evaluation

The following attributes have been picked up from World Bank project appraisal documents. Suppose you want to identify risks in a new project by referencing a previous project's risk analysis information, which attributes and at what level, do you think would contribute best in selecting similar projects? Please rank attributes from 0-10 (0 means no contribution, 10 represents significant contribution).

Note: Please ignore the co-relationship among features. For example, the total project cost is the sum of the other Finance Plan components.

Attribute	Score	Attribute	Score
Region		Main Loan Conditions	
Country		Implementation Period	
Borrower		Economical Analysis Method	
Guarantor		Financial Net Present Value	
Commitment Fee		Financial Internal Rate of Return	
Service Charge		Front End Fee of Bank Loan	
Sector		World Bank Support Amount	
Sub Sector		Economical Internal Rate of Return	
Grace Period		Economical Net Present Value	
Target Population		Environmental Category	
Proposed Terms		Environment Mitigation Measurement	
Theme		Potential Environment Impact	
Effective Date		Institutional Support	
Completion Status		Procurement Issue	
Team Leader		Resettlement of People	
Year to Maturity		Financial Plan Total	
Finance Source		Finance Management Issues	
Lending Instrument		Financial Plan Government	
Land Acquisition		Financial Plan IBRD	
Shareholder		Financial Plan IDA	
Beneficiaries		Financial Plan Other	
Sustainability		Project Objective Category	
Technical Issue		Program of Target Intervention	

3. Please indicate any other attributes that you think should also be considered, along with significant level.
4. At what level do you think the risk identification is accurate in the Project Appraisal Document? Please indicate with percentage.
5. Commend

Thank you very much for your cooperation, please send the complete questionnaire to: cenyt@leeds.ac.uk

Should you have any questions regarding the questionnaire or the research, please feel free to contact me by email to the address above.

You cooperation is greatly appreciated! It will be possible to share the results of this research when it is completed.

Appendix D: Correlation Analysis for Case Retrieval Attributes

Attribute		Grace Period	Year to Maturity	Government Committee	Implementation Duration	FNPV	FIRR	ENPV	EIRR	Land Acquisition	Resettlement People	Year	% Transportation	IBRD Commitment	IDA Commitment	Total Project Cost	Finance Source
Grace Period	Pearson Correlation	1.000	0.984	-0.324	-0.282	.	.	-0.220	0.026	-0.234	.	0.153	-0.407	-0.494	0.874	-0.409	0.291
	Sig. (2-tailed)	.	0.000	0.122	0.182	.	.	0.325	0.909	0.334	.	0.474	0.048	0.014	0.000	0.047	0.168
Year to Maturity	Pearson Correlation	0.984	1.000	-0.312	-0.268	.	.	-0.203	-0.023	-0.185	-0.112	0.171	-0.402	-0.462	0.864	-0.378	0.297
	Sig. (2-tailed)	0.000	.	0.138	0.206	.	.	0.366	0.920	0.448	0.639	0.423	0.051	0.023	0.000	0.069	0.158
Government Committee	Pearson Correlation	-0.324	-0.312	1.000	-0.083	-0.127	-0.279	0.507	-0.366	0.782	0.214	0.156	0.259	0.306	-0.263	0.836	-0.527
	Sig. (2-tailed)	0.122	0.138	.	0.530	0.652	0.314	0.007	0.061	0.000	0.305	0.234	0.046	0.017	0.043	0.000	0.000
Implementation Duration	Pearson Correlation	-0.282	-0.268	-0.083	1.000	0.632	0.130	0.046	-0.159	-0.480	-0.563	-0.381	-0.015	0.043	-0.131	-0.050	-0.267
	Sig. (2-tailed)	0.182	0.206	0.530	.	0.011	0.645	0.819	0.428	0.017	0.003	0.003	0.908	0.746	0.317	0.706	0.039
FNPV	Pearson Correlation	.	.	-0.127		1.000	0.251	0.179	0.002	0.170	0.241	-0.465	0.319	0.475	.	0.217	-0.248
	Sig. (2-tailed)	.	.	0.652	0.011	.	0.367	0.524	0.993	0.578	0.428	0.081	0.246	0.074	.	0.437	0.373
FIRR	Pearson Correlation	.	.	-0.279	0.130	0.251	1.000	0.039	0.801	-0.192	0.418	-0.789	0.533	-0.075	.	-0.121	-0.379
	Sig. (2-tailed)	.	.	0.314	0.645	0.367	.	0.892	0.000	0.529	0.155	0.000	0.041	0.791	.	0.668	0.163
ENPV	Pearson Correlation	-0.220	-0.203	0.507	0.046	0.179	0.039	1.000	0.351	-0.007	0.225	-0.230	0.216	0.466	-0.185	0.679	-0.242
	Sig. (2-tailed)	0.325	0.366	0.007	0.819	0.524	0.892	.	0.073	0.975	0.302	0.248	0.280	0.014	0.356	0.000	0.224
EIRR	Pearson Correlation	0.026	-0.023	-0.366	-0.159	0.002	0.801	0.351	1.000	-0.421	0.141	-0.343	0.174	0.366	0.061	-0.135	-0.156
	Sig. (2-tailed)	0.909	0.920	0.061	0.428	0.993	0.000	0.073	.	0.046	0.521	0.079	0.387	0.061	0.761	0.504	0.436
Land Acquisition	Pearson Correlation	-0.234	-0.185	0.782	-0.480	0.170	-0.192	-0.007	-0.421	1.000	0.385	0.321	0.110	0.045	-0.078	0.751	-0.039
	Sig. (2-tailed)	0.334	0.448	0.000	0.017	0.578	0.529	0.975	0.046	.	0.085	0.126	0.608	0.833	0.716	0.000	0.858
Resettlement People	Pearson Correlation	.	-0.112	0.214	-0.563	0.241	0.418	0.225	0.141	0.385	1.000	0.242	-0.229	0.489	0.585	0.400	0.171

Attribute		Grace Period	Year to Maturity	Government Committee	Implementation Duration	FNPV	FIRR	ENPV	EIRR	Land Acquisition	Resettlement People	Year	% Transportation	IBRD Commitment	IDA Commitment	Total Project Cost	Finance Source
	Sig. (2-tailed)	.	0.639	0.305	0.003	0.428	0.155	0.302	0.521	0.085	.	0.244	0.271	0.013	0.002	0.047	0.413
Year	Pearson Correlation	0.153	0.171	0.156	-0.381	-0.465	-0.789	-0.230	-0.343	0.321	0.242	1.000	0.059	0.064	0.144	0.190	0.237
	Sig. (2-tailed)	0.474	0.423	0.234	0.003	0.081	0.000	0.248	0.079	0.126	0.244	.	0.655	0.626	0.272	0.145	0.069
%Transportation	Pearson Correlation	-0.407	-0.402	0.259	-0.015	0.319	0.533	0.216	0.174	0.110	-0.229	0.059	1.000	0.318	-0.139	0.348	-0.465
	Sig. (2-tailed)	0.048	0.051	0.046	0.908	0.246	0.041	0.280	0.387	0.608	0.271	0.655	.	0.013	0.288	0.006	0.000
IBRD Commitment	Pearson Correlation	-0.494	-0.462	0.306	0.043	0.475	-0.075	0.466	0.366	0.045	0.489	0.064	0.318	1.000	-0.383	0.621	-0.319
	Sig. (2-tailed)	0.014	0.023	0.017	0.746	0.074	0.791	0.014	0.061	0.833	0.013	0.626	0.013	.	0.003	0.000	0.013
IDA Commitment	Pearson Correlation	0.874	0.864	-0.263	-0.131	.	.	-0.185	0.061	-0.078	0.585	0.144	-0.139	-0.383	1.000	-0.079	0.366
	Sig. (2-tailed)	0.000	0.000	0.043	0.317	.	.	0.356	0.761	0.716	0.002	0.272	0.288	0.003	.	0.548	0.004
Total Project Cost	Pearson Correlation	-0.409	-0.378	0.836	-0.050	0.217	-0.121	0.679	-0.135	0.751	0.400	0.190	0.348	0.621	-0.079	1.000	-0.469
	Sig. (2-tailed)	0.047	0.069	0.000	0.706	0.437	0.668	0.000	0.504	0.000	0.047	0.145	0.006	0.000	0.548	.	0.000
Finance Source	Pearson Correlation	0.291	0.297	-0.527	-0.267	-0.248	-0.379	-0.242	-0.156	-0.039	0.171	0.237	-0.465	-0.319	0.366	-0.469	1.000
	Sig. (2-tailed)	0.168	0.158	0.000	0.039	0.373	0.163	0.224	0.436	0.858	0.413	0.069	0.000	0.013	0.004	0.000	.

Appendix E: Adjusting Matching Weight Based on Pearson Correlation Parameters

Attributes	Match	Mismatch	Absent	Pearson	Attributes	Match	Mismatch	Absent	Match1	Mismatch1	Absent1
Grace Period	4	2	2								
	2.03	1.02	1.02	0.984	Year to Maturity	4	2	2	2.03	1.02	1.02
	2.44	1.23	1.23	-0.407	%Transportation	5	2	2	6.02	2.41	2.41
	3.04	1.53	1.53	-0.494	IBRD Commitment	3	1	0	3.74	1.25	0
	1.71	0.86	0.86	0.874	IDA Commitment	3	1	0	1.69	0.56	0
	2.06	1.04	1.04	-0.409	Total Project Cost	7	3	2	8.43	3.61	2.41
Year to Maturity	2.03	1.02	1.02								
	2.5	1.26	1.26	-0.462	IBRD Commitment	3.74	1.25	0	4.6	1.54	0
	1.42	0.72	0.72	0.864	IDA Commitment	1.69	0.56	0	0.96	0.32	0
Government Committee	4	2	1								
	2.99	1.49	0.75	0.507	ENPV	6	1	0	4.48	0.75	0
	1.82	0.91	0.46	0.782	Land Acquisition	5	2	1	3.05	1.22	0.61
	1.58	0.79	0.4	0.259	%Transportation	6.02	2.41	2.41	5.24	2.1	2.1
	1.34	0.67	0.34	0.306	IBRD Commitment	4.6	1.54	0	3.9	1.3	0
	1.52	0.76	0.38	-0.263	IDA Commitment	0.96	0.32	0	1.09	0.36	0
	0.88	0.44	0.22	0.836	Total Project Cost	8.43	3.61	2.41	4.91	2.1	1.4
	1.11	0.56	0.28	-0.527	Finance Source	6	2	2	7.58	2.53	2.53
Implementation Duration	5	2	2								
	6.2	2.48	2.48	-0.480	Land Acquisition	3.05	1.22	0.61	3.78	1.51	0.76
	7.94	3.18	3.18	-0.563	Resettlement People	4	2	1	5.13	2.56	1.28
	9.45	3.79	3.79	-0.381	Approval Year	4	1	1	4.76	1.19	1.19
	10.71	4.3	4.3	-0.267	Finance Source	7.58	2.53	2.53	8.59	2.87	2.87

Attributes	Match	Mismatch	Absent	Pearson	Attributes	Match	Mismatch	Absent	Match1	Mismatch1	Absent1
FIRR	4	2	1								
	2.4	1.2	0.6	0.801	EIRR	6	2	2	3.6	1.2	1.2
	3.35	1.67	0.84	-0.789	Approval Year	4.76	1.19	1.19	6.64	1.66	1.66
ENPV	4.48	0.75	0								
	3.44	0.58	0	0.466	IBRD Commitment	3.9	1.3	0	2.99	1	0
	2.27	0.38	0	0.679	Total Project Cost	4.91	2.1	1.4	3.24	1.39	0.92
EIRR	3.6	1.2	1.2								
	4.36	1.45	1.45	-0.421	Land Acquisition	3.78	1.51	0.76	4.58	1.83	0.92
Land Acquisition	4.58	1.83	0.92								
	2.86	1.14	0.57	0.751	Total Project Cost	3.24	1.39	0.92	2.02	0.87	0.57
Resettlement People	5.13	2.56	1.28								
	3.87	1.93	0.97	0.489	IBRD Commitment	2.99	1	0	2.26	0.76	0
	2.74	1.37	0.69	0.585	IDA Commitment	1.09	0.36	0	0.77	0.25	0
	2.19	1.1	0.55	0.400	Total Project Cost	2.02	0.87	0.57	1.62	0.7	0.46
%Transportation	5.24	2.1	2.1								
	4.41	1.77	1.77	0.318	IBRD Commitment	2.26	0.76	0	1.9	0.64	0
	3.64	1.46	1.46	0.348	Total Project Cost	1.62	0.7	0.46	1.34	0.58	0.38
	4.49	1.8	1.8	-0.465	Finance Source	7.58	2.53	2.53	9.34	3.12	3.12
IBRD Commitment	1.9	0.64	0								
	2.26	0.76	0	-0.383	IDA Commitment	0.77	0.25	0	0.92	0.3	0
	1.56	0.52	0	0.621	Total Project Cost	1.62	0.7	0.46	1.12	0.48	0.32
	1.81	0.6	0	-0.319	Finance Source	9.34	3.12	3.12	10.83	3.62	3.62
IDA Commitment	0.92	0.3	0								
	0.75	0.25	0	0.366	Finance Source	10.83	3.62	3.62	8.85	2.96	2.96
Total Project Cost	1.12	0.48	0.32								
	1.38	0.59	0.4	-0.469	Finance Source	8.85	2.96	2.96	10.92	3.65	3.65

Appendix F: A Scenario Project *for Sensitive Analysis*

Attribute	Value
Grace_Period	5
Year_to_Maturity	20
Front_End_Fee_of_Bank_Loan	0.01
Government_Committee	200.7
Implementation_Duration	60
Financial_NPV	600
Financial_FIRR	0.08
Economical_NPV	400
Economical_EIRR	0.18
Land_Acquisition	9855333
Resettlement_People	9109
Approval_Year	2001
Percentage_Transportation	0.96
IBRD_Commitment	200
IDA_Commitment	0
WB_Amount	200
Total_Project_Cost	400.57
Project_Objective_Category	
Commitment_Fee	0.0075
Effective_Date	10/20/2001
Proposed_Terms	Variable Spread & Rate Single Currency Loan
Economic_Evaluation_Method	Cost-Benefit Analysis
Country	China
Sub_Sector	Roads and highways
Lending_Instrument	Specific Investment Loan
Finance_Source	Loan
Program_of_Targeted_Intervention	No
Project_Status	Active
Region	EAP
Environment_Category	A
Borrower	P.R. CHINA
Implementing_Agency	P.R.China
Targeted_Thematic_Outcomes	Pollution management and environmental health

Appendix G: CBRisk Performance Evaluation Questionnaire

The following section provides the initial risk analysis result listed in the P51705 Project Appraisal Document Form.

Risk Item	Risk Rating
Delays in procurement	N
Delays in necessary land acquisition/site readiness	N
Inadequate project objective plan	M
Implementation delays	M
Quality of construction will not comply with required standards	M
Inadequate counterpart funding	N
Staff will not use new skills	M
Traffic grows is lower than expected rates on toll roads	N
Delays in resettlement	N
Deficient engineering design	N
Human resources not available on time	N
Training program irrelevant to needs and objectives	M
Selection of incompetent contractors	M
Policy and Institutional Reform will not receive large ownership and will be unsuccessful	M
Overall Risk Level	**M**

M-Medium; N-Small or Negotiable

The risk items in shadow are in common with CBRisk output. Besides, some other risk items that identified through CBRisk are listed below:

Risk Item	Risk Rating
Natural disasters delay or increase the cost of road works	N
Problems with contractor or consultants performance	M

Case-based Reasoning is a problem-solving paradigm that solves a new problem by remembering a previous similar situation and by reusing information and knowledge of that situation. It is a methodology for modelling human reasoning and thinking, for building intelligent computer systems, and for storing previous experience in memory.

In CBRisk, information from previous projects is processed and stored in a knowledge base. When a new project comes, it is first entered into the project description form as the cases stored in the knowledge base. It then searches the knowledge base to find similar previous projects, reuses the risks identified, and revises these risk lists to come up with suggested risks for the project in question. This preliminary suggestion is reviewed by human experts, and then a new final report of identified risk for this project is generated. This solution is associated with the project description and is retained in the knowledge base as a piece of new knowledge for future use.

Please answer following questions:

1. In what level do you think the risk identification above is accurate compared with the real project performance? Please indicate with percentage If you have further commends, please describe in the space below:					
	Strongly agree	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
2. Please indicate the degree of satisfaction of the system output.					
a. compared with original risk identification	(	(	(	(	(
b. compared with the real project performance	(	(	(	(	(
3. Do you find the system useful					
a. for construction risk management?	(	(	(	(	(
b. for risk Identification?	(	(	(	(	(
4. In what ways might CBRisk be used for?					
a. Actual decision making	(	(	(	(	(
b. Supporting experts' decision	(	(	(	(	(
c. Information and knowledge management	(	(	(	(	(
d. Others, please specify					
5. What are the potential benefits of CBRisk?					
a. More objective decisions	(	(	(	(	(
b. Higher accuracy	(	(	(	(	(
c. Increase in consistency	(	(	(	(	(
d. Greater efficiency	(	(	(	(	(
e. Improve the application of risk management	(	(	(	(	(
f. Knowledge sharing	(	(	(	(	(
g. Improving communication among project participants	(	(	(	(	(
h. Rationalise the risk management procedures	(	(	(	(	(
i. Serve as a central risk management system	(	(	(	(	(
j. Others	(	(	(	(	(
6. Do you agree with the indices used in CBRisk?	(	(	(	(	(

7. Does CBRisk have any potential for practical use?	○	○	○	○	○
8. Will you consider using CBRisk or case-based reasoning for risk management?	○	○	○	○	○
9. In what size of the firms/organizations do you think this approach is particularly useful? ☐ Small, less than 20 projects ☐ Medium, between 21 to 50 Projects ☐ Large, Between 51 to 100 projects ☐ Huge, more than 100 projects					
10. Comments					

If you have any questions, please feel free to contact me.
After you finish, please send it back to: cenyt@leeds.ac.uk

You cooperation is great appreciated! Further research result will be able to share when it is finished.

Appendix H: Publication List

Published Paper:

Tan, Y., Smith, N.J. and Bower, D.A. (2003) Improving project risk management by utilizing hybrid intelligent reasoning, Association of Research in Construction management (ARCOM) 2003, ISBN: 0953416186, Brighton, UK, pp359-368.

Tan, Y., Smith, N.J. and Bower, D.A. (2004) Managing knowledge of project risks, Proceedings of the 4th international conference of construction project management, ISBN: 980505795, 4-5 March 2004, Singapore, pp253-262.

Tan, Y., Bower, D.A. and Smith, N.J. (2004) The Role of IT in Project Risk Management: A Review, Proceedings of the 1st international conference of the world of construction project management, 27-28 May 2004, Toronto, Canada, pp323-334.

Tan, Y., Smith, N.J. and Bower, D.A. (2005) A case-based reasoning approach for risk identification in construction projects, Proceedings of the eight international conference on the Application of Artificial Intelligence to Civil, Structural and Environmental Engineering, Italy, Rome, September, 2005.

Papers under reviewing:

Tan, Y., Smith, N.J. and Bower, D.A., A case-based reasoning approach for risk identification in construction projects, Journal of Advances in Engineering Software, Special Issue: Artificial Intelligence in Civil, Structural and Environmental Engineering.

Tan, Y., Bower, D.A. and Smith, N.J., Learning from experience of previous projects: risk identification, 2006 International Conference on Computing in Civil Engineering of ASCE, Montreal, Canada, June 2006.