KNOWLEDGE SEARCH FOR NEW PRODUCT DEVELOPMENT: A MULTI-AGENT BASED METHODOLOGY

Guo Jian

A thesis submitted in partial fulfilment of the requirements of the University of Greenwich for the Degree of Doctor of Philosophy

This research programme was carried out in collaboration with the Aircraft Manufacturing Company

May 2011

DECLARATION

“I certify that this work has not been accepted in substance for any degree, and is not concurrently being submitted for any degree other than that of Doctor of Philosophy being studied at the University of Greenwich. I also declare that this work is the result of my own investigations except where otherwise identified by references and that I have not plagiarised the work of others.”

Guo Jian

Candidate

Supported by:

Professor James Gao

Supervisor

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. James Gao for his help, guidance, encouragement and support for the duration of the research project. Dr Dele Owodunni has been very supportive. Working with them has had a tremendous positive impact on my intellectual and personal development.

The Commercial Aircraft Corporation of China, Aviation Industry Corporation of China, and Vaillant Group provided assistance with data collection. I would like to thank all the assistant people for their ideas and interest, particularly Mr Zhiqiang Xie and Mr Weiming Hu. I would also like to thank the anonymous contributors to the data collection and evaluation sessions who made the research a fascinating and engaging process.

I remain indebted to my mother and father for their love and support.

RELATED PUBLICATIONS

A Jian and J Gao, a multi-agent based knowledge search methodology for product development, Proceedings of the 4th International Conference on Digital Enterprise Technology, 19-21 Sept. 2007, University of Bath, UK. Ed. P G Maropoulos and S T Newman. ISBN 978-0-86197-141-1. Pp524-530.

Jian G. and Gao J. and Yinglin Wang, a multi-agent based knowledge search framework to support the product development process, International Journal of Computer Integrated Manufacturing, Volume 23 Issue 3, pp237-247, Pub. Taylor & Francis, March 2010.

Jian G and Gao J., Knowledge search for product development: a multi-agent based methodology, in Advances in Manufacturing Technology XXIV, 8th International Conference on Manufacturing Research, 14-16 Sept 2010, University of Durham, UK. Edited by Vitanov V.I. and Harrison D., ISBN 9781905866519. Pp50-55.

ABSTRACT

Manufacturers are the leaders in developing new products to drive productivity. Higher productivity means more products based on the same materials, energy, labour, and capitals. New product development plays a critical role in the success of manufacturing firms. Activities in the product development process are dependent on the knowledge of new product development team members. Increasingly, many enterprises consider effective knowledge search to be a source of competitive advantage.

This research presents an exploratory case study conducted at an aircraft manufacturer. This investigation uncovered six, empirically derived and theoretically informed, problems to enterprise knowledge search. They have been articulated as (i) the effectual web bandwidth limits search speed; (ii) less relevant search results based on word-frequency recognition models of search engine; (iii) un-useable techniques for enterprise search; (iv) rigour security, reliability, and company policy; (v) poor search performance about unstructured enterprise knowledge; (vi) the lack of tacit knowledge sharing. Existing search methodologies have focused on the internet search, rather than providing effective search for enterprise.

This research aim is developed to assist the manufacturing enterprise in meeting the industrial requirements in the following way: a methodology and system that can improve the information and knowledge search performance in new product development process. Based on the exploratory case findings, a knowledge search methodology and system has been developed. Agent technology is used to fulfil the requirements of enterprise search. Some initial tests were conducted to better understand implementation issues and future deployment of the methodology and system in practice.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABCDE: Agent-Based Design Concurrent Design Environment
AC: Airworthiness Certificate
ACL: Agent Communication Language
AO: Assembly Order
API: Application Programming Interface
AVIC: Aviation Industry Corporation of China
CACC: Commercial Aircraft Corporation of China
CAD: Computer Aid Design
CAPP: Computer-aided Process Planning
CoCAPP: Machining Cooperative Process Planning System
CORBA: Common Object Request Broker Architecture
CPC: Collaborative Product Commerce
DSOM: Distributed System Object Model
DSS: Decision Support System
EBOM: Engineering Bill of Material
EPR: Enterprise Resource Management
FAI: The First Aircraft Institute of AVIC
FTP: File Transfer Protocol
GPL: General Public License
HTML: Hyper Text Markup Language
IT: Information Technology
JDK: Java Development Kit
JITIR: Just-In-Time Information Retrieval
KM: Knowledge Management
LGPL: Lesser General Public License
MAML: Multi-Agent Modeling Language
MAS: Multi-Agent System
MIME: Multipurpose Internet Mail Extension Protocol
MIS: Management Information System
NPD: New Product Development
OEM: Original Equipment Manufacturer

OLE: Object Linking and Embedding

P-TÂTO: Precedence Cost Tâtonement

P2P: Peer-to-peer

PC: Production Certificate

PDM: Product Data Management

PLM: Product Life Management

RA: Remembrance Agent

RAPPID: Responsible Agents for Product Process Integrated Design

RDBMS: Relational Database Management System

RE: Reconfigurable Enterprise

SAIRE: Scalable Agent-based Information Retrieval Engine

SECI: Socialization, Externalization, Combination, Internalization

SNIS: Social Network-based Item Search

TC: Type Certificate

TCP/IP: Transmission Control Protocol/Internet Protocol

URL: Uniform Resource Locator

VPM: Virtual Product Management

XML: Extensible Markup Language

CHAPTER ONE: INTRODUCTION

This chapter first provides an introduction of manufacturing and new product development process. Based on existing knowledge management and knowledge search technologies, the problems for enterprise search are figured out. Research scope is confined. Aims and objectives are outlined. After that, research methodology is clarified. At the end of this chapter, the structure of the whole thesis is given.

1.1 Importance of Manufacturing

Manufacturing, or the process of converting raw materials into products, is the backbone of any industrialised nations (Kalpakjian 1991). Manufacturing has important implications for other industries, for example, it has significant ripple effects through related service industries. Many service sector jobs depend on manufacturing. If manufacturing falls, job lost cannot always be absorbed by employment in the service industries. A long term decline in industrial employment can create a major problem of structural unemployment, and change the labour market. Consumer confidence will be affected. There are strong inter-relationships between manufacturing and the rest of the economy. A report from the National Association of Manufacturers in the USA claimed that "manufacturing pays higher wages and has a higher multiplier effect than other economic sectors" (the Manufacturing Institute 2009). The multiplier effect shows how much additional output is generated by a dollar's worth of final demand for each industry. As figure 1.1 shows, every dollar in final sales of manufactured products supports $1.40 in output from other sectors of the economy. Manufacturing therefore has a powerful positive impact on economic development. In the UK, manufacturing contributes £150 billion per annum to the economy. Half of UK exports are manufactured goods (Hutton and Denham 2008).

Manufacturers are clearly the leaders in developing new products to drive productivity. These innovations are extremely important to other industries. For example, the Internet and telecommunication equipments manufactured have revolutionized business practices across every business sector. Higher productivity means more products based on the same materials, energy, labour, and capitals, which is the basis for higher wages.

Figure 1.1: Economic Activity Generated by $1 of Sector GDP (Source: U.S. bureau of economic analysis, 2007 annual input-output tables)

Manufacturing strategies are formulated to achieve competitive advantages. The first concern is high-commitment work systems, which embraces systems such as team working and new managerial techniques such as business process re-engineering. A second issue is the international context. For issues immediately arise as to the novelty of the developments identified and what identifiable social forces are actually causing them, globalization means new international competitive pressures. In the field of work, the related forces are international competition, the role of multinational companies, and integration into global value chains (Edwards 2003). A third set of issues of concern is economic performance, which includes energy efficiency, materials consumption, renewable energy and so on.

1.2 Importance of New Product Development

Business and retail customers have an ever-increasing range of products to choose from. Therefore, manufacturers face constant competition in order to retain or increase their market share. They need to be flexible and to consider new ways of working as well as new technologies that will allow them to respond rapidly and cost-effectively to customer demands. Business introduces innovation in manufacturing to a wide range of areas, including new processes, new materials, new equipment, and new products. Innovation impacts on production, employment, profit margin and market share. The revenue associated with innovation could be derived from three sources (Kannebley et al. 2010). One is the licensing from companies that do not develop new technology. Another is the incorporation of technologic advances in their own products, which leads to the share of new markets. The third part is from the introduction of new production processes that boost the productivity, saving cost, and increasing profit consequently.

Companies must innovate in order to survive. The power of innovation is revealed in numerous studies, which show that companies leading their industries attribute about half of their revenues to products developed in the most recent years. By comparison, companies at the bottom of their industries achieve approximately one-tenth of their sales from new products. The concept of new product development (NPD) has received enormous attention in the management literature over the past 20 years. Few business activities are approached with more justified optimism than the development of new product (Trott 2008).

Alarmingly, there is evidence that the failure of an industry to produce product innovations, or new products, may contribute to its downfall. Hart (1996), citing Ughanwa and Baker(1989), provided the UK shipbuilding industry as a case in point. As the number of product innovations in the UK fell, so the industry declined. Between 1890 and 1974, the UK's market share fell from eight percent to under four percent. Meanwhile, Japan's shipbuilding industry, which had presided over a rise in product innovations in the same period, saw its market share grow to around forty percent by 1969. This was not merely symptomatic of some decline in Western industry however. Germany, which like Japan increased their output of product innovations, saw their industry command a twenty percent market share by 1970.

Research shows that the top 20 per cent of companies have a new product success rate of 80 per cent, compared to 38 per cent for the bottom 20 per cent of companies. Furthermore, the profitability of these top new product offerings are two or more times higher than their counterparts in the bottom 20 per cent. New product development is a key point of competitive advantage. Furthermore, the process of product development is considered to be a critical factor for the manufacturing businesses that aspire to prosper in competitive markets.

The new product development process is most usually defined as idea generation, idea screening, concept development and testing, business analysis, product and marketing mix development, technical implementation, and commercialization. These steps may be iterated or eliminated as needed. Cooper (1994) defined NPD process as ‘a formal blueprint, roadmap template or thought process for driving a new product from the idea stage through to market launch and beyond’.

It is possible to classify the proposed models of NPD process into seven distinct categories (Trott 2008):

- departmental-stage models,
- activity-stage models and concurrent engineering,
- cross-functional models (teams),
- decision-stage models,
- conversion-process models,
- response models, and
- network models.

A commonly used model today is the cross-functional stage-gate model, which Griffin (1997) indicated that it is employed by almost sixty percent of firms in the United States. This model divides the NPD process into discrete stages, each of which is followed by a review gate. Each stage can be broken down into a collection of predefined, cross-functional and concurrent tasks, which are executed by cross-functional teams. The importance of such a formal process model and its connection to best practice is well established in the literature (Cooper 1994; Griffin 1997). Furthermore, Fredericks (2005) showed that cross-functional involvement in product development is dependent on a collective understanding of the tasks required at different phases of the NPD process.

1.3 Information/Knowledge in New Product Development

The NPD teams are cross-functional teams that are engaged in information/knowledge-producing and application activities. Team members bring to the situation their repertoire of skills, knowledge, and strategies, which affect and be affected by the situation. The interaction of individual and context is dynamic. From the idea generation step to the launch step, the creation of new knowledge could be viewed as the central theme of the NPD process. Knowledge is information possessed in the mind of individuals. It is personalized information, which may or may not be new, unique, useful, or accurate, related to facts, procedures, concepts, interpretations, ideas, observations, and judgments. Some of the knowledge is tacit knowledge, which cannot be explicated fully even by experts and can be transferred from one person to another only through a long process of apprenticeship (Polanyi 1967). In contrast, explicit knowledge is able to be articulated and communicated. It could be transferred between individuals and organisations.

Nelson and Winter (1982) argued that much organisational knowledge remains tacit because it is impossible to describe all the aspects needed for successful performance. Kogut and Zander (1992) distinguished information (e.g., facts) from know-how (e.g., how to organize factories). For example, the list of ingredients in a recipe consists of information. But the description of action steps is, at best, an imperfect representation of the know-how required. Badaracco (1991) pointed out that tacit knowledge exists in individuals or groups of individuals. He referred to such knowledge in individuals and social groups as embedded knowledge. In the NPD process, a cross-functional team is set up because its members have collective knowledge that cannot be held efficiently by any of its individual members. Once the team members get together, there is potential for the team to use and create new knowledge. This new knowledge is a result of a combination of tacit and explicit knowledge. Combining explicit knowledge is easy. Salomon (1993) conceived that individuals' inputs through their collaborative activities, affect the nature of the joint, distributed system, which in turn affects their cognitions such that their subsequent participation is altered, resulting in subsequent altered joint performances and products. Therefore, the potential for new knowledge is proposed to be embedded in the team and its interactions. The NPD team possesses embedded knowledge. The new product is embodied knowledge. The NPD process is to manage the transition from embedded to embodied knowledge.

Manufacturers must produce the customized, high-quality, and yet inexpensive products that are being demanded by their customers. These demands stress the product development process.

1.4 Overview of Knowledge Management Technologies

Knowledge management (KM) refers to identifying and leveraging the collective knowledge in an organisation to help the organisation compete (Vonkrogh 1998). It is work-in-progress, not a one-time search for an idealised state. Computer-enabled information systems are necessary but not sufficient elements of a comprehensive approach to KM. Holistic KM should be integral to the organization, working with not against the grain of its technical, social and cultural processes. Senior managers with titles such as “chief knowledge officer” may be crucial in establishing strategic priorities and change programmes, but all NPD personnel bear responsibility for effective KM.

Effective KM needs to: acknowledge the multiple organizational levels at which knowledge is deployed; support the production, elicitation and exchange of tacit knowledge as well as explicit, codified information; hence accommodate and enable both informal and formal, typically information technology (IT) enabled knowledge processes. KM technologies are necessary to be selected depending on the purpose for which knowledge is being managed.

IT does not apply to all the issues of KM. However, it can support KM in many ways. As Alavi and Leidner (2001) summarized, applications of IT to organisational knowledge management initiatives reveals three common applications: the coding and sharing of best practices; the creation of corporate knowledge directories (or referred to as the mapping of internal expertise); and the creation of knowledge networks. Figure 1.2 shows a variety of IT technologies in facilitating different KM processes.

Knowledge transformation within and between tacit and explicit forms can be supported by the technologies. The strongest contribution to current solutions is made by technologies that deal largely with explicit knowledge, such as search and classification. Contributions to the formation and communication of tacit knowledge, and support for making it explicit, are currently weaker, although some encouraging developments are highlighted, such as the use of text-based chat, expertise location, and unrestricted bulletin boards.

Knowledge Management Process	**Knowledge Creation**	**Knowledge Storage/Retrieval**	**Knowledge Transfer**	**Knowledge Application**
Supporting Information Technologies	Data mining; Learning tools	Electronic bulletin boards; Knowledge repositories; Databases	Electronic bulletin boards; Knowledge directories	Expert systems; Workflow systems
IT Enables	Combining new sources of knowledge; Just in time learning	Support of individual and organisational memory; Inter-group knowledge access	More extensive internal network; More communication channels available; Faster access to knowledge sources	Knowledge can be applied in many locations; More rapid application of new knowledge through workflow automation
Platform	Groupware and communication technologies			
Technologies	INTRANETS			

Figure 1.2: Knowledge Management Processes and the Potential Role of IT (Alavi and Leidner 2001)

1.5 Overview of Knowledge Search Technologies

Search engines mainly have two types of searching methods. One is robot crawler based indexes; the other is human powered directories (Sullivan 2002). When a requirement is received, the robot crawler based engine sends a crawler (or spider) to read a web page, make a copy, return the copy to the index of the search engine, and also crawl to other linked web pages in the site. The search engine software matches the web pages in the index with the search requirement and then lists the results. Human powered search engines, which are better known as web directories, need human assistance. For example, Website editors or owners produce a site description, on which the index of search engine is based. The search engine looks for matches with the descriptions submitted by the web site editors, not the actual content of the webpage.

Some search engines are purely crawler based, like Google. Such systems are powerful and efficient at locating matching terms and phrases. However, they are dumb and passive systems that require resourceful, active, intelligent human users to help produce acceptable results. Some search engines are human powered, like Open Directory. They are popular because the sites are hand-picked by the users themselves to be included in the index to provide the higher quality of results. However, such systems do not re-read a web page frequently because it is a manual process and time consuming. And many other search engines today present both crawler based results and human powered listings, like MSN Search. Some researchers have suggested that further improvement of information search and retrieval is to better index the Web documents and database records with items such as more key terms and conceptual indexing (Theobald and Weikum 2002). However, enhancing millions of web pages, documents and records would be extremely costly. Therefore, creating better search/retrieval engines provides a more realistic opportunity for further improving information search.

Current search engines also have problems with online databases. An online database is a database accessible from a network. Jansen (2005) summarized the obstacles facing the users searching for information in on-line databases as including: (i) a large number of hits along with failure to reduce the unnecessary retrieved information sets, (ii) zero hits and failure to increase the successful retrieval sets, and (iii) poor understanding of the cataloguing rules. In addition, there is a lack of understanding of the indexes, types of files, stop words, and hyphenation problems. It should also be noted that there is a rich source of knowledge which is not on the Web pages or on-line databases. They may be in individual's local database or file cabinets or in companies' secured information warehouses. These kinds of knowledge are not accessible through current Internet search engines. Some vendors provide information retrieval function in their knowledge management solution, like Autonomy (Autonomy 2009) which employs a search engine.

In this research, Agent technology is used for the proposed enterprise knowledge search engine. Agent is a computer system situated in some environments, and capable of autonomous action in this environment in order to meet its design objectives. These systems can act without intervention from humans and take control over their actions and internal state. Now, Agent technology is applied in the manufacturing domains like engineering design, process planning, production planning and resource allocation, and so on. Agent technology has been adopted in information search area as well as personal assistance, network management, software distribution, and so on. Some Agent based approaches for knowledge search have been developed or are currently under development in academic and commercial research labs. But they still have to wait to make it out to the real world of enterprise users broadly.

The applications of search today and tomorrow require more than exhaustive search, because these application domains become increasingly complex. Traditional methods fail to break the complexity barrier caused by the combinatorial explosion that characterizes these large, real-world domains. This research focuses on the application domain of enterprise search. The knowledge search in enterprise application currently has some problems which are summarized as follows:

i. Conventional information search based on client/server structure relies on effectual web bandwidth and limits search speed.
ii. The word-frequency recognition models used for search engine may result in less relevant search results.
iii. Some techniques broadly used for Internet search cannot be used for enterprise search, because the link structure of enterprise repositories is quite different from the one on the Internet.
iv. Security, reliability, and company policy are rigour for enterprise search.

v. Unstructured enterprise knowledge should also be searchable as well as structured knowledge.
vi. Enterprise search aims for better sharing of tacit knowledge as well as explicit knowledge. Traditional information retrieval only for relevant documents does not fulfil this requirement.

In manufacturing enterprise knowledge search domain, a Multi-Agent system can be applied and implemented in different ways:

The Agents can be used to encapsulate search activities or wrap legacy software systems in an open, distributed intelligent environment using a functional decomposition approach. Therefore, instead of moving large amounts of data to the search engine to create indexes, the Agents could utilise indexes built by other search engines to obtain sorted information. The refined search indexes are generated based on the sorted information. This addresses the first problem. Some special types of search are common in enterprise search, like term definition, homepages of groups or topics, employees' personal information and expert on topics. An additional search for these types of information is going to be considered. All the types of searches will be combined together finally. The Agent-based encapsulation and integration approach can facilitate this plug-and-play kind of integration of heterogeneous software and hardware systems. This addresses the fifth and sixth problems.

The Agents are inherent autonomous. It keeps track of user interests. The Agents extract, combine, and refine relevant keywords. The active knowledge search engine suggests to the designer items it estimates to be close to the target of the search. Moreover, it infers its suggestions from the designer's normal search actions, without any special input. As a result, the system can adapt to designer's individual interests, reduce the amount of information that are not required, and help designer to access right knowledge with minimal effort. This addresses the second problem.

The Multi-Agent based search engine is proposed to improve the enterprise search activity. The enterprise information is accessed by the enterprise users based on their access authorizations. This addresses the forth problem. The system is developed on the enterprise information structure. Techniques used adapt to enterprise. This addresses the third problem.

1.6 Research Scope

The focus of the knowledge management methodology to be developed is the knowledge search within manufacturing enterprises. The scope of its application in investigation is confined to aviation products in a complex product development environment, using a formally defined NPD process. It will be limited to new product development projects that are derivations or incremental improvements to existing products and will focus on tasks performed by the design functions in a single stage or phase of the NPD process. However, in principle, the proposed methodology and system is considered to be applicable to all the phases of the generic NPD process in manufacturing context.

1.7 Aims and Objectives

The aim of this research is to provide a methodology and system for improving the information, particularly knowledge, search performance in new product development process in manufacturing enterprises.

The research objectives are to:

- Explore the nature of knowledge and available approaches of information and knowledge search in new product development process;
- Identify key barriers to enterprise knowledge search, compared with general information search engines;
- Develop a Multi-Agent based knowledge search methodology which overcomes shortcomings of available search engines in enterprise applications such as limited search speed, unstructured knowledge, and various types of knowledge.
- Implement the enterprise search engine to fulfil the methodology, and conduct an initial test of the methodology with industrial data.

1.8 The Collaborating Company

In the early stage of this research project, an industrial survey was carried out with Vaillant Group as part of a larger scale project involving a number of researchers. The main purpose of the investigation is to identify requirements and problems in knowledge management (mainly knowledge sharing) in new product development. The questionnaires are included in Appendix A. And the answers taken during the interviews are included in Appendix B. Another company (Edwards Vacuum) was also visited by the author during the project to collect data and opinions of senior engineers involved in product development and manufacturing, and the questionnaires and responses of interviewees are included in Appendix D.

The collaborating companies of this research are Commercial Aircraft Corporation of China (CACC) and Aviation Industry Corporation of China (AVIC). The main business of CACC is the development of large civil passenger aircrafts. AVIC is an important cooperation partner of CACC and one of its shareholders. AVIC divides its production business into ten units, known as defence, transport aircraft, engine, helicopter, airborne equipment and system, general aircraft, aeronautical research, flight experiment, trade logistics, and asset management. These two companies provide raw data of industrial case study, example products, business process information and engineering knowledge. The overall domain, aim, and objectives of this project were devised and refined in cooperation with the collaborating companies. The company regards information especially knowledge management as one of its strategic priorities, and recognises that efficient information/knowledge search technologies would significantly improve their new product development process. Therefore it devoted valuable manpower to this project. A more detailed introduction to the company will be provided in a separate chapter later in this thesis.

1.9 Research Methodology

A robust research methodology is widely acknowledged as a critical part of a research project. Irani et al (1999) emphasised the importance of research strategy in developing a research methodology, since the research strategy is "the underlying construct upon which any robust methodology is built". A research strategy is considered to be "a way of going about one's research, embodying a particular style and employing different research methods" (Irani et al. 1999).

The likely data sources for this investigation would be qualitative in nature. It is because that knowledge to be searched in this project is available in a manufacturing company environment, and is mainly in records written in natural language and descriptive observations such as company reports or process documentation. Numerical or quantitative data is less of problematic in this research topic.

Case study is selected as the research strategy for this research investigation. Case study is defined as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between object of study and context are not clearly evident" (Yin 2003). Case study has been a common research strategy in social science. It is based on an in-depth investigation of a single individual, group, or event. It is widely employed in information systems research. Darke et al (1998) argued that case study "is well suited to understanding the interactions between information technology - related innovations and organisational contexts".

Case study is chosen in this research investigation because it provides a rich and deep understanding of the domain of interest, and suitable for application in a business environment. The sufficient academic credibility demonstrates its use in a doctoral research investigation. It has been used in similar research projects published in international, peer-reviewed literature. Moreover, it allows the use of multiple sources of data and data collection methods. Therefore more than one source is likely to be available in a product development environment.

The aircraft manufacturer would act as the setting for the case studies. The relationship that the researcher was able to cultivate with the company in the course of the investigation made it possible to gain access to confidential documentation and senior personnel that would otherwise be difficult to obtain.

As reported in section 1.7, the research aims to provide a prototype method and system for facilitating knowledge search in new product development process. Scientific or analytical approaches usually consist of five steps: observations or preliminary studies, hypothesis formulation, hypothesis test, hypothesis evaluation and hypothesis acceptance or rejection (Arbnor and Bjerke 2009). To fulfil the research aim, a five-stage research process is employed. At each stage of the process, detailed methods are described as follows.

Stage 1 is to set down the scope of the research. A review of the literature is conducted to establish the scope of the research and the research objectives.

Stage 2 is to investigate the real world. An industrial investigation is conducted to gain understanding from the real world. The researcher visited two large aircraft manufacturing companies, interviewed several design and development experts and IT people. Semi-structured interviews have been conducted with relevant personnel. The industrial investigation identifies knowledge search barriers and requirements. And it provides data and information need for the case study approach.

Stage 3 is to propose a methodology for knowledge search. The reviewed literature about methodologies and technologies has been used as the basis of the proposed methodology. Additionally, a methodology is proposed which is deemed suitable for better search activity of knowledge used and generated in the NPD process.

Stage 4 is to develop the knowledge search system. Using the methodology developed in stage 3, a knowledge search system is developed to support the search activity in a manufacturing NPD environment.

Stage 5 is to test the proposed methodology and system. Factual data is collected to demonstrate the functionality of the search system by showing how it could be used to retrieval knowledge in the NPD process of the aircraft company.

In summary, for this research, a knowledge search methodology is proposed based on existing theory and industrial investigation. Then an enterprise search system is developed and tested. Based on the results, the further work of the research is discussed.

1.10 Thesis Structure

Chapter 1 – Introduction: This chapter gives an introduction to the domain of research and states the aims and objectives and scope of the research project. Then it describes a research methodology to fulfil the research aim and objects.

Chapter 2 – Literature survey: This chapter reviews the published literature relevant to knowledge search, generic knowledge search barriers in new product development, and methodologies and technologies to facilitate knowledge sharing. Areas for further research are also identified.

Chapter 3 – Industrial investigation: This chapter describes the findings of investigation of knowledge search barriers in the new product development process in a large aircraft manufacturing company.

Chapter 4 – Proposed methodology: This chapter outlines the work carried out to design a method that tackles the knowledge search barriers identified in the industrial investigation in chapter three.

Chapter 5 – Development of Multi-Agent knowledge search system: This chapter described the development of a knowledge search system to facilitate knowledge search in the new product development process.

Chapter 6 – Test of the Multi-Agent knowledge search system: This chapter describes an initial test of the knowledge search system with search actions taken by the new product development practitioners.

Chapter 7 – Discussions, conclusions and further research: This chapter states the conclusions of the research project and explores areas for further research.

CHAPTER TWO: LITERATURE SURVEY

Last chapter introduced the trends and challenges in manufacturing industry, and emphasised the importance of knowledge search for new product development. Then aims and objectives were defined. Based on the research aim, a review of the literature was carried out examining issues related to information and knowledge management. This chapter presents a comprehensive review of the work carried out and reported by previous researchers and practitioners in information and knowledge management, search techniques, and applications in manufacturing especially in the product development process. Key concepts, issues, frameworks, systems and technologies will be introduced, in particular, the application of Agent technology in manufacturing and information/knowledge search will be emphasised. Main challenges, problems and barriers will also be examined to confirm the research gaps, and help define the aims and objectives of this project. The main research ideas and methodology are based on the result of the literature survey.

2.1 The Concept of Knowledge

As Stata (1989) observed more than twenty years ago, knowledge is the only sustainable source of competitive advantage. Peter Drucker (1993) has called knowledge, rather than capital or labour, the only meaningful economic resource of the post-capitalist or knowledge society. Knowledge management emerged as a scientific discipline in the earlier 1990s. In recent years, there has been a general recognition that knowledge is a fundamental enterprise resource.

In order to provide focus, consideration of the term here will mainly concentrate on how it is perceived in the knowledge management literature. It is very common to see a blurring of the meaning of data, information, and knowledge. Fahey (1998) argued that if knowledge is not something that is different from data or information, there is nothing new or interesting in knowledge management. It has often been pointed out that data, information, and knowledge are not the same.

As defined in the dictionary, data is “known facts or things used as a basis of inference or reckoning” (Jashapara 2004). Data is symbols which is raw (Bellinger et al. 1997). Spek and Spijkervet (1999) defined data as not yet interpreted symbols.

Ackoff (1989) defined information as “data that are processed to be useful”. Quigley and Debons (1999) took information as text providing answers to the questions “who”, “what”, “where”, and “when”. Information is data that has been given useful meaning by way of relational connection.

From a knowledge management perspective, Nonaka et al (2000) defined knowledge as commitments and beliefs created from a flow of meaningful messages, which is interpreted information by individuals. Quigley and Debons (1999) considered knowledge as text providing answers to the questions “why” and “how”.

There is implicitly a tight coupling between the concepts of information and knowledge. But there is no effectively distinguishing key. In view of the tight coupling, some authors feel that it is not necessary to distinguish clearly between the two concepts. The terms knowledge and information are often used interchangeably. Kogut and Zander (1992) defined information as “knowledge which can be transmitted without loss of integrity”, which implies that information is a form of knowledge. Keane and Mason (2006) argued that the distinction between information and knowledge should not be of concern to researchers involved with knowledge management issues. That is, that one should not be concentrated on at the expense of the other, since this increases the difficulty of managing knowledge. Nonaka (1994) also claimed such carelessness.

Davenport and Prusak (1998) offered a working definition of knowledge in organisations as follows:

> Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organisations, it often becomes embedded not only in documents or repositories but also in organisational routines, processes, practices, and norms.

This review will be confined to a discussion of the key ideas concerning the nature of knowledge and the identification of emerging trends pertinent to this research.

Approaches of categorising knowledge are varying. Different frameworks have been proposed. The most common notion of knowledge in the current knowledge management literature is based on the writings of Gilbert Ryle (1949) and Michael Polanyi (1967). From this perspective, knowledge exists along a continuum between tacit knowledge which is “know how” and explicit knowledge which is “know what”. Nonaka (1994) proposed a framework for categorizing knowledge in organisations to tacit knowledge and explicit. Tacit knowledge is deeply rooted in action, commitment, and involvement in a specific context, which makes it hard to formalize and communicate. Individual may not be consciously aware of it. Explicit knowledge could be codified and transmitted in formal, systematic language (Alavi and Leidner 2001).

Another framework proposed by Collins (1993) suggested that types of knowledge are embrained, embodied, encultured, embedded and encoded. Embrained knowledge is dependent on conceptual skills and cognitive abilities. Embodied knowledge is action oriented and is likely to be only partly explicit. Such knowledge depends on people’s physical presence, on sentient and sensory information, physical cues and face-to-face discussions, is acquired by doing, and is rooted in specific contexts. Encultured knowledge refers to the process of achieving shared understandings. It explores the significance of language, and hence to be socially constructed and open to negotiation. Embedded knowledge resides in systemic routines. Such knowledge is likely to depend heavily on relationships and material resources. It is analysable in systems terms, in the relationships between technologies, roles, formal procedures, and emergent routines. Encoded knowledge is information conveyed by signs and symbols. Such knowledge includes books, manuals, codes of practice, and information encoded and transmitted electronically (Blackler 1995).

Lyotard (1984) viewed knowledge as “a set of cultural practices situated in and inextricably linked to the material and social circumstances in which it is produced and consumed”. Based on this perspective, Hassard and Kelemen (2002) proposed a framework that production of knowledge relies on resources cut away from its original creation, and consumption of knowledge occurs through "being in the world" and social participation in a community of practice.

2.2 The Concept of Knowledge Management

One of the major challenges in knowledge management is converting internalized tacit knowledge into explicit knowledge through various ways in order to easily communicate and share it. However, the same effort is on the internalization of explicit knowledge for individual use. Nonaka and Takeuchi (1995) proposed a 'socialization, externalization, combination, internalization (SECI)' model about knowledge creation through the conversion process of tacit and explicit knowledge. Figure 2.1 displays the spiral knowledge conversion process of the SECI model (Nonaka and Takeuchi 1995; Nonaka and Toyama 2003). Tacit knowledge is shared in social interaction, i.e., 'Socialization' in the figure. Individuals accumulate and share tacit knowledge about the world that surrounds them through practical consciousness. Tacit knowledge is extracted to become explicit knowledge through the process of 'Externalization', to become concepts, images, and documents to be shared with others. Explicit knowledge is collected, combined, edited, or processed to form more complex and systematic explicit knowledge through the 'Combination' process. Explicit knowledge is applied and used in practical situations, converting into tacit knowledge by individuals, through the 'Internalization' process.

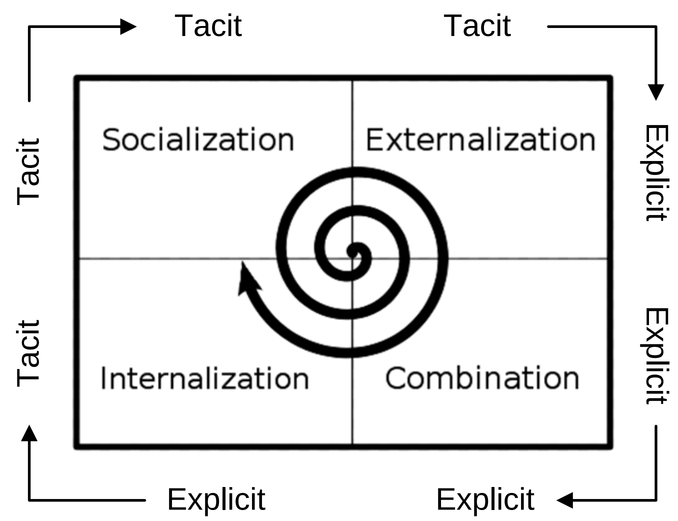


Figure 2.1: the Spiral Knowledge Conversion Process Between Tacit Knowledge and Explicit Knowledge (Nonaka and Toyama (2003)

In the context of an organisation, different interests and perspectives of knowledge lead to different perceptions of knowledge management. In 1993, Blackler et al (1993) investigated the issue of management of expertise, management of knowledge-based organisations, and management of knowledge workers. It was recognised as fertile subsoil for later growth. The explosive expansion of knowledge management literature came in just a year or two (Scarbrough and Swan 2001)., and diverse definitions of knowledge management have arisen.

Some came from a perspective of information technology and emphasised on knowledge capture and storage. Mertins et al (2001) took knowledge management as "all methods, instruments and tools that in a holistic approach contribute to the promotion of core knowledge processes". Some were from the perspective of human resource process and emphasised on knowledge creation and sharing. Swan et al (1999) scoped out knowledge management broadly as "any process or practice of creating, acquiring, capturing, sharing and using knowledge, where it resides, to enhance learning and performance in organisations". Some were from the perspective of the integration of information technology and human resource. Davenport and Prusak (1998) suggested that "knowledge management draws from existing resources that your organisation may already have in place – good information systems management, organisational change management, and human resources management practices". Later on, a more strategic management perspective was adopt by Newell et al (2002). They suggested knowledge management "improving the ways in which firms facing highly turbulent environments can mobilise their knowledge base (or leverage their knowledge 'assets' in order to ensure continuous innovation)". Therefore, knowledge management is best understood as complex, multilayered, and multifaceted concepts.

Most knowledge management work was concerned with groups, communities, and networks. In recent research, the term personal knowledge management was introduced and refers to the management of knowledge at the individual level (Wright 2005). Ideas about how to manage personal knowledge to increase innovation and productivity are investigated in firms. Grundspenkis (2007) argued that personal knowledge management is a collection of processes that an individual carries out to gather, classify, store, search, retrieve, and share knowledge in the daily activities. It is a bottom-up approach to knowledge management, as opposed to more traditional, top-down knowledge management (Pollard 2008).

For this research project, a working definition of knowledge management is most appropriate, which is proposed by Malhotra (2001):

> Knowledge management caters to the critical issues of organisational adaptation, survival and competence in the face of increasingly discontinuous environmental change. Essentially, it embodies organisational processes that seek synergistic combination of data and information-processing capacity of information technologies, and the creative and innovative capacity of human beings.

2.3 Knowledge Management in Manufacturing and Product Development

Knowledge management should link to the knowledge work process it is designed to support. The linkages must specify how knowledge should be imported to and exported from the process, when and how in the process this knowledge should be used, and what difference it should make in the outcome (Davenport and Prusak 1998). The strongest contributions to knowledge management by technologies are largely dealing with explicit knowledge, such as search and classification. Contributions to the formation and communication of tacit knowledge, and support for making it explicit, are weaker (Marwick 2001). Recent discussions of what techniques and technologies are appropriate for sharing different types of knowledge and in not only knowledge availability but also knowledge use (Prusak 2001).

Pricing pressures leave no room for inefficient production. The cycle time for developing new products and getting them on the market is becoming more and more compressed. New products have become a focus of competition for many manufacturers. The product development process consists of a sequence of steps of activities which an enterprise employs to conceive, design, and commercialise a product (Ulrich and Eppinger 1995). The knowledge-based activities of developing products and processes are becoming the primary internal functions of firms and the ones with the greatest potential for providing competitive advantage (Davenport and Prusak 1998).

New product design and development process is tightly linked to knowledge management. People of NPD transfer knowledge to or from work processes at the core of the organisation. Project management of NPD reviews what has been learned and what needs to be learned next at each phase of process. Knowledge creation and use is maximized in prototypes of NPD process. Process design of NPD specifies how knowledge is to be used at different points in the flow of work. And IT technologies in NPD supply the right form of knowledge.

Great attention has been focused in recent years on the application of knowledge management to new product development, a point emphasised by Zahay et al (2004). In order to achieve higher levels of performance in new product development, knowledge or information from past product development projects is needed to integrate as well as current project information between research and development and marketing (Adams et al. 1998). The diversity and complexity of NPD knowledge has inspired the contribution to academic research attempting to describe its nature, usage, management, and sources. Broens and Vries (2003) claimed that classifications of knowledge help NPD practitioners such as engineers find knowledge more easily.

Previous research has shown that knowledge from past product development projects affects performance in new product development (Lynn et al. 2000). Adams et al (1998) claimed that the processes of recording or acquiring information, retrieving or disseminating information, and reviewing or utilising information influence the effect of improvisation on NPD process outcomes.

2.4 Information and Knowledge Search Technologies

Advanced information technologies developed in the recent years help capture and share the valuable explicit and tacit knowledge or "knowledge how" in enterprise. The most common technologies employed by organisations were: creating an Intranet, creating data warehouses, implementing decision-support tools, and implementing groupware to support collaboration. Some authors believe that tacit knowledge leads to greater effectiveness in enterprises, and try to develop tools to enable tacit knowledge to be made explicit in a manageable manner (Jashapara 2004).

The management of organisational knowledge today certainly implies the sophisticated development of electronic corporate information systems, which enable a firm to abstract its activities and codify them in the form of generic rules (Gates and Hemingway 1999). Enterprises provide their members with the requisite propositional statements for acting efficiently and consistently. The member should have all needed information instantly, ideally (Tsoukas and Vladimirou 2001).

Search engines are the most common form of retrieving information on the web. The use of information systems in libraries has formed an important basis for how search engines now work. There is a long history of ideas about how to organise knowledge in the library. But the rise of computing in a library setting brought mathematics and linguistics to bear in new ways. Many of the techniques now used by search engines were first used by library indexes. Early computer systems provide approaches of encoding data digitally, and structures for holding that data. Therefore the collections could be rearranged and cross-indexed much more quickly than the physical counterparts. Over time, this evolved into its own art. Database design and database retrieval systems continued to be a rapidly advancing subfield of computer science (Halavais 2009). A traditional database was later enhanced by combining it with a knowledge-based subsystem to allow users to formulate database queries with natural language, and generate cooperative results for ill-formed queries.

As the Internet began its exponential increase in size during the 1990s, it became apparent that there was more information than ever before that could be browsed. Early technologies used for finding files or users were often built into the operating system. Once computers were networked via the Internet, it was possible to use the same functions from a distance. The first indexes on the Internet were created manually, often by the users of the systems as a guide to others. File Transfer Protocol (FTP) is one of the first ways of moving files between computers, which continues to be used today. A text document is created to briefly summarize the content of each file on a given server. Internet user could choose an FTP server from a list of public servers, and request that list to choose file to be downloaded.

The first search engine Archie periodically visited the existing FTP sites and indexed their directories in 1990 (Halavais 2009). It searches the titles of the files without using links to the information or content of each page. It represented a first effort to indexing a chaotic information resource, not by imposing order on it from above, but by mapping and indexing the disorder to make it more usable.

The World Wide Web emerged from 1991 (Ward 2006). Users began publishing their bookmark files, which is a collection of useful URLs (Uniform Resource Locator), to the web as pages. The updating and annotating of links to interesting new websites gave rise to the first collaborative directories and search engines. Modern web search engines are complex software systems using the technology that has evolved over the years. There are several categories of search engine software, including web search engines, database or structured data search engines, and mixed search engines. Web search engines are designed to allow searching through these largely unstructured units of content of web pages and documents (Teevan et al. 2004). They are built to follow a multi-stage process: crawling the pages or documents to discover their contents, indexing their content in a structured form (e.g., database), and finally resolving user queries to return results and links to the documents or pages from the index.

In the case of full-text search, the first step is to find and index the web pages. In the past, search engines started with a small list of URLs as seed list, fetched the content, parsed for the links on those pages, and fetched the web pages pointed to by those links which provided new links and the cycle continued until enough pages were found (Khare et al. 2004). Most modern search engines now utilize a continuous crawling method rather than discovery based on a seed list. The continuous crawling method is just an extension of discovery method without using a seed list, because the crawling never stops. The current list of pages is visited on regular intervals and new pages are found when links are added or deleted from those pages. Many search engines use sophisticated scheduling algorithms to decide when to revisit a particular page. These algorithms range from constant visit-interval with higher priority for more frequently changing pages to adaptive visit-interval based on several criteria such as frequency of change, popularity and overall quality of site, speed of web server serving the page and resource constraints like hardware and bandwidth of Internet connection. Search engines crawl many more pages than they make available for searching because crawlers find a lot of duplicate content pages on the web, and many pages don't have useful contents. Duplicated and useless contents often represent more than half the pages available for indexing.

Pages discovered by crawlers are fed into a server (often distributed) that creates a link map of the pages. Link map is a graph structure in which pages are represented as nodes connected by the links among those pages. This data is stored in data structures that allow fast access to the data by certain algorithms which compute the popularity score of pages on the web, essentially based on how many links point to a web page and the quality of those links. One such algorithm, PageRank, proposed by Google founder Larry Page (Vise and Malseed 2005), is well known and has attracted a lot of attention. The idea of doing link analysis to compute a popularity rank is earlier than PageRank, and many variants of the same idea are currently in use. These ideas can be categorized in three main categories: rank of individual pages, rank of web sites, and nature of web site content (Kleinberg 1999). Search engines often differentiate between internal links and external links, with the assumption that links on a page pointing other pages on the same site are less valuable because they are often created by web site owners to artificially increase the rank of their web sites and pages. Link map data structures typically store the anchor text embedded in the links because anchor text often provides a good quality short-summary of a web page's contents.

Indexing is the process of extracting text from web pages, tokenizing it and then creating an index structure that can be used to quickly find which pages contain a particular word (Grefenstette 1997). Search engines differ quite a lot in tokenization process. The issues involved in tokenization are: detecting the encoding used for the page, determining the language of the content, finding word, sentence and paragraph boundaries, combining multiple adjacent-words into one phrase and changing the case of text and stemming the words into their roots. This phase also decides which sections of page to index and how much text from very large pages (such as technical manuals) to index. Search engines also differ in the document formats they interpret and extract the text from.

Some search engines go through the indexing process every few weeks and refresh the complete index used for web search requests while others keep updating small fragments of the index continuously (Introna and Nissenbaum 2000). Before web pages can be indexed, an algorithm decides which node (a server in a distributed service) will index any given page and makes the information available as metadata for other components in the search engine. The index structure is complex and typically employs some compression algorithm. The selecting of compression algorithm involves a trade-off between on-disk storage space and speed of decompression when needed to satisfy search requests. The largest search engines use thousands of computers to index pages in parallel.

As the largest web search engines Google and Yahoo! process billions of web pages and return results for thousands of searches per second. High volume of queries and text processing requires the software to run in highly distributed environment with high degree of redundancy. A listing of the most popular search engines globally appears in table 2.1.

Search Provider	Searches (000)	Share of Searches (%)
Total	9,716,488	≈100
Google	6,387,932	65.7
Yahoo	1,304,427	13.4
MSN/WindowsLive/Bing	1,183,268	12.2
AOL	245,810	2.5
Ask.com	183,975	1.9
My Web	120,713	1.2
Comcast	52,011	0.5
WhitePages.com Network	29,868	0.3
NexTag	28,633	0.3
Yellow Pages	23,857	0.2
others	155,994	1.6

Table 2.1: Top 10 Search Engines Use As of March 2010 (source: Nielsen MegaView Search)

Searching for text-based content in databases presents some special challenges and opportunities to a number of specialized search engines (Silberschatz et al. 1996). Databases are slow when solving complex queries (with multiple logical or string matching arguments). Databases allow logical queries (e.g., multi-field Boolean logic) which are not allowed by full-text search. There is no crawling necessary for a database search, since the data is already structured, but it is often necessary to index the data in a more compact form designed to allow for faster search.

In database search systems, relational databases are indexed by combining multiple tables into a single table containing only the fields that need to be queried (or displayed in search results). The actual data matching engines can include any functions from basic string matching, normalization, transformation. Database search technology is heavily used by government database services, e-commerce companies, web advertising platforms, and telecommunications service providers.

In cases where the data searched contains both database content and web pages or documents, search engine technology has been developed to respond to both sets of requirements. Most mixed search engines are large Web search engines (e.g. Google) or enterprise search software products (e.g., Autonomy). They search both through structured and unstructured data sources. Pages and documents are crawled and indexed in separate indices. Databases are indexed also from various sources. Search results are then generated for users by querying these multiple indices in parallel and combining the results according to the research rules.

Much of the incremental value of these search systems comes from their ability to connect to multiple sources of content and data and their ability to interpret their multiple formats. Information and knowledge can come in a variety of forms, as structured, unstructured, and semi-structured. Financial data, sales data, customer data, and demographic data are usually structured. Documents, e-mails, presentations, videos are normally unstructured. Cases, policies, procedures, and action plans are normally semi-structured. In order to organise this various information and knowledge, efforts are made on a common schema conceptualising a vocabulary of terms and relationships to represent the information and knowledge. This is described as knowledge map or ontology. Ontologies are metadata schemas, providing a controlled vocabulary of concepts, each with an explicitly defined and machine processable semantics. By defining shared and common domain theories, ontologies help both people and machines to communicate concisely, supporting the exchange of semantics and not only syntax (Maedche and Staab, (2001). This implies that domain ontology provides users with a formalised vocabulary for describing a given domain.

2.5 Advantages and Shortcomings of Existing Search Technologies

Information and Knowledge search engines provide intelligent multi-dimensional searches, in which the search engines always present complete, holistic results, and in which the search engines present knowledge (linked facts) and not just information (facts). It is preferred to use knowledge search to information search and retrieval in this research project. With Internet search, a user asks a question and gets a link to a “top ten list” of sites that might have what the user needs. Providing the search environment, knowledge search tends to provide the answer. The generic term knowledge search is used, in its broad sense, to collectively refer to these fields and application areas. Domain knowledge is needed to interpret queries and generate responses. “One category of queries which requires domain knowledge beyond data consists of questions asking about things which reflect general rules which apply in the domain, rather than things which happen to be accidentally true of the current data.” Kao at el. (1998) note.

Most enterprises use client/server network architecture. Compared to peer-to-peer (P2P) architecture, it is more secure, and easy to provide system-wide services. Some classified government may not have need for extensive file sharing and may prefer the more isolated architecture environment offered by a traditional mainframe or stand alone power workstation environment. Computers in P2P architecture act both the client and the server. The models of P2P searching are not commonly used in enterprise search because files stored in the P2P directory are accessible by people outside the local network. The indexing process of client/server reads the text of the documents, while P2P search reads the name of the files. Within client/server architecture, additional permissions and business rules can be set to allow different types of access to different users. Information search based on client/server structures solved the problem of information resource location to certain extent. However, it presents some shortcomings. To download a lot of futility information contained in web pages, it affects search speed, and waste web information resource. Frequent changes of web pages and websites decline the actual effect of search results. In addition, since client/server structure is based on message-passing and remote procedure call, the connection between auto search program and website should be ensured during search process. Therefore, information search based on client/server structure rely on effectual web bandwidth. It is not suitable for enterprise condition which has narrow web bandwidth, unstable connection with distributed systems.

Agent technology plays an important role in decreasing network dependence. To apply Agent technology in web information search, it could dynamically move to a server to search according to user request, and send back search results of the server. In this way, an Agent accesses to a server's resource directly, avoiding a mass of data transferred in the web. Therefore it declines the dependence of system on bandwidth, and keeps stable work under the unstable network connection. Meanwhile, Agents could do parallel searches in different network nodes to reduce search time. Therefore, this Multi-Agent knowledge search methodology solves the problem of inefficiency of information retrieval across sites. Note that more detailed introduction to the Agent technology is given in the following section. For the established enterprises, to move away from or to upgrade their search systems is not so easy. They require for the continuous operation of their critical functions. The Agent-based encapsulation and integration approach can facilitate this plug-and-play kind of integration of heterogeneous software and hardware systems.

Most broadly used search models use word-frequency for recognition, neglecting real content of information results in mis-recognition of information meaning (semantics). Search engines require users to search using standardized language and format to express their requirements, which may result in incomplete requirements or biases. Engineers may get documents which contain the key word but not relevant to the subject, and miss some relevant ones. Agent is feature of learning, which supports the efficient knowledge search. It collects, indexes, and filters information using domain knowledge structures, user behaviours and interests. Search will be more intelligent with the assistance of Agents.

In manufacturing applications, the available powerful web search engines are not designed for enterprise information and knowledge search. The link structure of enterprise repositories is quite different from the one on the Internet. The lack of highly hyperlinked nature hinders the use of some techniques based on link analyse. Or the application of these techniques is not as effective on enterprise search as on the web. The low relevance of retrieved information requires the employment of other techniques to improve.

For example, in product catalogues, each item has unstructured text, as well as structured attributes. For example, an automobile typically has a description and attributes, such as year, model, and price. A typical query is a conjunction of an arbitrary text query ("Leather Trim" AND "All-wheel Drive") and a parametric query on structured fields (Manufacturer = Toyota AND price < $30,000 AND year > 2000). Whereas the text query is within the purview of conventional information retrieval systems, the parametric query is traditionally handled using relational database systems. Modern search engines perform both functions for applications such as e-commerce and marketplaces where scalability and cost-performance are critical.

Using a relational database management system (RDBMS) to solve the problem would result in unacceptably poor query responses. Text search extensions in RDBMSs do not support powerful free text query capabilities (e.g., fuzzy search), and are not cost effective for search. In addition to being able to sort along attribute values, it is crucial to be able to rank the results based on the query. This enables efficient guided navigation of results, allowing the user to progressively refine (or relax) the query.

Another key characteristic of data is the structure within the data itself. With the increased adoption of XML (Extensible Markup Language), the ability to search and retrieve specific portions of documents (e.g., specific elements in XML) is mandatory. Query semantics like XQuery (an XML query language) will be supported, but with the added ability to handle unstructured text and fuzzy constructs that databases do not handle elegantly (e.g., spelling errors). The ability to dynamically construct virtual documents that consist of relevant portions of many documents will be critical. What the end users want in the future is not just a matching document, but something that represents an answer.

In manufacturing companies, there are additional security, reliability, and company policy, which even complicate the problem. The privacy of individuals' personal data is protected and unable to obtain without authorization. Information spans organisational boundaries. Not all the information required for a task is available as indexed content. Even if an organisation has access to relevant content (e.g., on the Web), there are cases where it cannot be indexed (e.g., security) or is forbidden from being indexed because of legal constraints. Further, in large organisations, different departments commonly index silos of information via different software systems or applications.

In such cases, Multi-Agent search is a better way to provide a single point of access to data from enterprise repositories and applications, as well as external subscription sources and real-time feeds. Multi-Agent systems add further value via ranking, filtering, duplicate detection, dynamic classification, and real-time clustering of results from disparate sources that may not be under the jurisdiction of the enterprise. Database vendors provide federation across disparate relational databases, but Multi-Agent search for unstructured data provides different challenges and opportunities.

Enterprises have good traditional search methods for structured information. However, unstructured information is not efficiently searched. A specially designed enterprise search engine could solve this kind of problem. However, there is less research effort devoted to the development of advanced search systems for engineering applications. It would benefit enterprises significantly if the technologies in the powerful modern search engines can be applied to engineering applications.

2.6 Agent Technology for Design and Manufacturing

Agents are a paradigm for developing software application. Jennings and Wooldridge (1998) suggested that Agent is a computer system situated in some environment, and capable of autonomous action in this environment in order to meet its design objectives. These systems can act without intervention from humans and take control over their actions and internal state.

This new technology met well the requirements in all domains of manufacturing where problems of uncertainty and temporal dynamics, information sharing and distributed operations, or coordination and cooperation of autonomous entities had to be tackled (Monostori et al. 2006). A Multi-Agent system (MAS) contains an environment, objects and Agents, relations between all the entities, a set of operations that can be performed by the entities and the changes of the universe in time and due to these actions (Ferber 1999).

Long ago, manufacturing system theory suggested cautious organising principles for systems exposed to substantial internal and external uncertainties (Hatvany 1985). These principles say that it is better to recognise ignorance than to presume knowledge, and it is better to regard the future as unpredictable than only to be prepared for expected events. Accordingly, manufacturing systems should be organised as loosely coupled networks of communicating and cooperating components or Agents (Duffie 1990). Agents address autonomy and complexity problems. They are adaptive to changes and disruptions. Thanks to its properties as autonomy, responsiveness, redundancy, and openness (Bernon et al.

2005), Agents could be designed to work with uncertain and incomplete information and knowledge. Now, Agent technology is applied in many areas of the manufacturing domain. This section gives examples of Agent technology applications in the domain of engineering design, process planning, and production planning and resource allocation. It is applied in other domains such as production scheduling and control, process control, monitoring and diagnosis, enterprise organisation and integration, supply network, and assembly and life-cycle management.

Computer supported engineering design utilizes the Agent-based approach to satisfy new requirements like collaborative design and customized design. The Agents do not need to be co-located as normal modular system structure for collaborative design. The Agent-based structure relies on shared ontology for sharing knowledge and requesting services across disciplines and among different representations. Multi-Agent system design problems and the available knowledge can be structured in more complex ways than single-Agent design. Collaborative and life-cycle approaches to design are utterly based on interaction. Early examples in distributed collaborative design system like PACT (Cutkosky et al. 1993) based on interacting engineering tools, which were wrapped up as Agents. Other examples include SHARE (Toye et al. 1993), First-Link (Park et al. 1994), ABCDE (Balasubramanian et al. 1996), RAPPID (Parunak et al. 1999), and Facilitator (Sun et al. 2001). SHARE considered a wide range of information exchange technologies in order to help engineers and designers collaborate in mechanical domains. First-Link was a prototype system for concurrent design of aircraft cable harnesses, based on the SHARE infrastructure. ABCDE (Agent-Based Design Concurrent Design Environment) was an architecture for the integration of design, manufacturing, and shop-floor control activities. The RAPPID (Responsible Agents for Product Process Integrated Design) system relied on three strategic mechanisms: autonomous Agents as a way to distribute decision-making among a community of human beings and computers, market-based control as a mechanism for coordinating distributed decision-making and set-based design as a means of making decisions in parallel, regarding partial design. Facilitator presented a MAS with a facilitator Agent, a console Agent and some service Agents. The facilitator was responsible for the decomposition and dispatching of tasks, and for resolving conflicts of poor designs. The console Agent acted as an interface between designers and the system. Each service Agent is used for modelling different product development phases.

Agent technology has been applied in process planning to handle inconsistent pieces of design and technology related knowledge. In manufacturing, computer technology has been used to plan the process of producing a design part. Manual process planning is based on a manufacturing engineer's experience and knowledge of production facilities, equipment, their capabilities, processes, and tooling. This process is very time-consuming and the results vary based on the person doing the planning (Crow 1992). Computer-aided process planning (CAPP) simplifies and improves process planning and achieves more effective use of manufacturing resources. For example, with Agent model, Machining CoCAPP (Machining Cooperative Process Planning System) (Zhao et al. 2000) developed a extensive industrial system to generate flexible and upgradable process plans. In recent research, CAPP focused on conflict between available resources and actual lot-sizes or due dates. Process planning knowledge was used for short-term scheduling decisions at the shop floor. MAS concepts were particularly well suited to this integration.

Production planning and resource allocation is the process of selecting and sequencing activities, and allocating available resources, so that they should achieve the goals of an enterprise and satisfy a set of domain constraints. Lee et al (2003) suggested a market-based negotiation mechanism, called precedence cost tâtonement (P-TÂTO). The system was composed of a project manager Agent, task Agents, resource manager Agents, resource Agents, and coordinator Agents. The project manager Agent maintained the project milestones, the project activity network and each task's resource allocation information. A task Agent was in charge of its own single task. A resource manager Agent was in charge of monitoring and coordinating a set of resources. A coordinator Agent was responsible for coordinating multiple resource allocation markets in the virtual market model. The simulation results based on P-TÂTO indicated high levels of solution quality and computational efficiency.

The exploitation of Agent-based technology in production planning was addressed in the ExPlanTech project (Říha et al. 2002). There, one of the five intra-enterprise Agents introduced was the production planning Agent which was in charge of constructing an exhaustive, partially ordered set of tasks to be carried out in order to accomplish the given project.

Five levels of production planning in a reconfigurable enterprise (RE) were distinguished by (Bruccoleri et al. 2005) (see Figure 2.2). The authors claimed that the successful tools for operation management in RE needed to be based on the decentralization of the decisions where each entity in charge of specific planning decisions makes its own decision autonomously, while global planning decisions were achieved by means of coordination and negotiation among them. The various levels differ from each other on the planning horizons, the planning issues, producers and mechanisms. The negotiation model was compared with a centralized solution, showing the benefits of the Agent-based approach. However, for the continuity, there should be an additional level indicating the negotiation between order agent and source agents.

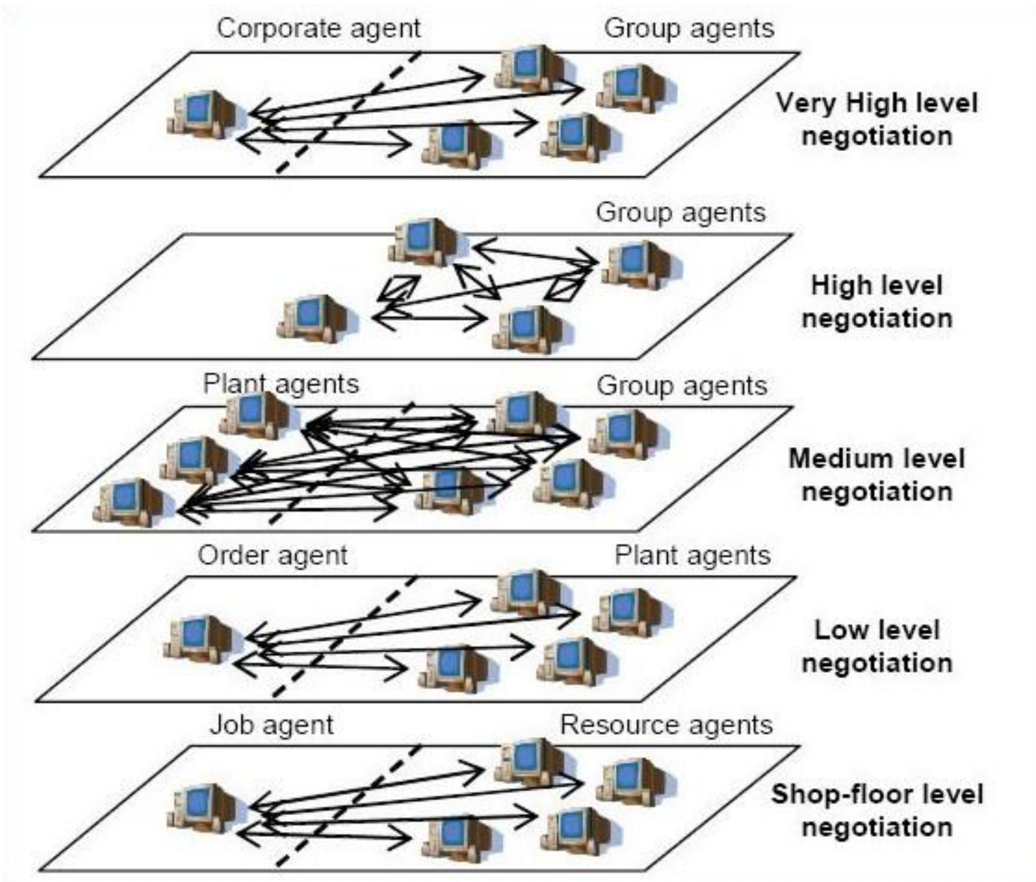


Figure 2.2: Negotiation Models for Different Planning Levels of an Enterprise (Bruccoleri et al. 2005).

Above examples of the Multi-Agent approaches for manufacturing reveal the new and interesting possibilities of Agent technology to satisfy a number of fundamental manufacturing requirements. Agents encapsulate manufacturing activities or wrap legacy software systems in an open, distributed intelligent environment. Product design, process planning, and production planning examples above all show such functional Agents. Such an Agent-based encapsulation and integration approach can significantly facilitate the integration of heterogeneous software and hardware systems. It can also enhance communication and cooperation among departments within an enterprise and among enterprises. Agents are implemented to represent physical manufacturing resources (e.g., machines, tools, and operators), products, parts, and operations. Human operators represented by software Agents and interacted with Agents through user interfaces will cooperate with manufacturing resources, particularly those intelligent machines to complete manufacturing tasks more effectively. Different from centralized approaches, Agent-based manufacturing scheduling and control systems can respond quickly to dynamic changes through local decision making. This approach also provides much better fault tolerance than traditional approaches because of its loosely coupled system architecture and the autonomy of individual resource Agents. By using this approach, manufacturing enterprises will be able to reduce their response time to market demands and therefore to win in a globally competitive market.

The rich variety of Multi-Agent approaches clearly shows the application potential of Agent technology (Monostori et al. 2006). However, the adoption of Agent technology has not yet entered the mainstream of commercial enterprises. Only a relatively small number of industrial applications of Agent technology are visible (Luck 2005). A considerable potential exists for other enterprises to apply Agent technology.

2.7 Agent Technology for Knowledge Search

Information search and retrieval is one of the fundamental applications of IT. However, the applications of today and tomorrow require better than the traditional exhaustive search technologies. Traditional methods fail to break the complexity barrier caused by the combined information explosion that characterises these large, real-world environments. Knowledge search, as referred to in previous sections, is used to collectively address these fields and application areas.

Agent technology is deployed in knowledge search area as well as personal assistance, network management, software distribution, and so on. A number of Agent approaches for knowledge search have been developed or are currently under development in academic and commercial research labs. But they still have to wait to make it out to the real world of enterprise users broadly. However, the ambitious and promising goal to satisfy the requirements of enterprise users appears to be realizable in the future. This rest of this section reviews the efforts made by previous researchers on facilitating knowledge search by software Agents, Multi-Agent systems, and Multi-Agent approaches.

To support conceptual information retrieval capabilities, natural language input, as well as individual user preferences, a Scalable Agent-based Information Retrieval Engine (SAIRE) was developed for public accessing to Earth and Space Science data maintained by NASA and NOAA over the Internet (Odubiyi et al. 1997). Figure 2.3 displays the Multi-Agent architecture in the framework of three Agent groups: the User Interface Agents, a Coordinator Agent, and the Domain Specialist Agents. SAIRE employed the Agent technology to realise natural language query, conceptual search, and collaboration among multiple Agents to access and deliver information to the public from distributed sources over the Internet.

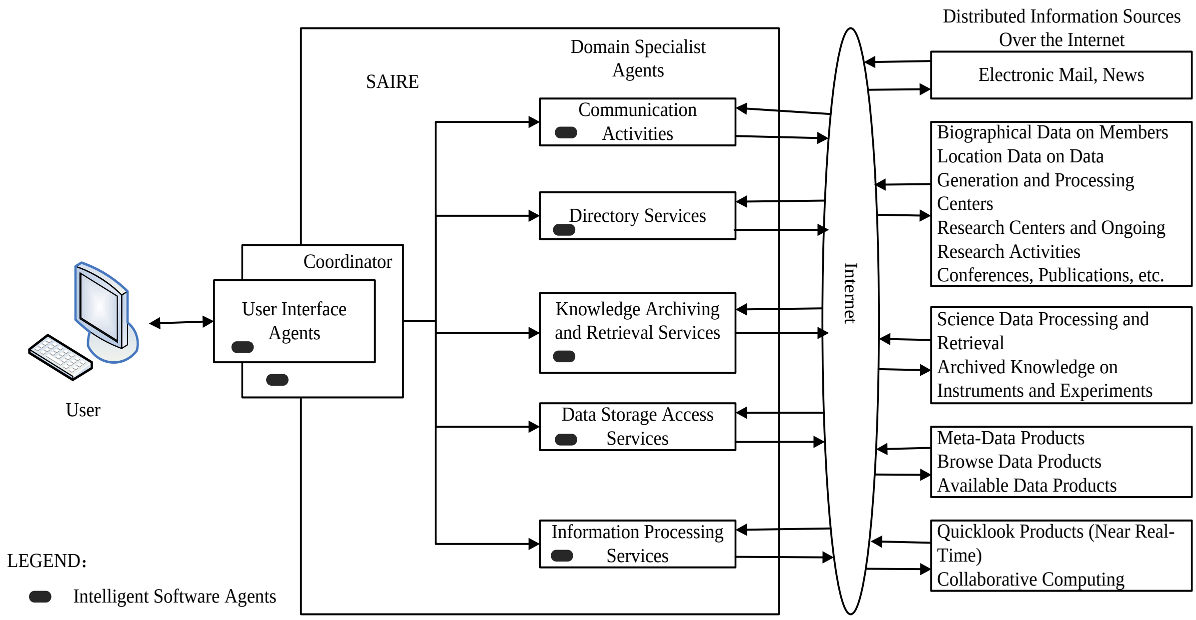


Figure 2.3: Architectural Overview of SAIRE (Odubiyi et al. 1997)

In 1996, a Remembrance Agent (RA) was implemented to display a list of documents which might be relevant to the use's current context (Rhodes and Starner 1996). It ran continuous searches for information that might be of use in the user's current situation without user intervention. It was an early Just-In-Time Information Retrieval (JITIR) Agent. In 2000, Margin Notes was implemented (Rhodes 2000). It was a web-based Agent that automatically annotates web pages and adds hyperlinks to related documents as they are being loaded into a browser. Both RA and Margin Notes provided potentially useful information based on a user's computational environment (e.g., email, documents, and web pages being viewed). The Jimminy system (Rhodes 1997) provided information based on use's physical environment (e.g., location, people in the room, time of day, and subject of the current conversation). All these three implemented JITIR software Agents used the same information retrieval back end.

The open Internet and WWW allow users to access multimedia data and knowledge located throughout the world. An algorithm was presented by Yan et al (2003) for video retrieval. It retrieved a text description plus images or video and relevant shots by fusing the retrieval results of multiple retrieval Agents.

Social networking sites as Facebook, Flickr, and MySpace include large volumes of user information. An Agent-based framework called Social Network-based Item Search (SNIS) was developed to mine the social network of a user to improve search results (Gürsel and Sen 2009). An Agent-based photo searching system was developed following the framework for flickr.com. By mining past user activities in the social network, the user Agent determined the likelihood that the user will query about a certain item. The framework was effective of presenting search results in social networking sites.

Agent technology would play a significant role in the evolving knowledge infrastructure, if it is proven to be useful to users and enterprises for knowledge search and management. Thus, Agent-based knowledge search attracts the attention of both the community in industry and academia, and professional and private Internet users.

2.8 Summary

This chapter introduced knowledge, knowledge management, knowledge search technologies, and enterprise knowledge search. Agent technologies and its applications in manufacturing domain and knowledge search areas have also been introduced. The literature survey confirms that the fast developing and successful internet search engines have significant potential in enterprise knowledge search, especially in the new product development process of the global manufacturing business. By comparison with general information search engines via the web, there is by far insufficient research and development work devoted to enterprise knowledge search. There is an urgent need to understand the specific problems, challenges and requirements of product development processes so that the lasted search technologies can be further developed for the manufacturing sector which is critically important to the world economy. The advanced Agent technologies have the potential to significantly improve the search capabilities in the dynamic environment of the real-life manufacturing business. Based on the literature survey, an advanced knowledge search framework will be proposed and developed which is new and also complimentary to existing information search technologies.

CHAPTER THREE: INDUSTRIAL INVESTIGATION

This chapter introduces an in-depth industrial investigation carried out during the project. In the early stage, the main purpose and effort devoted to is the understanding of the industrial collaborator's business processes, product range, organisational structure and its AS-IS situation of information and knowledge management. Then a more focused investigation into industrial problems and their requirements for knowledge management and search systems were carried out. In the later stage, examples of actual knowledge have been captured and used for case studies to evaluate and test the proposed knowledge search methodology. This Chapter will present the findings of the industrial investigation. Firstly, an overview of the collaborating company is presented. Then, the new product development project teams, products, and processes are introduced, followed by the introduction of the company's AS-IS IT systems. Finally, the knowledge management and knowledge search requirements in the industrial background are analysed. The example knowledge for case study will be presented in later Chapters.

3.1 Overview of the Collaborating Company

The company which provided industrial case study, example products, business process information and engineering knowledge is Commercial Aircraft Corporation of China (CACC). The company was recently established in May 2008. Its main business is the development of large civil passenger aircrafts (CACC 2009). As a state-holding limited company, CACC was founded with registered capital of 19 billion RMB (Chinese currency). Its headquarters is located in Shanghai. The corporation has over 6000 employees. The key business of CACC at present includes design and manufacturing of new regional aircraft ARJ21 and a new large passenger aircraft C919, involving aircraft design, aerodynamics design, structure design, avionics, flight control system, manufacturing process, aviation engineering, flight training and supplier management.

CACC was established by drawing expertise from a super state-owned enterprise: Aviation Industry Corporation of China (AVIC), which is an important cooperation partner of CACC and one of its main shareholders. As a state empowerment investment corporation, AVIC has nearly 200 subsidiary companies, involving 21 listed companies. It has about

400 thousand employees (AVIC 2009). In 2009, it generated a turnover of 191 billion RMB, and 9.7 billion RMB profit (Yan 2010). AVIC divides its production business into ten units, known as defence, transport aircraft, engine, helicopter, airborne equipment and system, general aircraft, aeronautical research, flight experiment, trade logistics, and asset management (CACC 2009).

The First Aircraft Institute of AVIC (FAI) was established in the 1970s. It is now responsible for the development of fighter-bombers, civil aircrafts, special mission aircrafts and transport aircrafts. It has around 3000 employees, including about 2000 technicians. Its headquarters is located in Xi'an, and a branch of the institute is located in Shanghai. There are more than 20 design and research departments in the institute, covering the following specialties: aircraft configuration, aerodynamics, structure design, strength analysis, landing gear, avionics, flight control and hydraulic system, electronic system and instruments, armament and weapon system, reliability, quality control, airworthiness, standards and materials, flight test, ground test facility, market development, computer center, technical information and metrology center, and so on. Main aero products of FAI are military aircrafts, civil aircrafts, light aircrafts and its series, aircraft ground support equipment, and airborne software and hardware (FAI 2007). The branch of the institute of FAI based in Shanghai was reconstructed and merged into CACC in 2008. It is responsible for the design and development of new regional aircraft ARJ21 at present, and support large aircraft project development. This research is supported by this part of the company's business.

The overall structure of the business relationship is illustrated in Figure 3.1. CACC headquarters is located in Shanghai. The branch of FAI in Shanghai belongs to CACC at present. Although they are in the same city, they are on different sites. Shuttle Bus runs every working day between them. The industrial investigation and case study have been untaken in the two site of CACC in Shanghai.

As Figure 3.1 shows, AVIC has 14 high level departments including management department, strategy planning department, operation management department, human resource department, finance department, and so on. The corporation has many subsidiary companies and institutes. CACC is cooperation partner of AVIC. AVIC is one of the major shareholders of CACC. CACC has 13 departments including development planning department, marketing department, quality technology department, project management department, and so on. On top of these departments, there are general manager, board of directors, and shareholders.

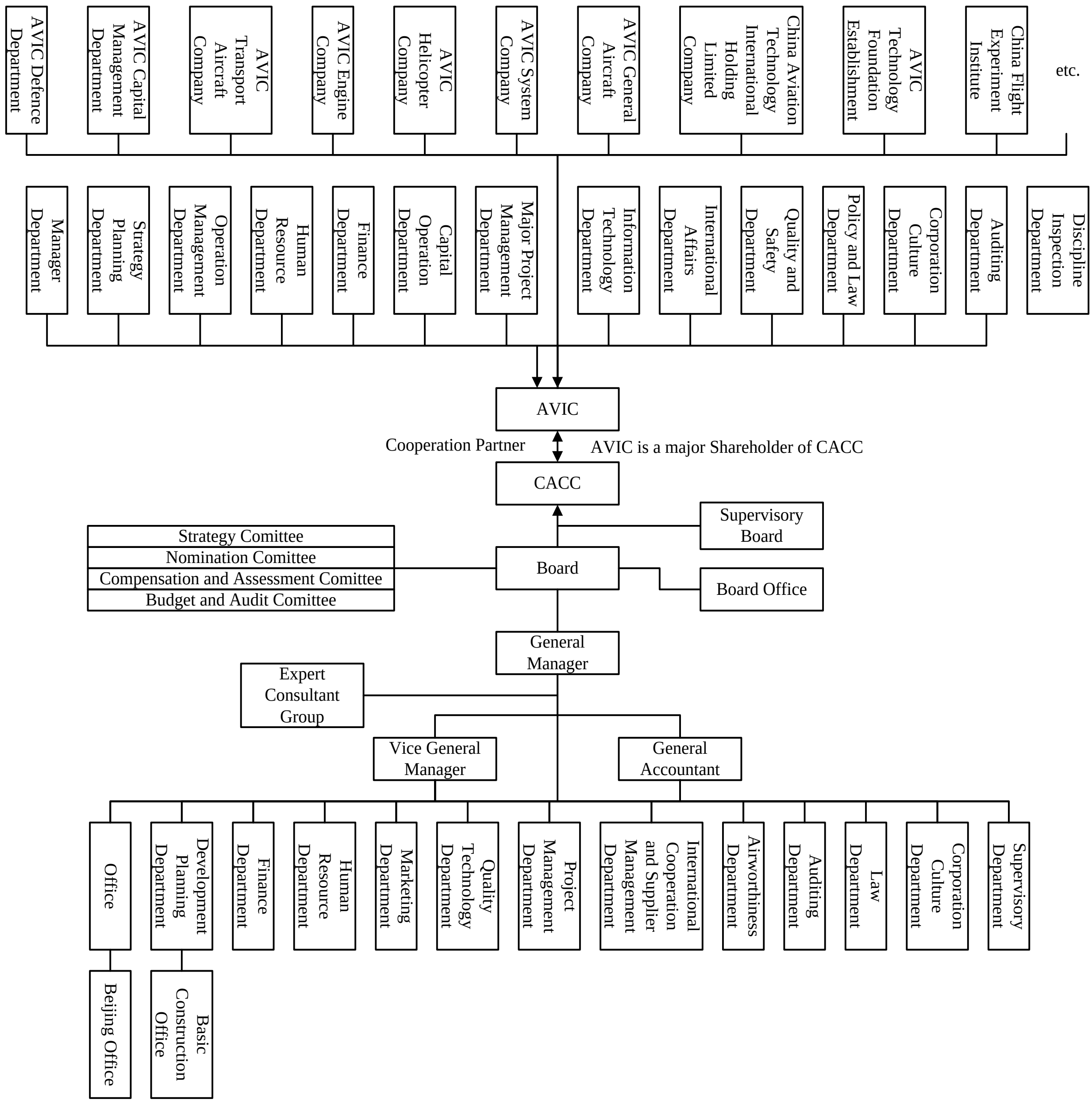


Figure 3.1: Organisational Structure of the Company

3.2 New Product Development Project Teams

To develop a new aircraft is an extremely complex task. In CACC, there are nine project teams. Senior managers of these teams report to vice manager of the aircraft project. The vice project manager reports to senior executive vice manager. The nine senior managers are in charge of different projects, namely aircraft design project, preparatory work project, flight and experiment project, manufacturing and standard project, material and processing project, engineering technology information project, project engineering project, technical support project, and administration and business operation project (Wang, 2007). Each project refers to several domains involved in the aircraft development. A list of the domains involved in each project is given in Table 3.1.

These nine project teams are in charge of the research, design, and experiment of the aircraft, and flight tests. With the Type Certificate (TC) based on airworthiness management requirements, the final produced aircraft would satisfy market requirements and customer requirements. In reality, the project teams are not fixed. Project members change according to different types of aircraft and task requirements.

Aircraft design project team is the largest one. As seen in Table 3.1, this team takes charge of aircraft structure design, system design, decoration, assembly, and on-board equipment. It manages the cooperation of supports as well. Preparatory Work Project team is responsible for project plan, including the new technology development, new or modified aircraft pre-analysis and model development, aircraft overall design, layout and performance target decision, aircraft energy efficiency, compatibility of aircraft and airport, system analysis, and overall cooperation. Flight and experiment project team does work about structure mechanics, weight mechanics, aerodynamics, acoustics and environment control, lab experiment, system performance, structure experiment, vibration experiment, human resource factor, and flight experiment.

This project will mainly focus on the Aircraft Design Project and use its structure design function as example for case study. The structure design function is the basis of the aircraft development and the structure design team cooperate with all other functional teams.

Project	Domain
Aircraft Design Project	Structure, System, Decoration, Assemble, On-board Equipment, Coordination Support.
Preparatory Work Project	New Technology Development, Advanced Project (pre-analysis and developing model), Overall Design (performance placement), Aircraft Energy Efficiency, Compatibility of Aircraft and Airport, System Analysis, Overall Cooperation.
Flight and Experiment project	Structure Mechanics, Weight, Aerodynamics, Acoustics and Environment Control, Lab Experiment, System Performance, Structure Experiment, Vibration Experiment, Human Resource Factor, Flight Experiment.
Manufacturing and Standard Project	Producibility, Commissioning System, Standard, Design Service, Sample Exhibition, Simulator.
Material and Processing Project	Metallurgy, Non-destructive Test, Mechanical Performance, Metal Bonding (Fastener, Welding), Decoration Bonding, Transparent Material, Composite Material, Plastic, Structural Joint, Coating Sealant, Chemistry (Processing/Analysis) Electric, Rubber, Composite.
Engineering Technology Information Project	Application of CAD/CAM, Engineering Calculation Center, Data System, Geometry System, System Integration, Project Design Definition.
Project Engineering Project	Project Architecture Management, Information Management System, Approval and Distribution of Blueprint, Project Process Plan Management, Validity Management.
Technical Support Project	Airworthiness (Type Certificate), Law Affair, Safety and Reliability Approval, Maintainability, Engineering Procedure (Work Standard).
Administration and Business Operation Project	Administration Management, Contract Service, Engineering Facilities, Design Cost, Business Strategy, Project Budget, Project Introduction, Training, Documents, and Library.

Table 3.1: Project Teams and Project Domains.

3.3 The Products of the Company

FAI has designed and jointly developed more than 10 types of aircrafts. At present, the institute is undertaking several national projects, as well as ARJ21 regional jet and small business aircraft LE-500. Aero products developed by FAI include military aircrafts, civil aircrafts, light aircrafts and its series, aircraft ground support equipment, airborne software and hardware. Non-aero products developed by FAI are ACM Series Aircraft Air-conditioner, 603FD Series Vacuum-freeze-dry equipment, Series Airport System Products, Automobile Fittings, Electronic and Electrical Equipment, Aircraft Ground Trainer, Computer Network Construction and Relational Software Development, intelligent localization of the airport light malfunction, power on-line control system, oil-field slurry lifting equipment, high-tech agriculture intelligent greenhouse, automatic fluid filling machine, Baosteel spare parts, IT software training and service, computer auxiliary design, and so on (FAI, 2007).

AVIC develops several types of aircrafts, like military aircrafts, transport aircrafts, training aircrafts, helicopters, general aircrafts, and unmanned aerial vehicles. In addition, it is doing developing new series engines of turboprop, turboshaft, turbojet and turbofan.

CACC is a recently established corporation. It focuses on the development of Chinese first turbofan regional aircraft ARJ21 with independent intellectual property. ARJ21 is designed for economical use. It satisfies the requirements of take-off from and landing on airport with high temperature and high altitude. And it realises obstacle navigation on complex air routes. This aircraft is the result of extensive cooperation. The airframe parts are manufactured in several different aircraft manufacturing companies in China. The engine, avionics, power systems are purchased globally. Except ARJ21, CACC emphasizes on the development of large passenger aircraft. C919 is a mid-short distance commercial aircraft. The project was set up in November 2008. It plans to have C919 first flying in 2014 and obtain airworthiness certificate and deliver to customers in 2016.

This project will use the structure design of aircraft nose of C919 as the environment for knowledge search. This is recommended by the enterprise, because of the complex process. A conventional aircraft has nine principal structural units: the fuselage, engine mount, nacelle, wings, stabilizers, flight control surfaces, landing gear, arresting gear, and catapult equipment. Figure 3.2 shows the construction of the nose of C919, consisting of clapboard, frame, floor, doorframe, skylight frame and air damper, nose cap, and hatch cover.

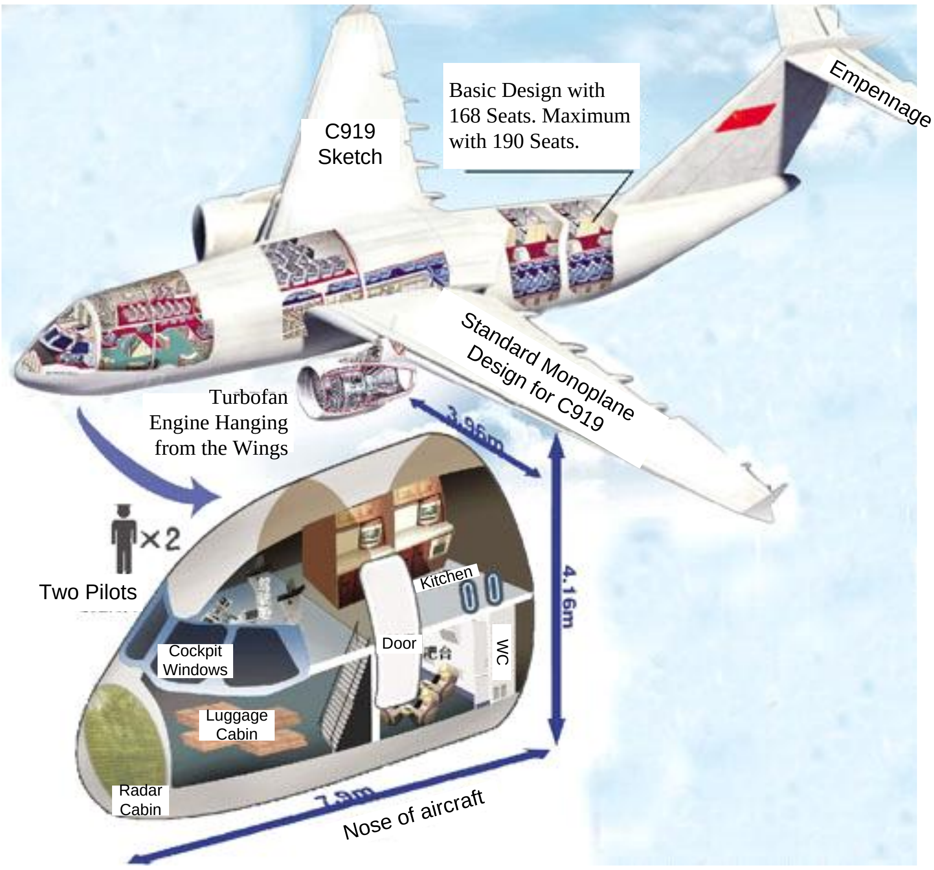


Figure 3.2: Construction of the Nose of C919 (Source: CACC)

3.4 Product Development Business Process

The development of an aircraft is a complex process. Figure 3.3 displays various jobs that are conducted in parallel during one phase of the design and development of the aircraft. Before the more 'detailed' design and development stage, there is a lot of pre-arrangement work to do. Concept design is done during this period, starting from layout of the scheme of the aircraft. Engineers reason about the overall concept (configuration) in consideration of aerodynamics arrangement, propulsion device, system, weight and barycentre, as well as calculating intensity and aircraft price. Then the concept is optimized and new one is generated, expressing the technologies, processes and materials. This progress may be repeated to adjust and optimize the concept with respect to each factor. Once the concept is approved, engineers develop a sketch with evaluation technology. Customers evaluate the design sketch with their requirements. The sketch is then set down. And the process comes into detailed design and development stage. As shown in Figure 3.4, engineers produce blueprints, generate calculation reports, test reports, flight-test reports ad training documents. This progress is repeated as well. Related work is done on parallel. An aircraft assembly drawing (product model) is produced after the individual blueprints are approved. Then different tests and experiments are run till the model is satisfactory. Manufacturing stage starts after the validation of Type Certificate (TC), Production Certificate (PC), and Airworthiness Certificate (AC).

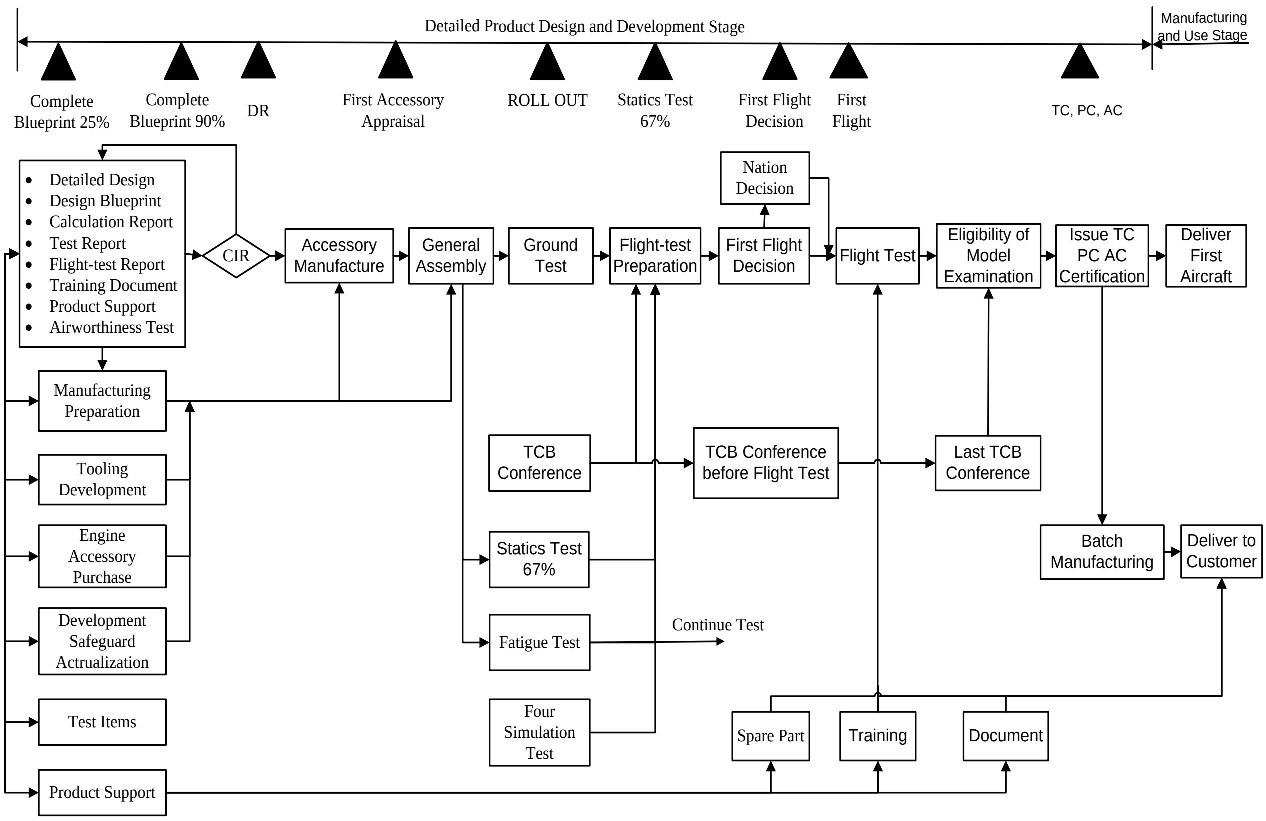


Figure 3.3: Business Process of the Aircraft Development (Source: FAI)

For the large passenger aircraft C919, there are ten tasks formulated at the detail design and development stage shown in Figure 3.3. The first step is interpreting the technical requirements of aircraft design. Secondly, generate preliminary overall technical structure of C919. Third, select wind tunnel test type, and design preliminary loading test. Fourth, select domestic suppliers, and roughly select foreign suppliers. Fifth, select the engine technical model. Sixth, roughly digitalize sample aircraft. Seventh, manufacture physical sample of aircraft nose. Eighth, manufacture display sample. Ninth, begin to tackle key technologies. And tenth, carry out the economic feasibility evaluation of technology.

The above are main tasks during this stage for the development of C919 aircraft. This study is focused on the nose of the aircraft. Figure 3.4 is the design process of C919 nose. The initial structure design is conforming to airworthiness standard. It has various requirements about space, horizon, layout, aerodynamics, and so on. The shape of the nose affects the icing issue. The Structure Group cooperates with the System Group for better de-icing solution. The corrosion of deicing fluid relates to the Material Group. Materials used need the Intensity Group to ensure rational intensity. Material intensity influences aerodynamics design. The Aerodynamics Group collaborates with the Structure Group on docking. Structure design requires large windows for pilot horizon. The Decoration Group considers the ergonomics design to satisfy it. The Decoration Group reckons for the barycentre problem as well to adapt to aerodynamics design. Therefore, each Group has its main focus, and has to compromise with other Groups on the development. There are various conflicts as well. For example, structure design requires large windows for pilot horizon. Aerodynamics Group designs the shape of the windows. And Intensity Group limits the windows size. Groups have to interact on each other.

The product development is an iterative process which could be described as follows. The Structure Group generates the initial structure of the nose. The Aerodynamics Group calculates and provides suggestions for structure modification. The Loading Group analyses the load. The structure is verified by the Construction Group and the Intensity Group. Then the Decoration Group conducts internal layout, reckoning in ergonomics. Groups interact on each other. All requirements, conflicts and impact factors are taken into consideration. Normally, after three to five iterations, engineers will have optimized a reasonable design which is acceptable to every Group.

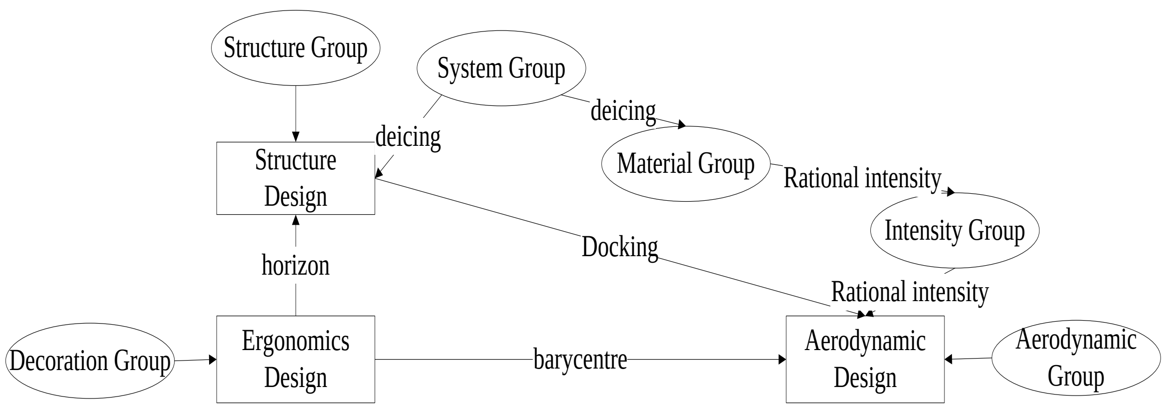


Figure 3.4: C919 Nose Design Process

3.5 The Information Systems Used in the Company

The empirical investigation was taken in CACC and FAI. CACC is the main manufacturing factory which takes charge of disseminating assembly and manufacturing data. Data include file data and drawing data. File data is usually formatted as Excel, Word, and PDF files. Drawing data is in the form of CAD (Catia) Files. In practice, physical files required include blueprint and related engineering order, and engineering bill of material (EBOM). Figure 3.5 shows the distribution of physical and electronic files in CACC.

The Archive Department of CACC receives data mainly from FAI through the Collaborative Product Commerce (CPC) platform which is a Product Data Management (PDM) or Product Lifecycle Management (PLM) software platform. New types of files are assigned to the responsible group by C919 project office. Distribution routes are planned by the Planning Group. Then files are sent to responsible sites. Physical files are sent to the Drawing Office for printing.

For CACC, product data management system is operated on CPC platform. FAI uses the Virtual Product Management (VPM) system to manage drawing data, and the Assembly Order (AO) system for technique orders. Data transferred between FAI and CACC is based on CPC platform. Data from CACC to other three manufacturing sites, like Xi'an Aircraft Industry Company, is disseminated through CPC platform as well. These three systems, i.e., CPC, VPM, and AO, run on three different servers. CPC and VPM systems are developed by the PTC Company.

Data is collected through questionnaires and semi-structured interviews. The questionnaires are designed for designers and archive people separately (refer to Appendix E). The questions are approved by the manager. Twelve participants are included in this investigation. The participants are selected according to the recommendation of the manager, based on their responsibility in the group.

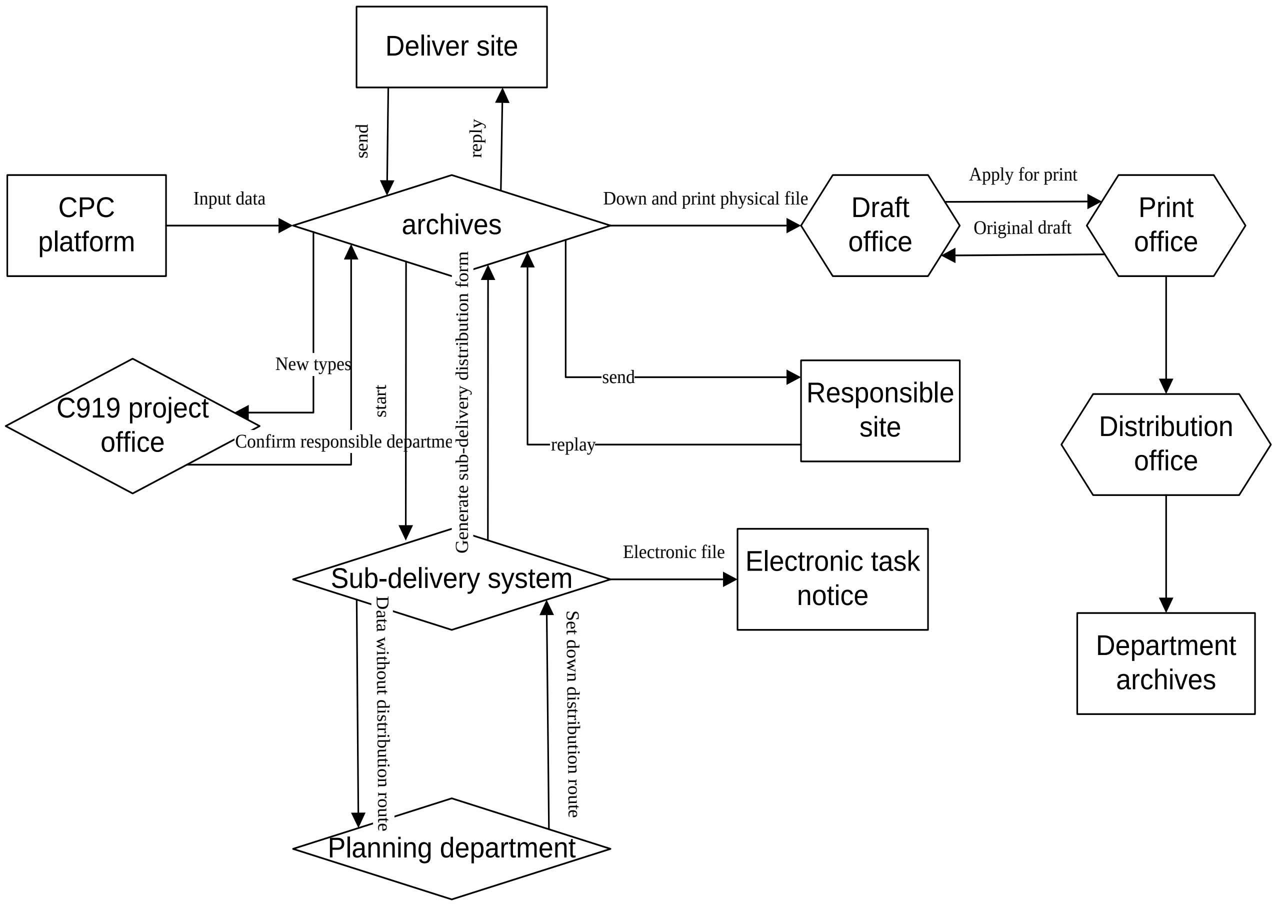


Figure 3.5: Distribution of Physical and Electrical Files.

Data collected from CACC show that some members of staff in the Development Department are unsatisfied with the efficiency of searching for project management documents. Some members of staff in the FAI institute complained that documents are difficult to find if they are unfamiliar with exact area. And some mentioned that the system should be more human friendly since they would not like to spend much time accepting new systems.

A high-level view of the computer network collocation is presented in Figure 3.6. C919 Group employs CPC platform as the main information integration and management tool for development sites and manufacture partners based on different provinces of China, as the network layout displays.

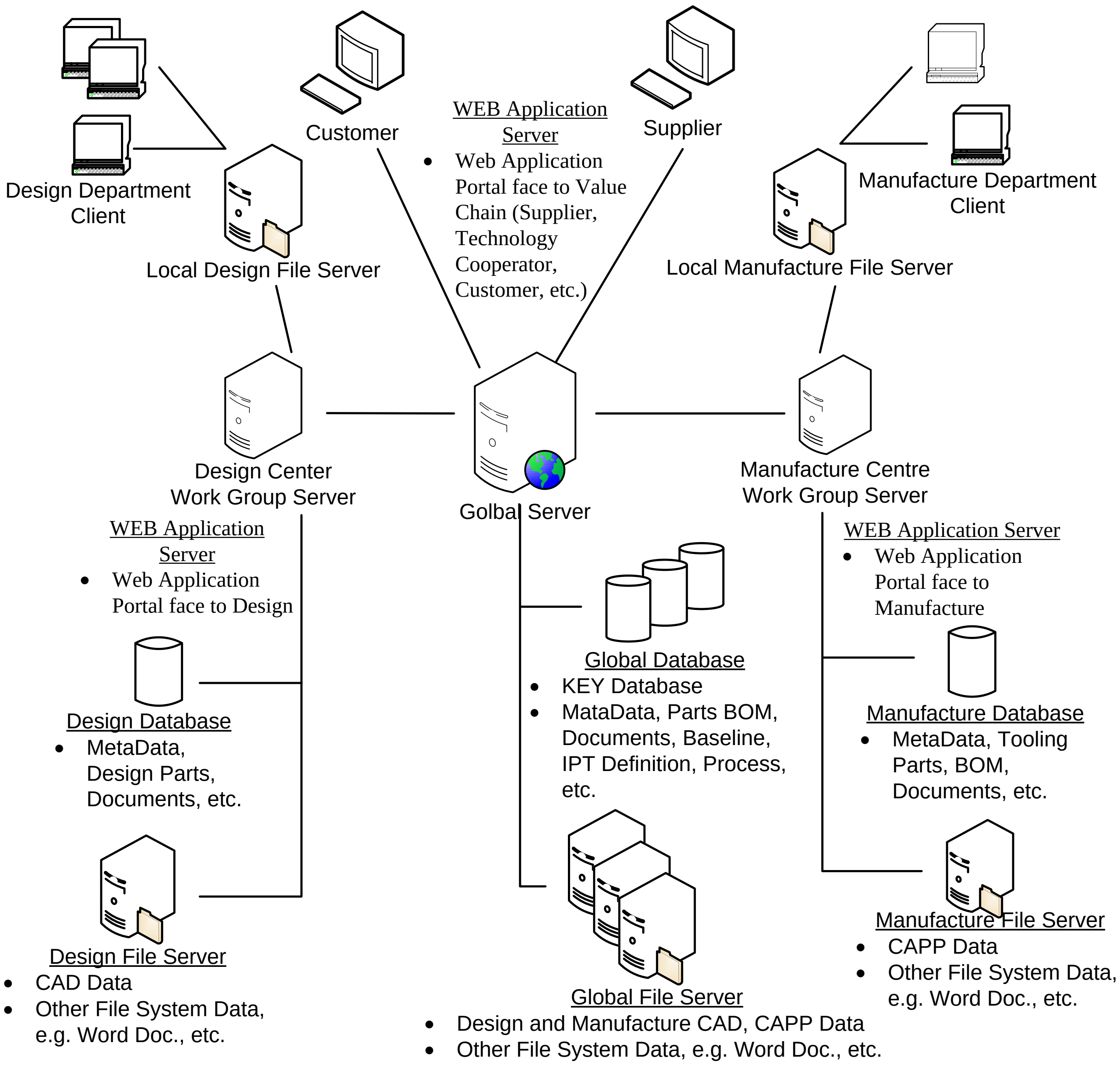


Figure 3.6: Typical Computer Network Collocation in Engineering Enterprises.

3.6 Knowledge Management and Knowledge Search Problems and Requirements

Different investigation methods were used in the industrial investigation to collect data. A survey of the CACC company and its FAI Institute was carried out. Interviews were conducted with the people catalogued as Design, Development, Archive, and Organisation. Example quotations of the interviewees for each problem are summarised in Table 3.2.

Problems	**Supporting Example Quotations**
Inefficiency of information retrieval from the network.	• "Documents from FAI could be found on the net. But documents from AVIC and Xi'an aircraft manufacturing could not been found, though we are cooperators of this project." (Source: Project Engineering Project Team Leader). • "The retrieved information is not complete sometimes. Maybe missed when they arranged data." (Source: Aircraft Design Project Engineer, aerofoil). • "Documents could be unread because of authorisation problems." (Source: Aircraft Design Project Engineer). • "When you try to search information you want, you usually couldn't. Otherwise you know where it is." (Source: Aircraft Design Project Engineer, assembly). • "Ask related engineer if I couldn't get it from the net. Especially when I have no idea where to find the information from the net." (Source: Aircraft Design Project Engineer, coordination support).
Lack of knowledge sharing through the network.	• "They should update knowledge regularly. Otherwise I still need to ask other people. But data on the net is more rapid and accurate. So if I could get it on the net I won't turn to others." (Source: Aircraft Design Project Engineer). • "It satisfies our requirements. But not powerful. Why? You could get required information that you know you could get it on the net." (Source: Aircraft Design Project Engineer, structure). • "I am wondering if the information from our meetings and seminars could be put on the net. We will be happy to see that." (Source: Project Engineering Project Lead). • "We know the code very well. I get the document I required immediately. Maybe that is not enough. But I could go to the library." (Source: Aircraft Design Project Engineer, structure).

Table 3.2: Knowledge Management Problems and Supporting Quotations.

Two main knowledge management problems emerged from the investigation. Though other problems were exposed, like the identifying the knowledge assets, encouraging knowledge sharing, and so on, these two major ones should be focused on in this study. They are:

- Inefficiency in information retrieval from the network.
- Lack of knowledge sharing through the network.

Knowledge management problems identified from the investigation are discussed separately. The knowledge management within an enterprise is hindered by various barriers. These knowledge management problems are caused by factors which could be social, technological or the combination of both. People are reluctant to organisational change. In response to the question about company organisation, an interviewee opined that it is unlikely to change because "it is defined when it is established. Besides time, economy is a main constraint". Indeed, another respondent stated that the current organisation "facilitates knowledge processes best". As Wick (2000) suggested, the priority of organisational approach to knowledge management is growing and nurturing a knowledge-sharing culture. Besides organisational change, technological approach is worth considering. Information technologies facilitate disseminate and application of knowledge. Two key problems are extracted from the investigation, which are described below.

The first problem is inefficiency of information retrieval from the network. There is not an entirely shared drive. Each Department has one computer and fileshare. This kind of model aims for easy control and management. All data input from uniformed data interface. The advantage of this structure is that once the computer is affected by computer virus, physical isolation could control the problem.

Shared documents are divided into Public Folder, Department Folder, and Personal Folder. Users access different folders with different privileges. For each project there are special interfaces for related users. Regular data from FAI is packed and sent to the Archive Department of CACC every working day. The Archive Department then releases the data into CACC. These data are all structured. Formats of the shared documents are mainly Catia Drawings, Excel and Word files.

Users need multimedia or hypermedia documents to learn the work flow and operation methods. But these documents are unable to be retrieved from the network. Conference video, courseware, and product demo are in those formats as well. Besides multimedia and hypermedia documents, semi-structured and unstructured documents like email are not retrievable.

The second problem is the lack of knowledge sharing through the network. Users are not active in updating their knowledge to the network. In general, regular structured information can be updated to the network and is easy to get. This kind of data is classified and coded. File folders are separated as Airworthiness Folder, Design Folder, Model Folder, Manufacturing Folder, Product Supporting Folder, and so on. In the Design Folder, documents are divided into Directive Order Folder, Technique Coordination Folder, Technique Standard Folder, Test Folder, Engineering Order, and so on. Engineers could get information through path access.

If the user would like to know what components of the first flight are completely designed, he/she could get documents by path access. The first flight folder contains two sub-folders, displayed as System Folder and Structure Folder. Unfold the Structure Folder could get sub-folders like Door, Fuselage, Nacelle, Empennage, Window, and Wing. The Fuselage Folder includes sub-folder of Nose, Fore-body, Medium-body and Rare-body. But if the user searches the network, he/she will get several documents directly.

Almost all engineers are quite familiar with the document code. They get the exact document by searching its code directly. The document is coded according to information it contains. The information is about aircraft type, sequence, part number, document type (e.g. fabrication order and assembly order) and so on. The code is usually more than 10 digital numbers.

Disadvantage of searching by document code is that users do not update their knowledge to the network for knowledge sharing. One reason is that others do not get information through path access to others' folders. The search engine used retrieves official documents only. All these documents are dealt with and guarded by the Archive Department. Engineers do not think that providing knowledge to the network is their work. Personal knowledge is usually not so formal. It may be semi-structured or unstructured. Engineers with knowledge are unlikely to format it and update it to the official documents. There is no incentive to share knowledge. Engineers do not have the permission to publish information freely. When other people need the knowledge, they talk to them face to face, or by telephone, or by email. This kind of information is not kept either. Therefore, the knowledge lives with the person before and after he/she retires. It might be even more trouble if the searcher does not know who knows the knowledge he/she needs. An idea is that the searcher may discuss it on BBS (Bulletin Board System). But he/she could not get the answer immediately, or even worse, no answer at all.

The limitation (or boundary of the scope) of the investigation is that it emphasises on technological approach of knowledge management. The majority of its effort is concerned with knowledge sharing as knowledge distribution and retrieval, instead of knowledge creation.

3.7 Summary

This Chapter described the industrial investigation carried out in two sites of an aircraft design and development corporations, CACC and FAI, both located in Shanghai, China. CACC focuses on aircraft design integration, assembly, marketing, customer service, and airworthiness and certificate obtainment. FAI is focused on the research and development of aircraft including flight experiment. CACC is the cooperation partner of AVIC while AVIC is a major shareholder of CACC. They are responsible for the projects of regional aircraft ARJ21 and large passenger aircraft C919. It is the development of the passenger aircraft that will provide the industrial case study for this PhD project. Nine project teams are involved in the passenger aircraft development. Each project team has its main focus, and cooperates with other teams.

Through the industrial investigation, two key knowledge management problems have been identified. The first problem is the inefficiency in information retrieval/search from the network. The second problem is the lack of knowledge sharing through the network. The industrial investigation, together with literature survey, provided justification of the project aims and objectives, and a real-life case study for the evaluation and testing of the proposed methodology. The industrial investigation concluded that the proposed enterprise knowledge search framework could much improve the two problems identified and thus their performance in product design and development. Some other problems are found in the industrial investigation. In the future, the work for better enterprise knowledge management may include an enterprise-wide vocabulary to ensure that the knowledge is correctly understood, and creating a culture that encourages knowledge sharing.

CHAPTER FOUR: THE PROPOSED MULTI-AGENT KNOWLEDGE SEARCH METHODOLOGY

After the industrial problems and requirements for enterprise knowledge and information search are identified in Chapter 3, this Chapter introduces the concepts, principles and technologies directly related to the problems and requirements. The proposed enterprise knowledge search framework will be described in detail. The selection of software tools and environment for the development of the framework will also be described. Section 4.1 gives an overview of the various techniques used in current search engines, most of them are used for Internet search. Section 4.2 introduces the principles and applications of Agent (and Multi-Agent) technologies and Section 4.3 discusses the use of Agent technologies in information/knowledge search. The detailed description of the proposed methodology is provided in Section 4.4. Section 4.5 discusses the different types and formats of information/knowledge that will be searched and some examples will be given in later Chapters. Section 4.6 assesses available tools and gives the reason for the selection of the tools used to develop the methodology. Finally, Section 4.7 summarises the main emphasises and advantages of the proposed methodology over previous/existing search engines.

4.1 Techniques Used in Current Search Engines

There are a number of search techniques used in traditional information retrieval systems or search engines that are fairly standard and well understood, including Boolean search, truncation operations, wildcard search, term highlighting, database field search, proximity search, and relevance ranking (Notess 2006). Information retrieval systems typically incorporate a number of the common search techniques with additional capabilities, such as saved searches, that allow for easier and more effective searching. There are other types of functionalities which are fairly recent in inception, or have been available but only now play a central role in retrieval. For example, concept-based searching, fuzzy queries, scalability, query expansion and term weighting. Each of these expanded search functions are further discussed below.

4.1.1 Concept-Based Searching

The phrases "content-based retrieval" or "concept-based retrieval" are used as technical descriptions by many practitioners. Almost half of the practitioners offering information retrieval systems on UNIX-based machines support content-based retrieval in some form. The notion of content-based or concept-based retrieval can be informally described as a method of searching for information that is based on meaning. The implication is that the technique uses some notion of semantic abstraction that goes beyond searching for terms explicitly present in the text, i.e., the information cannot be found by a keyword search. This loose definition fits a number of searching methods and satisfies all the users' notions of concept searching. However, although there are some aspects of concept searching that are shared by some users, in general, the way in which concept-based retrieval is implemented varies across users.

Concept-based retrieval presupposes some sort of semantic representation that reflects relationships between terms. These terms and their representations can be general or domain specific, manually created or machine generated, and products differ in terms of the types of representations and facilities for constructing the relationships. Verity, for instance, identifies concept-based searching with queries using topic trees that are manually constructed for a particular domain, whereas HNC (HNC software, Inc) (Ilgen and Rushall 1996) uses context vector associations automatically derived from text as a basis for conceptual searching. IDI's (Information Dimensions, Inc) (Solskinnsbakk and Gulla 2008) notion of concept searching is founded on term expansion using an application-specific thesaurus that includes the notion of concept hierarchies, where the levels represent the degree of specificity of terms. ConQuest's (ConQuest Software, Inc) (Nelson 1994) major emphasis is on its dictionary, which contains word meanings and semantic relationships.

Although it can be extended for different domains, the semantic network that is delivered with each of ConQuest's products is for general purpose. Verity topics and IDI's concept hierarchies must be developed by the customer for each application.

4.1.2 Fuzzy Searching

Fuzzy searching is another term that is used extensively in marketing and technology specifications of information retrieval users. Moreover, the use of the technique is not restricted to the high-end systems. Many PC- and Macintosh-based systems can perform fuzzy searching as well.

Fuzzy searching can be loosely described as a search technique that allows for term matches that are not exact. In practice, fuzzy searching has a variety of behaviours. For instance, in Concordance by Dataflight (Dataflight 2010), fuzzy queries allow phonetic variants (e.g., "jim" and "gym") and orthographic variants (e.g., "US" and "U.S."). Excalibur (Excalibur 2010) supports letter substitution, transposition, and a form of stemming, while Odyssey's ISYS (Odyssey 2010) only supports letter substitution. Even more of a contrast is Fulcrum (Fulcrum 2000), which permits fuzzy Boolean searches, a relaxation technique primarily used for the Boolean operator AND, and having nothing to do with phonetic or orthographic variants.

4.1.3 Scalability

Most users claim that their technology is capable of indexing and searching large document collections, that is, their technology scales. In fact, many users identify the ability to scale as one distinguishing feature of their technology. Since the amount of on-line information is growing rapidly, scalability is certainly important and will likely be a necessary condition for an information retrieval user to survive. Here are some examples of sizes of document collections that can be indexed and searched.

CLARIT from CLARITECH (CLARITECH 2010) can index up to 8GB of source text, while BASIS plus (IDI) (BASIS 2010) can handle a collection as large as 126GB. Moreover, since users can search up to 32 collections in a single session in BASISplus users, in principle, can search 4 Terabytes of data. ConQuest has been indexing and searching collections up to 100GB. Recently, ConQuest has been testing its system with 500GB collections. Open Text claimed that there are no inherent size limitations on document collections for indexing with their system. Open Text has indexed 50GB collections. TextWare, a company with a personal computer product, can support 2GB of text in a single database.

Speed also plays a factor in scalability. Large collections must be indexed in reasonable periods of time. CLARIT can index 80-100 MB per hour on a DEC Alpha 3000/400 (133Mhz and 128MB RAM). On mid-range UNIX servers, BASISplus has an indexing throughput of up to 120MB per hour, while TextWare can index up to 32MB per hour on a 80-386 PC.

Even with the divergence in the size of collections that can be indexed and searched and the divergence in indexing time, it is clear that today's technology is capable of handling tens of gigabytes of text with a processing speed on the order of tens of megabytes per hour.

4.1.4 Query Expansion (Stemming and Thesaurus-based)

Query expansion refers to the process whereby a query is augmented by terms which are in some way related to the terms in the original query. Two methods that are commonly used are word-based stemming or morphologically and thesaurus-based. In a system which supports stemming or morphology, inflected terms can be generated and used for uninflected terms in a query. For instance, a query with the simple term "tree", which is an uninflected form, would be broadened so as to include a search for the plural form "trees", which is an inflected form. A similar process can generate singular forms from plurals as well as third person singular, past tense, and participial forms of verbs. Such a system takes the burden off the searcher in remembering to use inflected forms or in formulating more complex forms containing inflected and uninflected forms.

Most systems do not use morphological processing or stemming, but give the user an option of using a wildcard. Proper use of wildcards can have approximately the same effect as stemming. However, wildcard search places a demand on a searcher to introduce a wildcard in a term in a way that will not be too restrictive, as documents will be missed, or too unrestrictive, as irrelevant documents will be returned.

Another form of term expansion exploits a thesaurus. This method differs significantly from stemming and other expansion processes in that expansion is accomplished via semantic relationships rather than word structure operations. The process of semantic expansion involves a collection of words or phrases that are linked through a set of relationships including synonymy ("happy" and "glad"), antonymy ("happy" and "sad"), and is-a ("birch" is a "tree"). Thus, a search for "happy" would also yield documents with the term "glad." Users can invoke any of the query expansion relationships (e.g., synonymy and antonymy) while doing a search in order to broaden searches as appropriate.

Thesaurus construction is a labour-intensive task and it is not surprising that few information retrieval users offer a large thesaurus with their products. However, given the potential utility of thesaurus-based term expansion, users often support this type of expansion with user-supplied thesauri with the proper format. Cuadra (Cuadra 2010) is one such user that does not supply a thesaurus but incorporates query expansion functionality in its products. ConQuest is one user that offers a large set of semantic relationships with its products for supporting term expansion. Currently, ConQuest network has over 400,000 concepts (terms) and 1.6 million word relationships. Dataware (Dataware Technologies, Inc.), which has licensed and adapted technology from Inso (Inso Co.), also includes thesaurus support as well as a semantic library API (Application Programming Interface). It is worthwhile to note that Inso's primary focus is to develop multi-lingual tools for supporting morphological processing and thesaurus functionality using a variety of techniques.

4.1.5 Term Weighting

In a query, usually each term is equal in value to any other in that one term match does not impact the scoring more than another during a search. However, certain systems allow users to adjust weights, that is, assign preferences to particular terms in a query. By doing so, the user will influence the relevance ranking of hits, and documents will be prioritized differently than the case where all terms were of the same weight. CLARIT, Excalibur, Fulcrum, and Verity all support term weighting.

4.1.6 Summary of Current Search Techniques

The search functions described above, i.e., concept-based searching, fuzzy searching, scalability, query expansion, and term weighting, are playing central roles in current information retrieval applications. However, they are mainly suitable for static and inefficient in real life enterprise applications. The proposed methodology will use Agent technology, to be introduced next, to locate enterprise-wide information, to gather new, time-critical enterprise information, and to support team-building and communication in enterprises' Intranets. This will be a big step forward from the above traditional search techniques.

4.2 Agent Technologies

There are several definitions of Agents in the literature. One can describe rather than define Agents in terms of their task, autonomy, and communication capabilities. Some main definitions and descriptions of Agents are given below.

Agents are semi-autonomous computer programs that intelligently assist the user with computer applications. Agents employ artificial intelligence techniques to assist users with daily computer tasks, such as reading electronic mails, maintaining a calendar, and filing information. Agents learn through example-based reasoning and are able to improve their performance over time (Klusch 2001).

Agents are computational systems that inhabit some complex, dynamic environment. They sense and act autonomously in this environment. By doing so, they realise a set of goals or tasks (Franklin and Graesser 1997).

Agents are software robots. They can think and will act on behalf of a user to carry out tasks. Agents will help meet the growing need for more functional, flexible, and personal computing and telecommunications systems. Applications of intelligent Agents include self-contained tasks, operating semi-autonomously, and communication between the user and systems resources (Jansen 1997).

The definition and description of an Agent for this research is that Agents are software programs that implement user delegation. Agents manage complexity, support user mobility, and lower the entry level for new users.

Two issues concerning Agents are trust and competence. Concerning trust, the user and other members of the user community must be able to trust that the Agent does only what the user wants done. The user must feel comfortable delegating tasks to the Agent. As for competence, the Agent must first acquire the skills to accomplish the delegated tasks. The Agent must also be able to decide when to help the user and how to help the user.

There are three major paradigms for building Agents. The first approach makes the Agent an integrated part of the end-program. The advantage here is that the user trusts the Agent because the rules are set within the end-program of the user. The problem is with competence (in terms of user and also the Agent itself), because a combined Agent and end-program requires too much insight from the user. The user must have the knowledge to effectively employ the Agent. The second approach is a knowledge-based approach, where the Agent has extensive domain-specific information and knowledge about the application. Therefore, the advantage of this approach is competence. Whilst trust is a problem since the Agent is usually autonomous from the start, which gives users a feeling of loss of control and lack of understanding. The third approach is a learning approach, where the Agent has some knowledge of the domain but learns what the user would like it to do base on user actions. The learning approach has the advantages of the other two approaches while minimizing their disadvantages. The "learning" approach is the architectural paradigm that this research will use. Learning means that the system evolves behaviours based on empirical data.

Generally, an Intelligent Agent could not exist in the hardware and software environment alone. It has to cooperate with other intelligent Agents in the environment. On the other hand, each Agent is active and operates autonomously. Therefore in real applications Multi-Agent systems are commonly used. A Multi-Agent System is composed of several Agents, collectively capable of reaching goals that are difficult to achieve by a single Agent or a monolithic system. MAS could be claimed to include human Agents as well. In fact, all Agents are under active human supervision in practice. Furthermore, the more important the activities of Agents to human is, the more supervision that they receive.

Human organisations and society in general could be considered as examples of Multi-Agent systems.

Considering a Multi-Agent manufacturing system, there are basically three different approaches of Agent organisation: hierarchical approach, federation approach, and autonomous approach (Shen, Hao et al. 2006). The hierarchical approach has a centralized appearance, which is criticised. The federation approach is able to coordinate Multi-Agent activities through facilitation. Therefore, it is a good foundation upon which to develop open, scalable Multi-Agent system architectures. And it is quite suitable for developing distributed manufacturing systems which are complex, dynamic, and composed of a large number of resource Agents. For autonomous approach, Agents interact directly with any other Agents in the system and also with other external systems. Moreover, Agents have their own goals and associated set of motivations. Therefore, this approach is demonstrated to be well suited for developing distributed intelligent design systems where existing engineering tools are encapsulated as Agents and connected to the system for providing special services. A hybrid approach is a combination of the above approaches. It is employed for developing more flexible, modular, scalable and dynamic manufacturing systems.

One typical Agent and its relationship with other Agents in a Multi-Agent System can be illustrated as Figure 4.1. It contains an Internal Database, an Internal Message Transport Module, an External Message Transport Module, a Process Execution Machine, and Multi functional Modules. The Process Execution Machine contains decision method and procedures used to control the Functional Modules to perform actions and tasks, making use of the Internal Database and external input (through the Communication with other Agents as shown in the bottom of Figure 4.1). Each Agent communicates with other Agents and the environment by the External Message Transport Module. The Internal Message Transport Module supports communication among Internal Functional Modules. Multi-Agent Systems can manifest self-organisation and complex behaviours even when the individual strategies of all their Agents are simple.

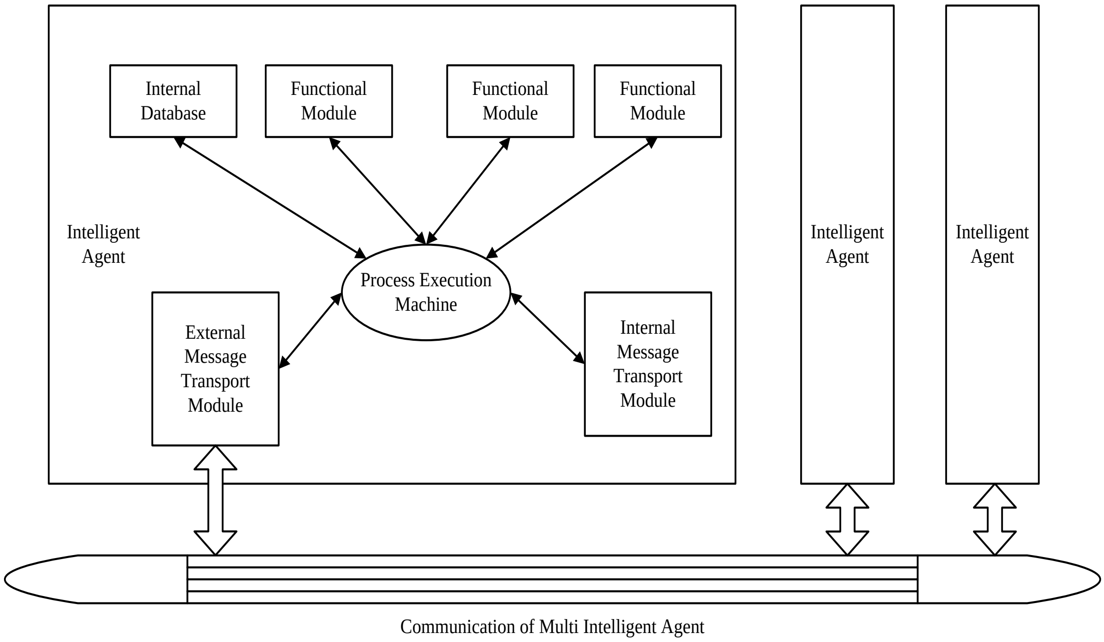


Figure 4.1: Basic Elements of a Multi Agent System

Agents could be organised to enhance collaboration into two very different approaches. The first approach is direct communication, in which Agents handle their own coordination. The other is assisted coordination, in which Agents rely on special system programs to achieve coordination.

The advantage of direct communication is that it does not rely on the existence, capabilities, or biases of any other programs. Two popular architectures for direct communication are the contract-net approach and specification sharing.

In the contract net approach to interoperation, Agents in need of services distribute requests for proposals to other Agents. The recipients of these messages evaluate those requests and submit bids to the originating Agents. The originators use these bids to decide which Agents to task and then award contracts to those Agents.

In the specification sharing approach to interoperation, an Agent supplies other Agents with information about its capabilities and needs; other Agents can then use this information to coordinate their activities. The specification sharing approach is often more efficient than the contract net approach because it decreases the amount of communication that must take place.

One disadvantage of direct communication is cost. So long as the number of Agents is small, e.g. two to three Agents, this is not a problem. But, in a setting like the Internet, with millions of programs, the cost of broadcasting bids or specifications and the consequent processing of those messages is prohibitive. In this case, the alternative is to organise the Agents in some way that avoids such broadcasts.

Another disadvantage is the complexity of implementation. In the direct communication schemes, each Agent is responsible for negotiating with other Agents and must contain all of the code necessary to support this negotiation. If all these capabilities can be provided by the system, this would lessen the complexity of application programs. In practice this is not the case.

A popular alternative to direct communication is the assisted approach that eliminates both of these disadvantages. Agents are organised into what is often called a federated system. Each system has its own machine and several Agents (each Agent has its own machine as well). Agents do not communicate directly with each other. Instead, they communicate only with system programs called facilitators, and facilitators communicate with each other. In a federated system, Agents use ACL (Agent Communication Language), in practice, a restricted subset of ACL, to document their needs and abilities for their local facilitators. In addition to this meta-level information, they also send application-level information and requests to their facilitators, and accept application-level information and requests in return. Facilitators use the documentation provided by these Agents to transform these application-level messages and route them to the appropriate places. In effect, the Agents form a “federation” in which they surrender their autonomy to their facilitators and the facilitators take the responsibility for fulfilling their needs. The proposed methodology adopts the assisted approach and has been implemented as a federated system.

System services in support of software interoperation are provided in Multi-Agent Systems. For example, directory assistance programs facilitate software interoperation by providing a way for programs to discover which programs can handle which requests and which programs are interested in which pieces of information. Distributed object managers, like CORBA (Common Object Request Broker Architecture), OLE (Object Linking and Embedding), DSOM (Distributed System Object Model), provide location transparency for object-oriented systems, routing messages to objects without requiring senders to know the locations of those objects. The primary difference between these approaches to software interoperation and Agent-Based software engineering lies in the sophistication of the processing done by facilitators.

Using Agent Communication Language, Agents can express their needs and capabilities more accurately than using pattern-based meta-languages, and facilitators can use this added needs and capabilities information to be more discriminating in routing messages. In order to deal with notational incompatibilities, facilitators can translate messages from one vocabulary to another using definitions supplied by Agents or retrieved from the ACL dictionary. In so doing, they can decompose messages into sub-messages and send them to different Agents. When necessary, they can combine multiple messages. In some cases, this assistance can be rendered interpretively (with messages going through the facilitators); in other cases, it can be done in one-shot fashion (with the facilitators setting up specialized links between individual Agents and then stepping out).

In order to provide these capabilities, current implementations of facilitators take advantage of automated reasoning technology developed in the Artificial Intelligence and Database communities. Powerful search control techniques are used to enhance normal message-processing performance; and automatic generation of message routing programs and pairwise translators is used for cases requiring greater efficiency.

From the above description and discussions, it can be assumed that traditional search engines could employ Agent technology to achieve active and autonomous search operations. And the physical width of the network does not restrict the search operations as previous methods do.

4.3 Agent Technology for Information/Knowledge Search

A number of researchers have explored the use of Agents for information retrieval, filtering, cataloguing, and delegation. Information filtering is similar to information retrieval. In information retrieval, one views the user actively searching for relevant information in a mass of largely irrelevant information. With information filtering, one views the user as largely passive as most relevant information flows past the user.

For example, Amalthaea (Moukas and Maes 1998) is a Multi-Agent ecosystem for personalised filtering, discovery and monitoring of information sites. Its main goal is to assist the users in finding interesting information on the Web. There are two types of Agents in Amalthaea: filtering Agents that model and monitor the interests of the user and discovery Agents that model the information sources. Both the users' interest and retrieved documents from Web sites are represented by weighted keyword vectors. The information Agents pick one document from the downloaded set passed by the discovery Agents and calculate how a confidence level is that specific document will satisfy the user's needs. The confidence measure is not different from the typical normalized similarity measure (cosine) used in the vector space model in information retrieval. A particular feature of Amalthaea is that it provides a market-like ecosystem in which Agents evolve, collaborate and compete to survive. Agents that are valuable (useful) to the users and to other Agents are allowed to reproduce while low-performing Agents are destroyed to save system resources. Though its Multi-Agent architecture was not explicitly specified, Amalthaea appeared to be a kind of specification sharing systems.

The application of intelligent Agents to the design of information retrieval/search systems has drawn some attention in recent years and their benefits have been demonstrated in several publications.

4.4 The Proposed Methodology

Agent technology provides a new method for information search through the Internet, with the advantages of dynamic execution, asynchronous computing, and parallel problem solving. However, both traditional and Agent-based search engines discussed so far are for general information retrieval in the public domain through the Internet. Search for enterprise applications differs from internet search in many ways. Enterprise search includes: search of the organisation's external website, search of the organisation's internal websites, and search of other electronic text held by the organisation in the form of emails, database records, documents on file shares and the like. The link structure of enterprise repositories is quite different from the one on the Internet. The lack of highly hyperlinked nature hinders the use of some techniques based on link analysis. The application of these techniques is not as effective on enterprise search as on the web. The low relevance of retrieved information requires the development of new techniques to improve information especially knowledge search. Another reason is that in the company, there are additional security, reliability, and company policy, which even complicate the problem. The privacy of individuals' personal data is protected and cannot be obtained without authorization.

To reduce communication cost as well as system workload, the proposed system will utilize a federated system consisting of multiple interacting Agents of three types, namely Dispatch Agent, Search Agent, and Learning Agent. To be clear on the description, the system is separated into two parts, i.e., searching information/knowledge through an enterprise's Host Computer directly (Part 1 of Figure 4.2) and searching information/knowledge through the Search Server (Part 2 of Figure 4.2). The Search Server is an enterprise search platform. It shares its architectural underpinnings with the Windows Search platform for both the querying engine as well as the indexer. The Host Computer is a computer connected to the Internet. It hosts information resources as well as application software for providing network services. Figure 4.2 shows the structure of the overall proposed system with Part 1 at the right hand side and Part 2 at the bottom left (the top left is the users of the product development functions).

Users, in this project the product development team members, communicate with the Dispatch Agents through TCP/IP (Transmission Control Protocol/Internet Protocol). The TCP/IP protocol suite coordinates all kinds of computers to communicate with each other. Once the Dispatch Agents get search requests from the users, they dispatch the queries to the Search Agents and the Learning Agents. In Part 1 of Figure 4.2, the Search Agent obtains access to the relevant enterprise information and do search. The information is stored in shared databases, DSS (Decision Support System), CAD (Computer Aid Design), Expert System, Knowledge Base, Data Warehouse, and MIS (Management Information System). In addition, the Search Agent is responsible for reporting the search results to the Dispatch Agent. The Learning Agent is assigned to consider the user interest model.

The Agent Server is a virtual machine on which Agents execute. Agents connected with the Agent Server could obtain references to Java objects that are associated with the server and available in the Server. The Agent Server that provides services to connected Agents could export references to Java objects associated with the services. While the Host Computer with the Agent Server in the Internet is not so common, providing a uniform Agent Server is not so difficult in an enterprise. The application of Agents will be more optimistic in the future along with the development of the Agent technology.

The System platform describes the hardware architecture and software framework of users' computer that allows software to run. It normally includes a computer's architecture, operating system, programming languages and so on.

In Part 2 of Figure 4.2, the Dispatch Agent, the Learning Agent and the Agent Server work the same way as in Part 1. The difference is that the Search Agent does not itself mobile to network to search. It gathers search results from other search engines like Alta Vista search engine, WebCrawler search engine, Excite search engine, and so on. It filters the search results and sends them back to the Dispatch Agent. The Dispatch Agent provides search results as a Web page to the user.

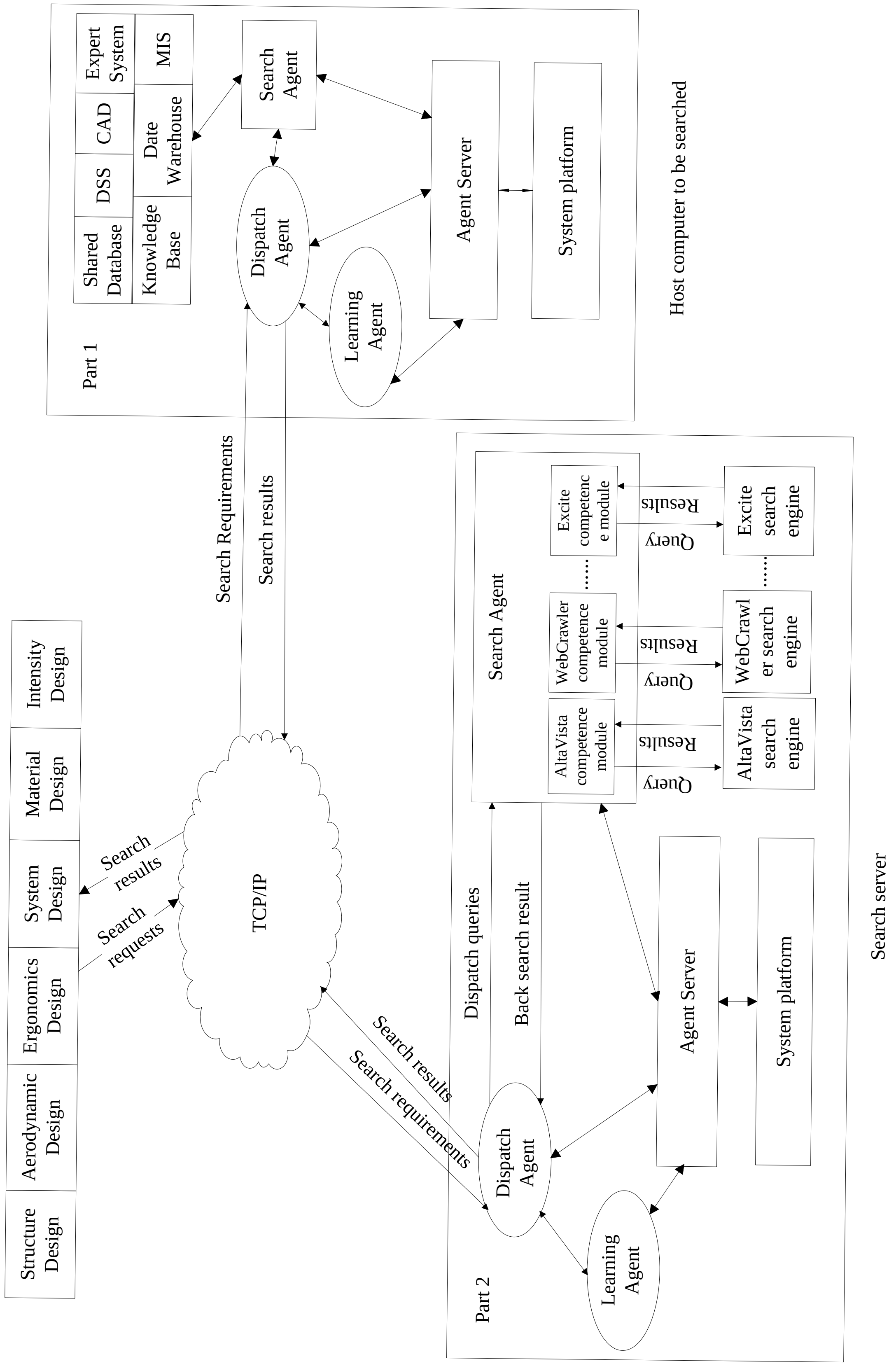


Figure 4.2: the Proposed Methodology

The whole system is based on TCP/IP, which could be separated into four layers as Figure 4.3 displays: the application layer, the Agent layer, the platform layer, and network layer.

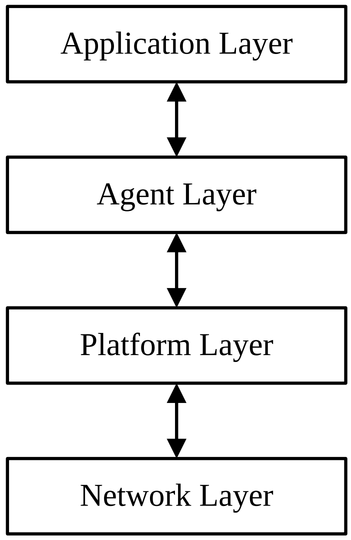


Figure 4.3: Four Layers of the System

The Application Layer is the top one, interacting with users. It has two functions. One is to accept user search requests and related parameters. The other is to provide search results as TXT (text) files or HTML (Hyper Text Markup Language) files, according to user requests. The Application Layer passes user requests and parameters to the Agent Layer as it needs.

The Agent Layer is the core and also the most complex part of the system. It possesses Agent object corresponding to user requests. It transfers the Agents to the Platform Layer. At the same time, it manages all local Agent objects, recording their current positions and states. Information returned by Agents are extracted and reorganised, in the format requested by users, and submitted it to the Application Layer.

The Platform Layer is the transmission pipe of the whole system. It provides existing environment for Agents, which supports Agent mobile, Agent communication, and connection between Agents and the user's computer.

The Network Layer is based on an enterprise's existing network communication protocol. It transmits information through protocol ports of peer Layers. It receives anomaly information from under layer and decides the delivery, acceptance, and storage of the information.

In summary, the proposed framework provides facilities for searching information and knowledge related to product development teams from both within and outside an enterprise. Tests will be given in Chapter six to illustrate how the system works with the collaborating company.

4.5 Types and Formats of Knowledge/Information

From the enterprise characteristics view point, content sources and formats are diversified. Enterprises must ingest and extract structured and unstructured information/knowledge from heterogeneous content sources, such as Microsoft Exchange, Lotus Notes, Documentum, as well as file systems and databases. Furthermore, documents exist in a myriad of file formats and several languages. A single document could contain multiple languages, or attachments with multiple MIME-types (Multipurpose Internet Mail Extension protocol). As at present, less than 10 percent of enterprise content by volume has been reported to be HTML (Mukherjee and Mao 2004). An individual's role in an enterprise dictates what documents can be accessed. Sophisticated enterprises demand a more stringent notion of security in which search result lists are filtered to display only the documents accessible to the user. Doing this in conjunction with the native security of the repositories is a particularly difficult challenge.

Information/knowledge that is considered to be unstructured is in fact semi-structured, with metadata such as author, title, date, size, and so forth. Conversely, much structured information in a Relational Database Management System is unstructured—for example, blobs of text and VARCHAR fields. Extensible Markup Language is ubiquitous in content and applications. It is essential to provide high-performance parametric search that allows the user to navigate information through a flexible combination of structured and unstructured data.

No single scoring and ranking function will work for all enterprise search contexts. Many of the powerful link-based scoring and ranking algorithms that have been honed for the Web are unlikely to be germane to the enterprise.

Enterprise content is fundamentally different from Web content, enterprise users are different in their goals and expectations from Web users, and enterprise search imposes layers of complexity that the Web doesn't have.

While the Internet tends to grow democratically, intranets are often governed by bureaucracies. Content creation on an intranet is normally centralized to a small number of people. While content published on an intranet may need to comply with specific policies (reviews, approvals), consistency is not guaranteed, since there may be multiple organisational units whose policies differ.

It is well known that some of the most valuable knowledge in an enterprise resides in the minds of its employees. Enterprises must combine digital information with the knowledge and experience of employees. An important distinction between the enterprise search and the Internet is that while Internet users are anonymous for the most part, enterprise search users are answerable and guided by specific controllable processes. Privacy issues are also very different in an enterprise, since people are usually engaged in enterprise-specific behaviour and are being compensated for their engagement.

Examples of the various information/knowledge types and formats that exist in the collaborating company will be provided in later Chapters to demonstrate the capability of the proposed framework in its current level of implementation. The framework has the potential to search for most types and formats of information/knowledge in real life engineering applications.

4.6 Selection of Potential Tools to Implement the Methodology

There are many powerful, freely available software platforms for implementing sophisticated Agent systems. Wide use of Agent technology depends on the availability of development tools and platforms. Such tools and platforms, in turn, presume the existence of standards that reflect the agreement of developers on which basic functionality should be and how it should be presented.

The development environments for enterprise search and for Web search are different. The hardware and software of a Web search engine are fully controlled and managed by an organisation as a service. However, the software systems of enterprise search are licensed to and deployed by many organisations in diverse environments. Therefore requirements are varied including hardware constraints, software platforms, bandwidth, firewalls, heterogeneous content repositories, security models, document formats, user communities, interfaces, and geographic distribution (Mukherjee and Mao, 2004). Many research and commercial organisations are involved in the realization of Agent applications. An overview of some Agent construction tools are given below.

University of Toronto (2002) developed an Agent Building Shell – Programming Cooperative Enterprise Agents, that provided several reusable layers of languages and services for building Agent systems including coordination and communication languages, description logic-based knowledge management, cooperative information distribution, organisation modelling and conflict management. One example of the application of this approach is to develop Multi-Agent applications in the area of manufacturing enterprise supply chain integration.

The Gypsy Project (Jazayeri and Luqmayr, 2000) utilized Java for the implementation of a flexible environment for experimenting with mobile Agent programming. It was intended for applications in Internet information retrieval, Internet commerce, mobile computing and network management. Chauhan and Baker (1998) provided a framework to guide the development of Multi-Agent systems along with a set of classes for Agent deployment in Java. The framework is intended to help beginning and expert developers structure their ideas into concrete Agent applications. Their project provided a good comparison of Agent tools with a particular emphasis on mobile Agent projects.

The Adaptive Retrieval Agents Choosing Heuristic Neighborhoods for Information Discovery reported by Menczer and Degeratu (2000) featured an artificial life inspired model using endogenous fitness for information retrieval in large, dynamic, distributed, heterogeneous databases, such as the WWW. A population of Agents is evolved under density dependent selection for the task of locating information for the user. The energy necessary for survival is obtained from both the environment and users in exchange for relevant information. By competing for energy, the Agents robustly adapt to their environment and are allocated to efficiently exploit their shared resources.

Multi-Agent Modeling Language (MAML) is a programming language for building Agent-based models. The MAML modelling environment is aimed to support scientists, who are not experts in programming, to do research in computer-assisted modelling. Such a modelling environment will enable and encourage experimentation and exploration, providing freedom in all phases of modelling. The supported modelling discipline is Agent-based modelling. MAML is used to develop models, run simulations, search in parameter space, and analyze results. It has a graphical user interface, support work through the Web, and be based on an adequate programming language.

Java Agent Development Environment (JADE) (Telecom Italia Lab, 2000) is a software framework in compliance with the FIPA (Foundation for Intelligent Physical Agents) specifications for interoperable intelligent Multi-Agent systems. It includes all the mandatory components that manage the platform, that is the ACC (Agent Communication Chanel), the AMS (Agent Management System), and the default DF (Directory Facilitator). All Agent communication is performed through message passing, where the representing language is FIPA ACL.

MultiAgent Systems Tool (MAST) (Vrba, 2003) is a general purpose distributed framework for the cooperation of multiple heterogeneous Agents. The MAST architecture consists of two basic entities: the Agents and the network through which they interact.

Open Agent Architecture (OAA) is a framework for integrating a community of heterogeneous software Agents in a distributed environment. Its architecture allows a number of languages, namely C, Java, Prolog, Lisp, Visual Basic and Delphi, to create Agent systems. In this framework, an Agent is defined as a software process that registers its services in an acceptable form, speaks the Inter-Agent Communication Language (ICL), and shares functionality common to all OAA Agents, e.g., the ability to install triggers, manage data in certain ways (Cheyer and Martin, 2001).

The Social Interaction Framework (SIF) (Schillo, 1999) is a tool for rapid-prototyping of simulations involving multiple Agents. It supports the design by providing ready-to-use components and several debugging functions during run-time. With this software you can concentrate on Agent-oriented design and the study of the interactions between Agents. SIF takes care of issues such as parallel computing and visualization.

ZEUS (Nwana and Ndumu, 1998) is developed by British Telecommunications Labs. It is a 'collaborative' Agent building environment and component library written in Java. Therefore it allows the rapid development of Java Agent systems by providing a library of Agent components, supporting a visual environment for capturing user specifications. Each ZEUS Agent consists of a API layer, a definition layer, an organisational layer, a co-ordination layer, and a communication layer. The API layer allows the interaction with outer world. The definition layer represents the Agent's reasoning and learning abilities, its goals, resources, skills, beliefs and preferences. The organisation layer describes the Agent's relationships with other Agents. The co-ordination layer describes the co-ordination and negotiation techniques the Agent possesses. Communication protocols are built on top of the co-ordination layer and implement inter-Agent communication.

Although many Agent platforms have been available, most projects or research works still use traditional programming languages, such as C++, Java, Lisp, SmallTalk, Prolog, and Objective C to develop Agent-based systems. Other recently developed Agent systems use Jini network technology (Jini 2010). Most recently developed Agent-based manufacturing systems are Java-based systems. Some of them utilise existing Agent development tools for fast prototyping and standards compliance reasons.

Even though hundreds of open source or commercial Agent development tools have been available, this research is still developing proprietary Agent framework because most industrial companies cannot accept GPL (General Public License) or LGPL (Lesser General Public License) licenses of open source tools, and commercial tools are not cost effective enough for industrial application developments yet. Agent development tools are defined as a set of utilities for guiding a programmer through the steps required for defining and adding new Agents to the Open Agent Architecture (Cheyer, 2001).

Comparing current construction tools, Lucene is a good choice for this project. Lucene is an Open Source project which could be downloaded from its official website. It is written entirely in Java. It is recognised as a high-performance, full-featured text search engine library. It is not a complete search engine, but a framework of full-featured text search. The goal of using Lucene is to simplify the development of applications while ensuring standard compliance through a comprehensive set of system services and Agents. A full-featured text search environment is required to use Lucene. However, it has already significantly lowered the level of programming knowledge necessary. The core of Lucene's logical architecture is the idea of a document containing fields of text. This flexibility allows Lucene's API to be independent of the file format. Text from PDF, HTML, Microsoft Word, and OpenDocument documents, as well as many others, and all be indexed as long as their textual information can be extracted. Lucene does not have a class as index. Therefore, Eclipse is used to develop index. Eclipse is a multi-language software development environment. It is written mostly in Java and could be used to develop applications in Java and, by means of various plug-ins, other programming languages like C, C++, COBOL, etc. Eclipse is free and open source software which could be downloaded from its official website as well.

Eclipse does not contain Java Development Kit (JDK). JDK is available from Sun official site. Version 7 is used in this project to achieve the development environment.

4.7 Summary

Traditional information management systems in the manufacturing industry such as PLM and EPR systems manage structured information and knowledge, and search for required information and knowledge. However, the process is passive (i.e., only search information at requested), and has to go through well defined structures. In real life applications, such as the product development process, a lot of information is unstructured or not kept in structured systems. Active search without going through structured systems or search for information and knowledge that is not kept in structured systems is in greater demand in today's competitive business environment. This Chapter explained the principles and constituent functional elements of the proposed information/knowledge search methodology. The methodology uses Multi-Agent technology, and focuses on enterprise search requirements, whilst most previous search engines are for general information search in the public domain through the Internet. This methodology is proposed based on in-depth industrial investigation with a large scale aerospace manufacturing company and focuses on the specific problems and requirements in the new product development process, which is the most important and core business of OEMs.

CHAPTER FIVE: DEVELOPMENT OF THE MULTI-AGENT KNOWLEDGE SEARCH SYSTEM

The main objective of this research project is to develop a methodology for facilitating knowledge retrieval in enterprise. In Chapter 4, it was proposed that Multi-Agent technology could be used to achieve this end. It is worth emphasize that when 'knowledge retrieval' is referred to in the context of general knowledge and information management such as in media and public services, the requirements of 'knowledge retrieval or search' are quite different in the context of knowledge management in the context of engineering enterprise.

The knowledge search system is developed under the consideration of these aspects:

(i) An Agent building and editing tool to construct and maintain the Agents;
(ii) A mechanism to control and integrate Multi-Agents, including Agents communication, interaction and cooperation; and
(iii) A mechanism to disseminate the results to users (designers in this project).

This chapter presents how the Multi-Agent Knowledge Search System is developed, including the considering of the above three aspects.

5.1 The Industrial Context

As described in depth in Chapter 3, the product development process of an aircraft manufacturer is used as the industrial context for the system development and also for potential application of the prototype enterprise knowledge search system. In this company, there are a number of general stages in the design, development, manufacture and other operations on an aircraft, each with associated design methods and data requirements. Figure 5.1 shows the steps of a commercial wide-body aircraft design and manufacturing process. The process starts from the Preliminary Design which is based on common aircraft technology and requirements obtained from the Sales and Marketing Department. Then Design and Development follows the Preliminary Design. Design knowledge, system knowledge and knowledge about certification support this task. The next step is Airframe Manufacturing which is supported by manufacturing knowledge. This is followed by the Systems Integration task in which all systems are assembled and integrated in the airframe. The aircraft is then verified with test factors considered, and finally certification is issued.

From Figure 5.1, it can be seen that the certification information is shared between the Product Verification and the Design and Development tasks. The Sales and Marketing task communicates and shares information with the Preliminary Design task and the Design and Development task. The Customer Service task also communicates and shares information with the Design and Development task.

In summary, starting from the requirements and information about available technologies, the product development process consists of tasks such as conceptual design, preliminary design, detailed design, manufacturing, and testing, which are concurrent and overlapping. From the early stage of conceptual design, engineers need knowledge about design ideas, aerodynamics, structure and materials, propulsion, systems and equipment, weights, performance, stability and control, operating economics, and noise. Very general knowledge is first used which are then refined by progressively more accurate knowledge as the design evolves.

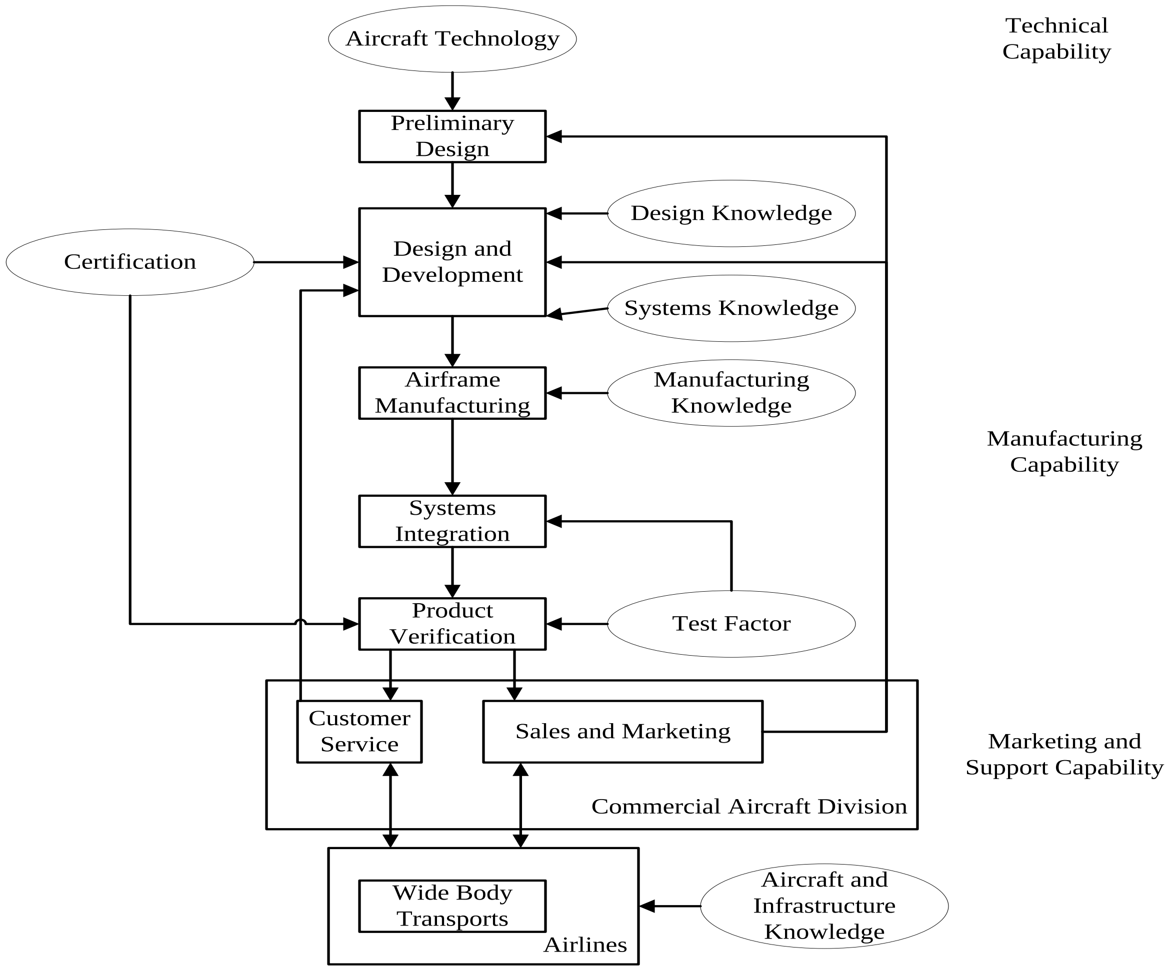


Figure 5.1: Steps of a Commercial Wide-body Aircraft Design and Manufacturing Process (Source: CACC)

The interior of an aircraft is restricted by its external shape. However, it is sometimes necessary to modify the external shape to accommodate the interior, with consequent aerodynamic changes. For successful operation, it requires a proper structure, propulsion, fuel, flying controls, avionics, furnishing, and so on. These all work together efficiently even when they have different functions. Each system is important, but each may have conflicting requirements. The aircraft designers must consider all of them. Taking the design of wing as an example, its structure should be designed to meet a number of requirements like, low weight, acceptable material and manufacturing costs. Meanwhile, the wing requires adequate strength to meet the maximum expected loads, with a suitable safety factor, adequate stiffness, good in-service properties such as fatigue and corrosion resistance together with tolerance of expected temperatures and other atmospheric conditions.

In general terms, flight of aircraft is simply the result of fluid flow about an aerofoil, or say, the result of moving a wing through the air. From a structural viewpoint, wing configurations could be classified as cantilever wing and braced (or strutted) wing. In terms of wing/fuselage arrangement, wings can be classified as high wing, mid wing, and low wing. From a sweep angle viewpoint, wings could be classified in the following manners: zero or negligible sweep, aft sweep (also called positive sweep), forward sweep (also called negative sweep), variable sweep (meaning symmetrically variable sweep), oblique sweep (meaning asymmetrically variable sweep).

An aerofoil is structure designed to obtain reaction upon its surface from the air through which it moves. The aerofoil is designed to produce lift. For a typical aerofoil profile, there are several obvious characteristics of design. Figure 5.2 shows the typical aerofoil section. It displays that there is a difference in the curvatures (camber) of the upper and lower surfaces of the aerofoil. The camber of upper surface is more pronounced than the lower one. And the two extremities of the aerofoil profile also differ in appearance. The leading edge is rounded while the trailing edge is narrow and tapered. A reference line drawn through the profile connecting the extremities of the leading and trailing edge is the chord line. The distance from this chord line to the upper and lower surfaces of the wing denotes the magnitude of the upper and lower camber at any point. Another reference line, drawn from the leading edge to the trailing edge, is the mean camber line. This mean line is equidistant at all points from the upper and lower contours.

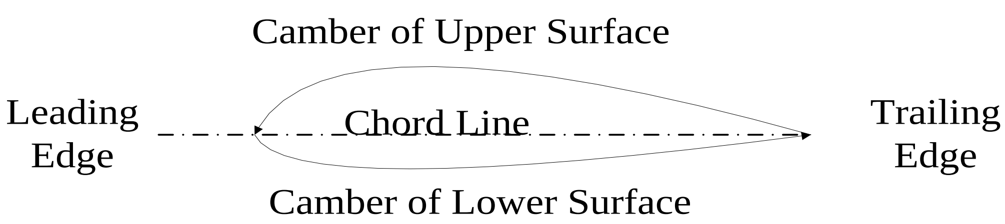


Figure 5.2: Typical Aerofoil Section (Source: FAI)

Design of the construction of the wing, so as to provide actions greater than its weight, is shaping the wing so that advantage could be taken of the air’s response to certain physical laws, and thus develop two actions from the air mass. One is a positive pressure lifting action from the air mass below the wing. The other is a negative pressure lifting action from lowered pressure above the wing.

Different aerofoils have different flight characteristics. Many thousands of aerofoils have been tested in wind tunnels and in actual flight, but no one has been found that satisfies every flight requirement. Therefore the weight, speed, and purpose of each aircraft dictate the design of its aerofoil.

The principal structural parts of the wing are spars, ribs, and stringers. These are reinforced by trusses, I-beams, tubing, skin, or other devices, as displayed in Figure 5.3. The wing ribs determine the shape and thickness of the aerofoil. Attached to the trailing edges of the wings are two types of control surfaces referred to as ailerons and flaps. To design the ribs designers need information of camber of upper surface, camber of lower surface, chord line, and mean camber line. Designers may search knowledge about the ribs.

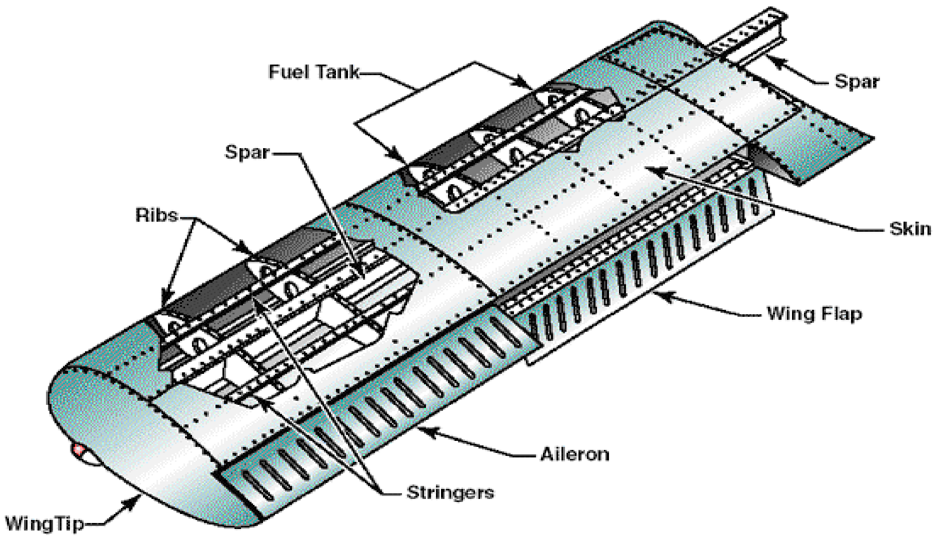


Figure 5.3: Wing Components (Source: FAI)

5.2 Agent Design of the System

5.2.1 Architecture of the Agent

To move from the proposed methodology to an implementation, some definitions should be considered. The software Agent is distinguished itself by its autonomy, cooperate, and learning. External stimulation motivate an Agent, meanwhile, an Agent is able to start new communicative acts by itself. Both messages received from external and its internal state could effect on the Agent's action. An Agent could interact with other Agents in order to pursue its goals and could even develop an overall strategy. Different from the client/server system, an Agent has the right to refuse requests.

The intrinsic Agent properties impact the concrete design of an Agent, which is reflected in Table 5.1 (Bellifemine, et al 2001).

Theoretical requirements	**Design solutions**
Agents are autonomous	Agents are active objects
Agents are social	Intra-Agent concurrency is needed
Messages are speech acts	Asynchronous messaging has to be used
Agents could say no	Peer-to-peer communication model is needed

Table 5.1: Design Solution to Meet Theoretical Requirements

To be autonomy, each Agent should be an active object. Each Agent must have at least one Java thread to be able to start new conversation.

An Agent may engage in several conversations at the same time, which requires capability of dealing with a number of concurrency.

Asynchronous message has to be used to represent an information exchange between two independent entities also has the additional benefit of producing more reusable interactions.

The last requirement suggests that an Agent could accept, refuse or ignore a received message. This characteristic replaces purely reactive message handlers and advocates by a pull consumer messaging model.

The autonomy requirement stresses that each Agent should have at least one thread. The Agent social character pushes towards many threads per Agent. However, the maximum number of threads has been clearly restricted by hardware and operation systems. Efficiency and cost are often at odds with each other. To balance them, several costs need to be evaluated, including thread creation and deletion, synchronisation between threads, and thread scheduling.

Although Agents may differ from each other, they have some generalities. The communication mode and process execution machine could be the same. The differences are decision strategy, action, knowledge representation, and so on. The agent architecture in the Multi-Agent System could be illustrated as Figure 4.1.

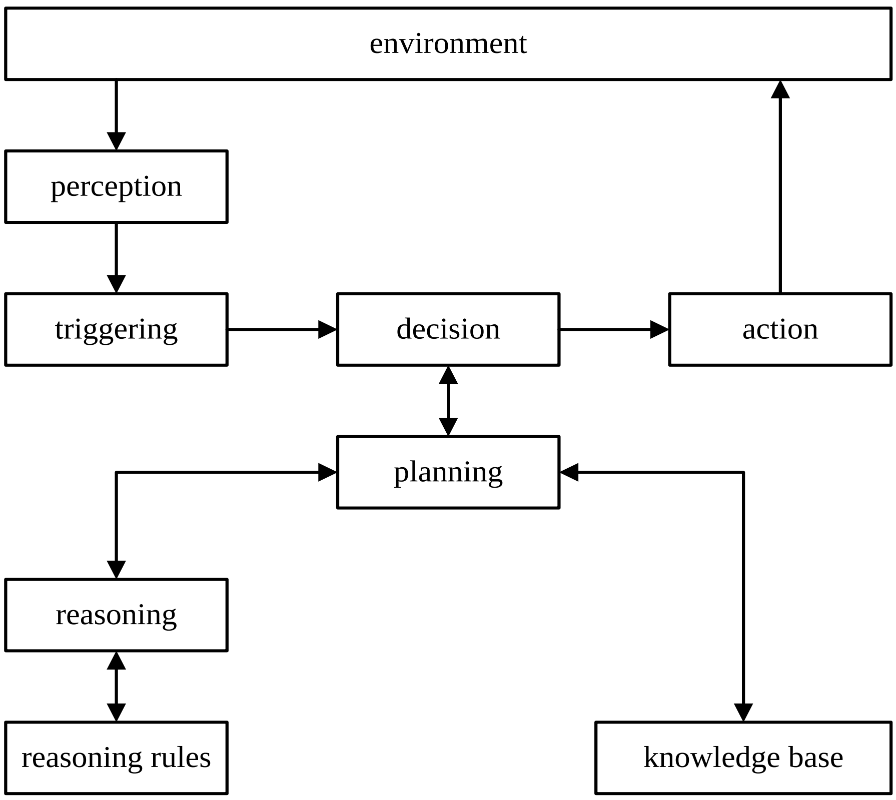


Figure 5.4: Architecture of the Agent

The environment module considers two aspects of the outer environment. One is the computer performance of the agent user, like the performance of the system resource. The other is the performance of the Internet, like the busyness of the network, the response time of the server. The perception module senses the events of the agent. For example, the Dispatch Agent senses the events like user registration, user login, reflection of user, search action of user. When the Agent senses one or more events, the triggering module triggers the decision module. The planning module processes the information from the knowledge base upon the reasoning rule, and passes them to the decision module. The knowledge base in the agent is the database stores the information of user and search results. The action module directs the action of the Agent based on the decision. For example, the actions of the Search Agent include: requesting the user interest model from the Learning Agent, obtaining primary search results by normal search techniques, filtering the results based on the contents, and filtering the results based on the cooperation with the Learning Agent. Each Agent of the Multi-Agent system is not isolated. They interact with each other.

5.2.2 Module Design of the Agent

Based on the proposed methodology in Chapter 4, different Agent corresponds to different events. However, the implementation of these Agents bases on the same Agent framework.

The Dispatch Agent interacts with the user directly in this system. It has the functions as follows:

- record and analyse the acts of user;
- obtain the reflection information from user and deal with it;
- extend and optimise user's query words;
- interact with the Learning Agent and Search Agent, dispatch the user's act reports to the Learning Agent for further processing.

According to the functions of Dispatch Agent, it needs to deal with the events like:

- user registration;
- user login;
- reflection of user;
- search action of user.

The actions of the Dispatch Agent corresponding to these events are:

- create basic information for user in the database, and notice the Learning Agent to create new user interest model for new user;
- obtain user login information, and send user information to the Learning Agent;
- record reflection information of user;
- record user's query, obtain user interest model from the Learning Agent, optimise user's query, and send it to the Search Agent and the Learning Agent.

The Learning Agent is an important part of the system. It has the functions as follows:

- obtain use's basic information from its interaction with the Dispatch Agent, create user interest model;
- improve user interest model by the analysis of historical data;
- cooperate with the Dispatch Agent, analyse user interest and habits, and update user interest model;
- manage the access and use of user interest model.

Corresponding to the functions of the Learning Agent, it needs to deal with the events like:

- new user registration;
- obtain historical data for creating model;
- obtain information about the reflection and action of user;
- receive requirements about user interest model from other Agents.

The actions of the Learning Agent regarding to these events are:

- create user interest model by artificial modelling;
- create user interest model by historical data;
- obtain user interest model according to other Agent's requirement, and send it to the Agent.

The process of creating user interest model could be described as:

- obtain the set of available model creation plans;
- choose the optimal model creation plan according to events and conditions. Once if the creation is failed, choose the next plan until all the plans are failed.

For most search engines, the search results are listed based on its relevance to the search requirement. It does not consider the user's interest. If different users send the same requirements, the results back to them are the same. Normally, the amount of the result documents is big. Users view these documents to pick up what they are interested in.

The Search Agent will cooperate with the Learning Agent to figure out the relevance of the documents to user interest model, besides the relevance to the requirement. It will filter the results and return organised results to the user. The events of the Search Agent are:

- user logs in but does not search;
- user logs in and do search.

Accordingly, the actions of the Search Agent are:

- request the user interest model from the Learning Agent;
- obtain primary search results by normal search techniques;
- filter the results based on the contents;
- filter the results based on the cooperation with the Learning Agent.

5.3 Implementation of the Agent

The Learning Agent and Search Agent are the un-replaceable parts of the proposed Multi-Agent system. The efficiency of them impacts the property of the proposed search system. Based on the module design which is described in last section, the implementation of the Learning Agent and the Search Agent is discussed in this section.

5.3.1 Implementation of the Learning Agent

To realise knowledge search, Learning Agent is an important component. Its main function is to create user interest model. To realise user interest model, the first step is to create basic model through user registration information, which is manual modelling. Then create model based on user's previous search history and perfect it during the process of using.

For each user, vector X is defined as a relevance mark eigenvector. x_i is the relevance mark of user for item i. Item i includes three classes, information type, keyword, and user basic feature. Therefore, one user has three eigenvectors to indicate his/her interest.

Based on statistical analysis, if the information types and keywords are listed based on the relevance, the 300 in front of the rank make sense to indicate the user's interests. The rest could be neglected. Therefore, the feature of type and the feature of keyword could be both defined as a vector whose dimension is 300. The dimension of user basic feature is fixed. It based on the amount of questions providing to new user during the registration process.

MySQL is selected as the database for the user interest models. The user interest could be represented by three tables. Table 5.2 shows the vector of information type. Table 5.3 shows the vector of keyword. Table 5.4 shows the user basic feature.

UserId	Type	Weight	Freq	Active Time

Table 5.2: tb_type_feature

UserId	KeyWord	Weight	Freg	Active Time

Tabel 5.3: tb_keyword_feature

UserId	UserName	UserInfo	Weight

Table 5.4: tb_basic_feature

"tb_type_feature" is the table of eigenvector of type. UserId is the unique ID of user, which is the primary key. Type is the title of the sort. Weight is the user's interest value of the current type. ActiveTime is the update time of the current weight. It is applied to delete outdated feature. The table will be deleted if the ActiveTime is longer than three months. Freg shows how many times the word appears.

“tb_keyword_feature” is the table of eigenvector of keyword. KeyWord is the searching key word. Other properties are the same as the eigenvector of type.

“tb_basic_feature” is the table of eigenvector of basic feature. UserName is the username of the user. UserInfo is the basic information of user. Weight is the weight value of the information. The value is defined during the creation of the user model. The information includes age, gender, education level, background, and so on. Since these information are fixed, ActiveTime is not applied.

The first step of creat the user interest model is to create basic model through user registration information, which is customised modelling. Basic user information is persistent and believable. This characteristic makes the manual customised modelling to be the best modelling method for the basic model. The Class ManualCreateUserProfile realise the modelling.

The process of the manual creation of model is as follows:

a) obtain user registration information from the Dispatch Agent;
b) sort the user registration information. Information like age, gender, education level, background are fixed information, which are defined as basic feature of user interest model. Weight value of them are set as 1;
c) set weight value of other registration information. For example, question like “are you involved with the project of ***”, if the answer is yes, set the value as 1, if the answer is no, set the value as 0, if the answer is partly involved, set the value as 0.5.

This information normally is persistent and believable. The system will not change the model automatically except the user change it.

The large amount historical data from pervious search system could be used to create user interest model. There are two types of modelling. One is the statistic method based on the user’s historical query log. The other is the modelling based on the historical accessing records.

User interest model could be created by the statistic of the user's historical query log. The algorithm is realised in the Class of StatisticalCreateUserProfile. The process is as follows:

a) pre-treatment of the data: the formats of query log are different for different application. Format them as "accessing times, query". Divide the query, combine the accessing times of the same keyword, and create the file with ranked "accessing times, keyword";
b) create user interest model after obtain the pre-treatment file:
   i. count the sum of the accessing times of all the keywords;
   ii. count the percentage of accessing times of each keyword, and rank from the high to the low;
   iii. the 300 keywords in front of the rank are set as the user feature, and the percentages are set as their weight value.

The vector of type modelling is similar to the keyword.

The historical accessing record reflects the interest of user to the information in one web page or document. The algorithm is realised in the Class of HistoryCreateUserProfile.

Since the vector of type modelling is similar to the keyword, only the process of vector of keyword modelling is described as below:

a) obtain the historical user access records;
b) obtain words in the content of one web page;
c) count the frequency of each word;
d) statistic the weight of these words, and rank them;
e) access accumulative database, sum the previous frequency and the new frequency of each word, statistic the new weight, and get new set N;
f) judge whether the previous eigenvector in the model is beyond 300;
g) if yes, compare the words in set N with existing words in the user interest model, replace with the new words whose weight and frequency are higher;
h) if no, put the high ranked words in set N into user interest model until the eigenvector reaches 300 if the frequency is larger than the threshold;
i) put all the rest words not in the model into the accumulative database;
j) complete update of user interest model.

The keyword frequency is considered as well as the weight. If the weight of one keyword is high, but its frequency is low, this keyword should not be updated to the vector.

5.3.2 Implementation of the Search Agent

Search Agent is the key component in the knowledge search system. Search Agent collects information from server. User provides search request through web browser one time. Then the Agent could search for the Intranet continuously and send back search results in web pages. A Search Agent relies on the encapsulation of the system and transparency to users. The efficiency of Search Agent impacts the property of the proposed search system.

The Search Agent deals with two events which are realised by Class LoginEvent and SearchEvent. The Class of SearchEvent is realised as below:

```
class SearchEvent{
private:
    int eventType;
    string attachmessage; // user's query
public:
    int getEventType();
    int getAttachMessage(); // obtain the user query
};
```

These two events are both created by the Dispatch Agent, and disseminated to other Agents.

The Search Agent is under the charge of the Class of SearchAgent. The key part of its realisation is as follows:

```
class SearchAgent {
public:
    static SearchAgent * getSearchAgent () {
        if(0 == _instance){
            _instance = new SearchAgent;
        }
        return _instance;
```

```
        }
        Void pleaseHandle(int EventType);
    Private:
        Event myEvents [];
        Plan myPlans [];
        SearchAgent (void) {
        }
        Virtual ~ SearchAgent (void) {
        }
        Static SearchAgent * _instance;
    };
```

Two important plans of the Search Agent to filter the information are ContentFilterPlan and CollaborativeFilterPlan. They realise the information filter based on the content and based on the collaboration.

The filtering approach based on the content is realised by the Class of ContentFilterPlan. It is detailed as follows:

a) calculate the relevance of the user interest eigenvector of keyword and the document index word;
b) set u as the vector of user interest, v as the eigenvector of web page index. If the relevance sim(u,v) is higher than the threshold, add the page to the result list based on the rank of relevance. The threshold should not be low, in order to ignore un-relevant results;
c) the relevance could be simplified, for example, represented as 1-5, if needed.

The filtering approach based on the collaboration is realised by the Class of CollaborativeFilterPlan. It is detailed as follows:

a) employ user's basic interest model, figure out the nearest neighbour of the user;
b) estimate user's interest model according to the keyword interest model and eigenvector of type model of the nearest neighbour. The method is to use the neighbour's average value of each vector. If more than 300, the values in the front are used;

c) use the method of ContentFilterPlan to the new user interest eigenvector to filter based on content, and add the new eigenvector to the database.

To obtain primary search results from the Internet, the Search Agent collaborates with the Internet search engines in order to get the URL information of the WWW servers. It realises part of the priority of utilising Agent technology in the proposed search system. This system does not need to send a crawler to read the web page, make a copy, or return the copy to the index of the search engine. It obtains the URL information of the WWW servers from collaborative Internet search engines. Dependence on the web bandwidth is declined. Search time is reduced. In parallel, the returned URL information is sorted according to the request, which makes the Multi-Agent search system more targetable and efficient.

One search engine could hardly cover all the web resources in the world. The user may need to use different search engines to achieve the goal. The Search Agent assembles the URL information from different Internet search engines in parallel to avoid the limits. However, each Internet search engine has its unique query interface, dissimilar search techniques, and various results.

The Internet search engines have diverse query input forms. Some have only one case to input the query. Some have complex query input fields besides text input case, like radio button, check box, menu selection, and so on. Although the complexity of the forms is different, the composing elements are the same. The basic composing method of an input form is extracted as follows:

```
<html>
<body bgcolor=”#FFFFFF”>
<h2 align=”center”>input the composing elements of the form</h2><hr>
<form action=”echo.asp”method=”post”>
<p>name: <input type=”text” size=”20” name=”Name”></p>
<p>interest: <input type=”text” size=”20” name=”Interest”></p>
<p> <input type=”submit” value=”submit”></p>
</form>
</body>
</html>
```

The <form> defines the important parameter of action and method. The parameter of action specifies the program in the search server that processes this form. The parameter of method sets the transfer method of the parameters. Post and get are the common methods.

The input field is the most diversiform part of the form. To provide friendlier operating interface, HTML provides input fields like text case, radio button, check box, menu selection, and so on. No matter what field is utilised, the important step is to name each field, like the “Name” and “Interest” in this form. The program in the server could then pick up the information in this field according to this name.

The submit button transfers the information to the search server when it is pressed. Therefore, the query despatched to the Search Agent should be transferred to the Internet search server. The method used to is to add the parameter after the Internet address, in the pattern of “? parameter 1=parameter value 1 & parameter 2=parameter value 2…”. The key approach is to analyse the web page input form, extract the parameters, compose the input query to character strings in pattern, and transfers to the server. The example the extract of the form of an Internet search engine is given as follows:

```
<td><formaction=”/cgi-bin/allsearch”
        method=”GET”name=”queryform”>
<input type=”hidden” name=”cdtype” value=”GB”>
<input type=”hidden” name=”dc” value=”ALL”>
<input type=”hidden” name=”display” value=”Standard”>
<input type=”hidden” name=”fno” value=”0”>
<input type =”hidden” name=”method” value=”Fuzzy”>
            <input type=”hidden” name=”operation” value=”And”><p><b>
        query string: </b></p>
            <p><input type=”text” size=”24” maxlength=”80” name=”word”></p>
            <p><input type=”submit” value=” search ”,>
            <input type=”reset” value=” reset ”,></p>
            </form>
```

In this case, the information should be transferred to the server is: http://the address of the search server/cgi_bin/allsearch?cdtype=”GB”&dc=”all” & display=”standard & fno=”0” & method=”Fuzzy” & operation=”And” & word=”the string about the user’s query”.

The user could decide the way of gathering the information by setting the parameter from the user interface. One way is gathering the information only if it is sought out by all search engines. The other way is gathering the information sought out by every search engine, and deleting the repeated ones. The Search Agent obtains the URL information according to the user's setting, and saves the information.

The empirical analysis of iProspect (Breese et al 1998) shows that 62% users read only the first result page of the Internet search engine. If there are 100 results per page, the upper limit of the filtering results amount is 300 and the lower limit is 100, the results displayed on the first page will not be filtered, and the maximum result pages are 3.

The process of Search Agent is realised by the Class of FilterContext as follows:

a) if it is SearchEvent, use QuerySearchPlan to realise the normal information search algorithm to obtain relevant documents to user's query;
b) if the results amount is less than 300, the search is completed;
c) if the results amount is more than 300, apply the ContentFilterPlan to filter the results;
d) if the amount of the results after filtering is more than 100, return the 300 results in front;
e) if the amount of the results after filtering is less than 100, it means that the filtering condition is too restrict. Apply the method of CollaborativeFilterPlan;
f) if the amount of the polished results is still less then 100, return the results to the user and the search is completed;
g) if the amount of the results is 0, do not filter, and return all information obtain by QuerySearchPlan to the user. It means there are errors of the user interest model, and filter all the results.

5.4 Summary

The development of a knowledge search system for facilitating knowledge management among product design and development team members has been explained. There are three aspects need to be considered: (i) the Agent building and editing tool to construct and maintain the Agents; (ii) the mechanism to control and integrate Multi-Agent, including Agents communication, interaction, and cooperation; (iii) the mechanism to disseminate the results to designers. In doing so, research objective of developing a knowledge search system has been met.

The industrial context is considered first. The construction tool is selected as Lucene and Eclipse. The architecture of the Agent is determined. Module design of the Dispatch Agent, the Learning Agent, and the Search Agent are developed based on the functions, the events, and the actions. In the module design of the Learning Agent part, the process of creating user interest model is described. Then the implementation of the Learning Agent and the Search Agent is built up separately. For the Learning Agent, it focuses on the creation of the user interest model by manual method, statistical method, and historical method. For the Search Agent, it focuses on the filter approaches based on the content and on the collaboration. The approach that the Search Agent obtains information from multiple Internet search engines is developed as well.

CHAPTER SIX: TESTING OF THE MULTI-AGENT KNOWLEDGE SEARCH SYSTEM

The previous chapter discussed the development of a methodology to search enterprise knowledge for designers during the product design process. This work formed the main part of the study. To conclude the distribution of the study, the methodology required testing to verify the system developed. In doing this, the work addressed the research objective: to test whether the methodology developed in previous chapters presents an effective way of searching and sharing the knowledge between engineering designers.

This chapter discusses the testing design, the testing sessions using different searching requirements. It details the aim and scope of the testing, and the consideration of the testing design.

6.1 Aims and Scope of the Testing Excise

Implementation involves demonstrating the functionality of the knowledge search system, and providing a test of the methodology developed. This was carried out by using the knowledge search system to capture knowledge and information used and generated by tasks in the new product development process of the case study company.

6.2 The Industrial Process that is Used for the Testing

As described in Chapter 4, the product development is a circular process which could be described as Figure 6.1. It is the design process of C919 nose. The initial structure design is conforming to airworthiness standard. It has various requirements about space, horizon, layout, aerodynamics, and so on. The shape of the nose affects the icing issue. The Structure Group cooperates with the System Group for better de-icing solution. The corrosion of deicing fluid relates to the Material Group. Materials used need the Intensity Group to ensure rational intensity. Material intensity influences aerodynamics design. The Aerodynamics Group collaborates with the Structure Group on docking. Structure design requires large windows for pilot horizon. The Decoration Group considers the ergonomics design to satisfy it. The Decoration Group reckons for the barycentre problem as well to adapt to aerodynamics design. Therefore, each Group has its main focus, and has to compromise with other Groups on the development. There are various conflicts as well. The Structure Group generates the initial structure of the nose. The Aerodynamics Group calculates and provides suggestions for structure modification. The Loading Group designs the load. The structure is verified by the Construction Group and the Intensity Group. Then the Decoration Group conducts internal layout, reckoning in ergonomics. Groups are adjusted to each other. All requirements, conflicts and impact factors are taken into consideration. Normally, after three to five iterations, engineers will have optimized a reasonable design which is acceptable to every Group.

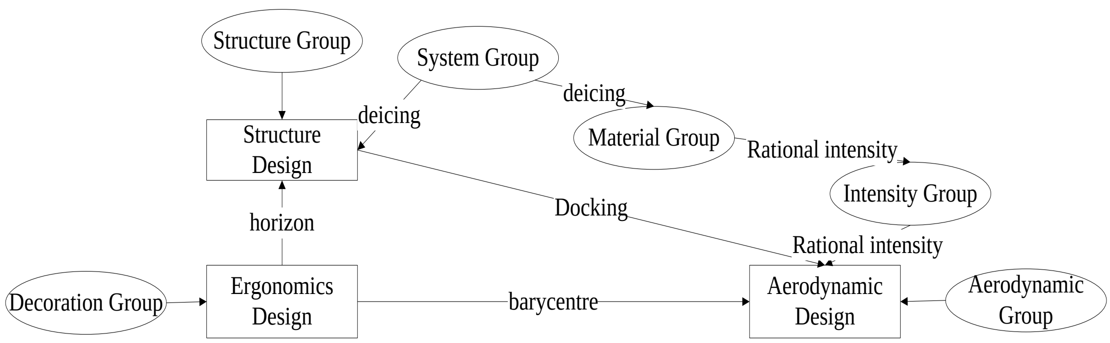


Figure 6.1: C919 Nose Design Process

The designers in structure group need knowledge including structure, system, decoration, assemble, on-board equipment, and coordination support.

6.3 **Knowledge to be Searched**

In the aircraft manufacturing enterprise, knowledge stored in different places and in different types need to be searched. As described before, knowledge stored in databases like shared database, knowledgebase, data warehouse, information systems like decision support system, expert system, management information system, and other computer systems like CAD. All the knowledge should be searchable. Here parts of the functions are realised. Knowledge tested includes knowledge in shared database, management information system, and web.

Search for knowledge within enterprise could be categorised into types, through an analysis of survey results and an analysis of search log data. The types include searching for definitions, persons, experts, and homepages. Search intentions are need to understand the underlying needs of search queries. Figure 6.2 is the user interface of the enterprise search system. If users would like to search in linked databases, shared files, and knowledgebase, they could input their queries in the input box. The results will display in the following box with file name or primary key value, where they are from, and the path to reach the original data. If users would like to search on the web, they could use this user interface as well. But the results will display in a popup page.

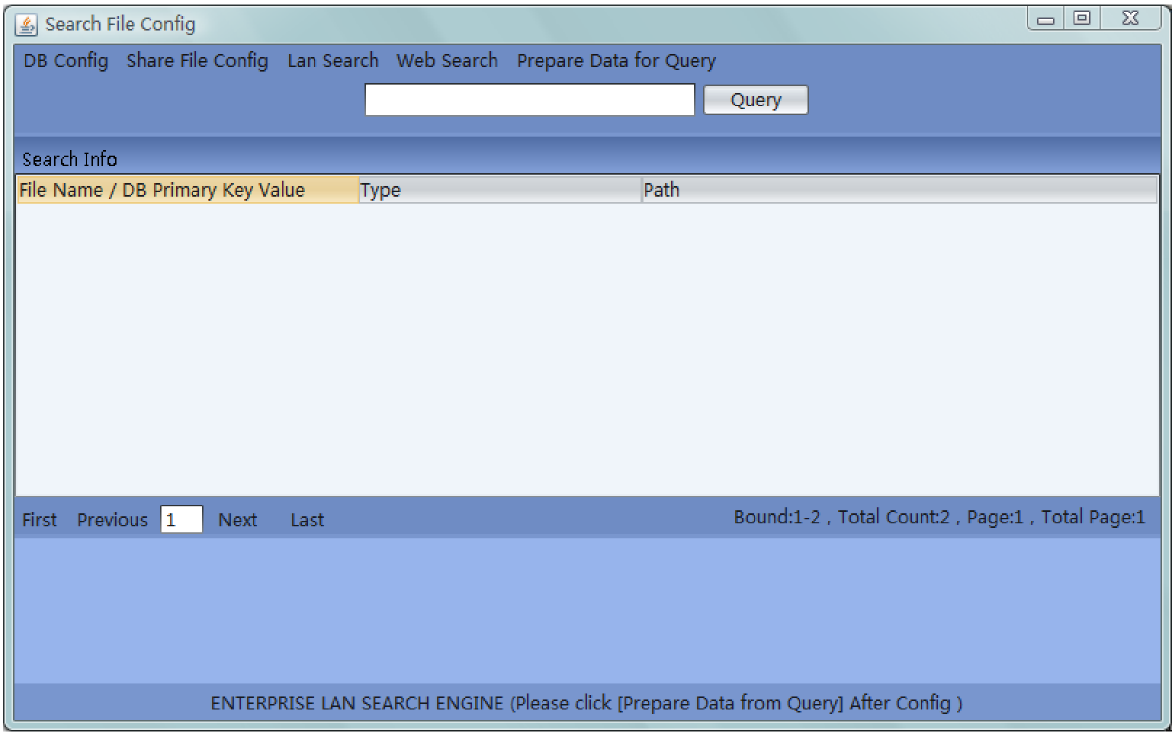


Figure 6.2: User Interface of the Enterprise Search System

6.4 Search for Knowledge Within Enterprise and Outside Enterprise

Within the enterprise, users' search needs are less diverse than those outside the enterprise, because the users are information workers and their motivations of conducting search are business oriented.

Most structured data is stored in the database. To search knowledge stored in database, the main focuses are the storage method of index document, the fields of index document, and the resources of data. Information of files and databases are abstracted as a class

SearchInfo. This class includes URL of database or server, File Path, Table Name, Column Name, Key Name, Search Info Type, and Search Info Value, as the following code does.

Figure 6.3 shows the connection of the database and the search system. Since this test is conducted in one computer, the localhost:3306 means this computer. However, shared documents and databases in different servers could be searched. For example, if the database is based in another computer whose IP address is 192.168.1.200, the URL is set as jdbc:mysql://192.168.1.200:3306/remote_user, in order to connect the database to the search system.

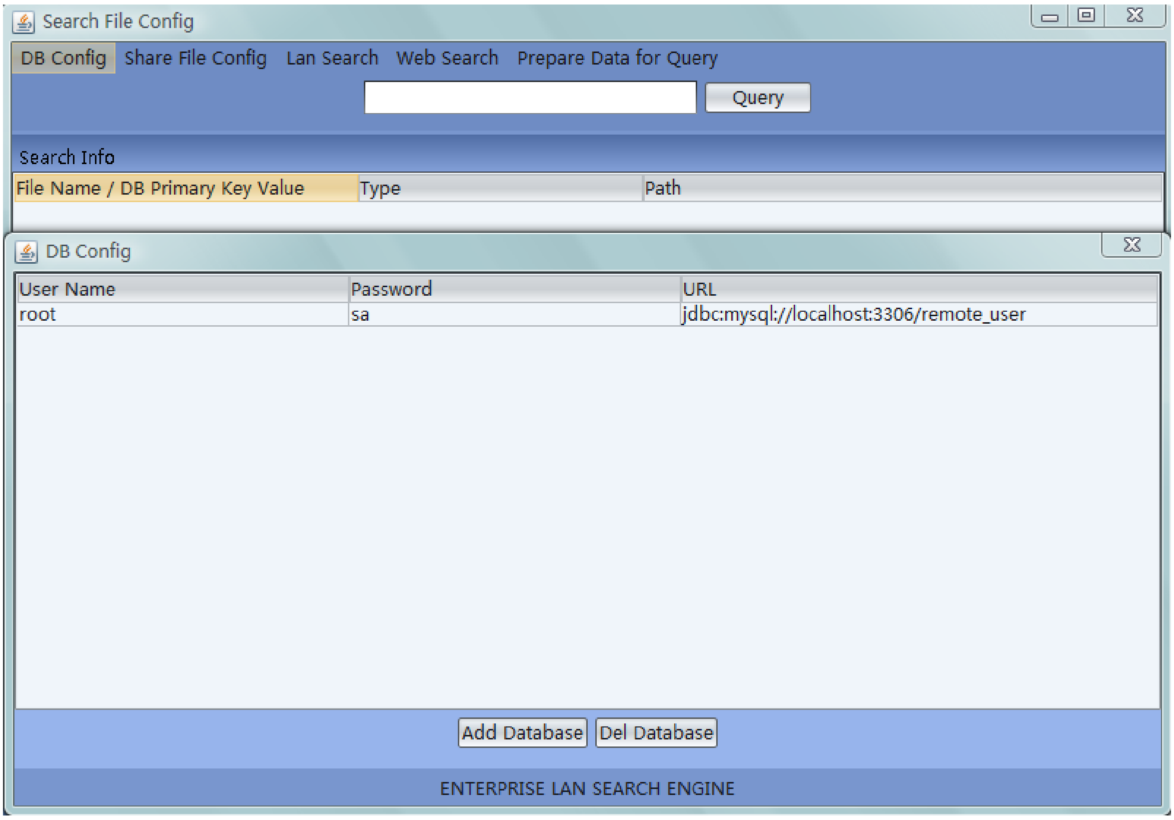


Figure 6.3: Set the Connection between the Database and the Search System

In this database, various information is stored. Users could use this search system to get information about designer, including their department, expertise, and contact number, as displayed in Figure 6.4. Besides searching by name and ID, Users could search by department or expertise to get designers' name and other information. Figure 6.5 shows the search results of an example of searching by department.

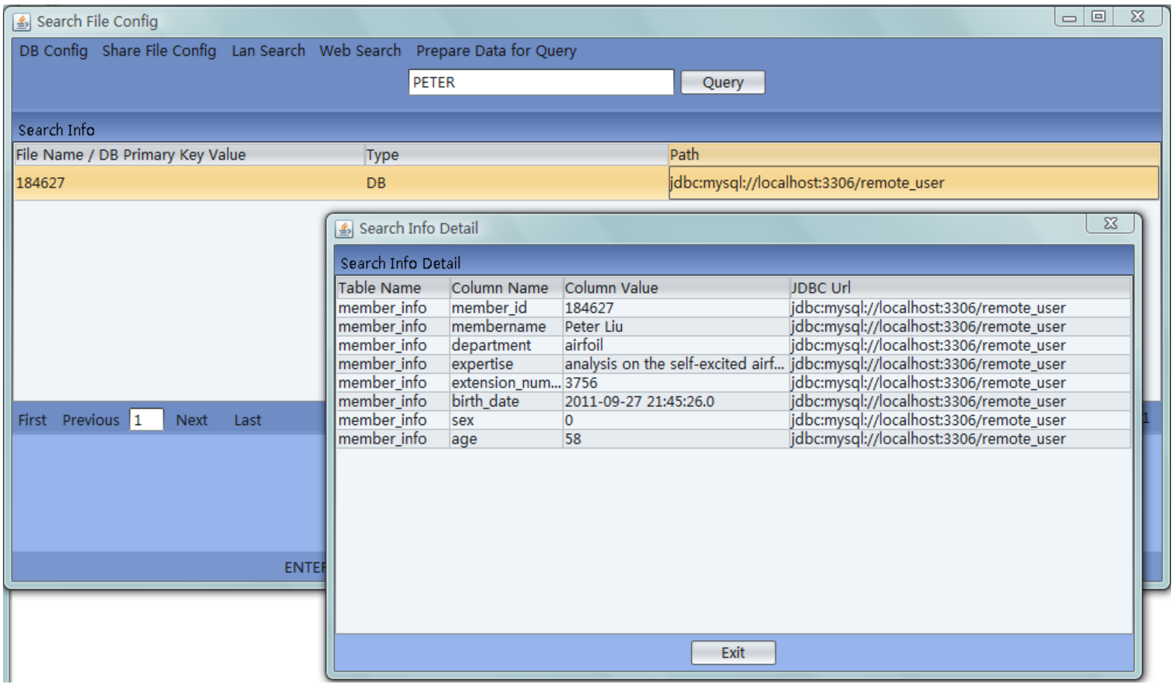


Figure 6.4: Search Result of Designer Information by Name

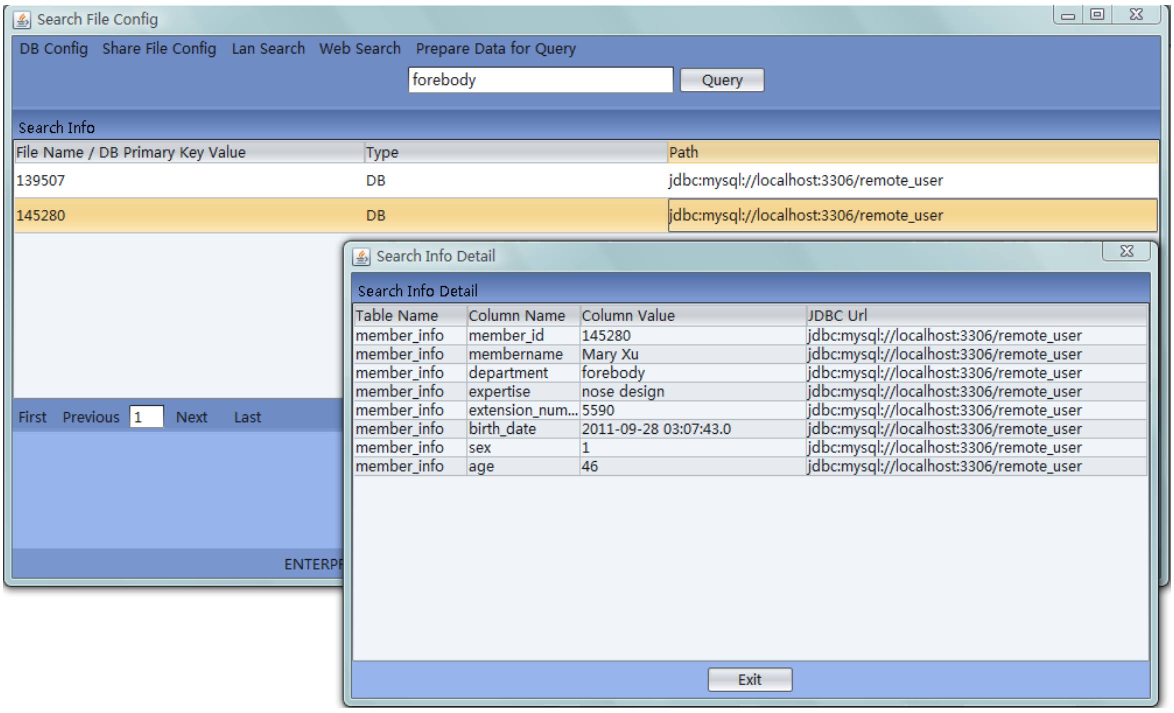


Figure 6.5: Search Results of Designer Information by Department

Figure 6.6 is an example of searching what documents are related to certain users. If the user is doing nose designer, the documents showed in Figure 6.5 are what he/she need to deal with frequently. The searching result shows the abbreviations of file names. Full name could be displayed by double clicking the results.

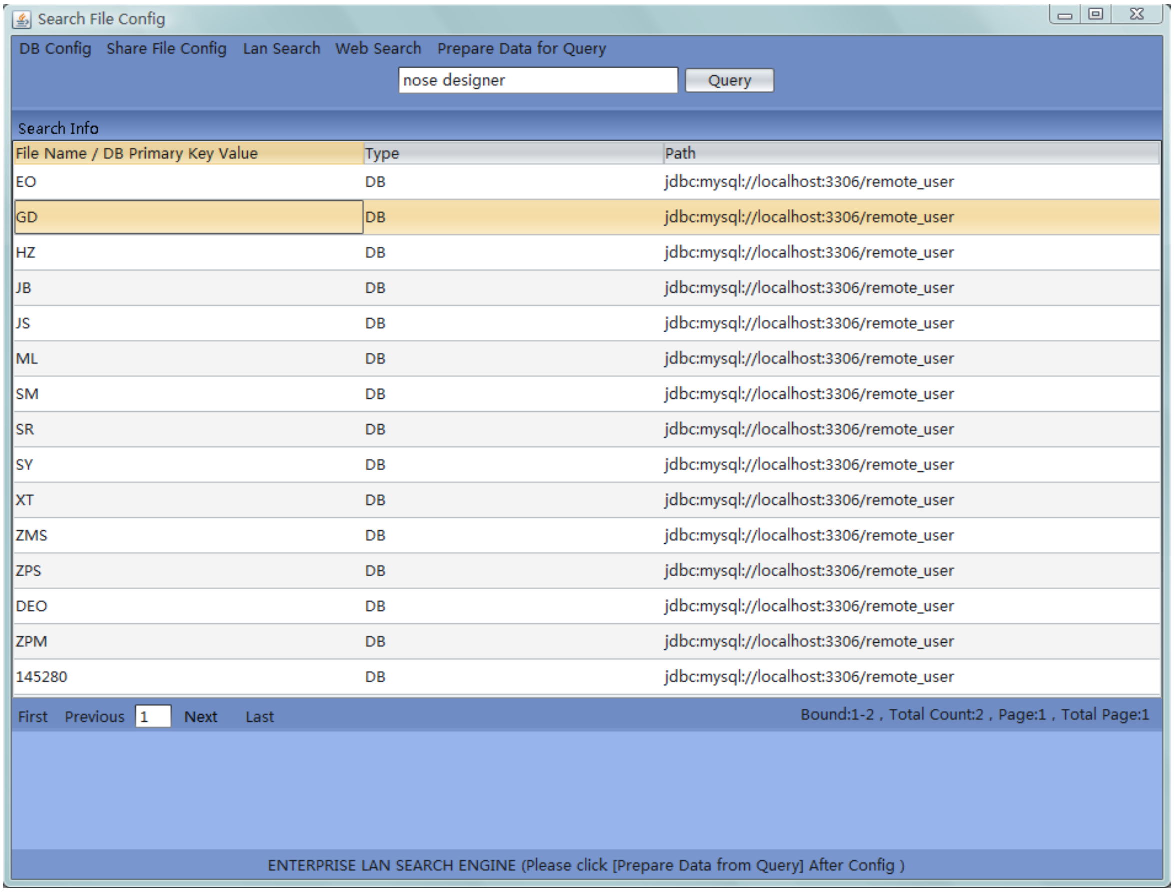


Figure 6.6: an Example of Searching Document Types by Users of the Document

A simple book management system is settled since there is no industrial MIS to be tested after the search system is developed. Figure 6.7 shows the login progress to the book management system. Figure 6.8 is search result of the book whose name contains aircraft.

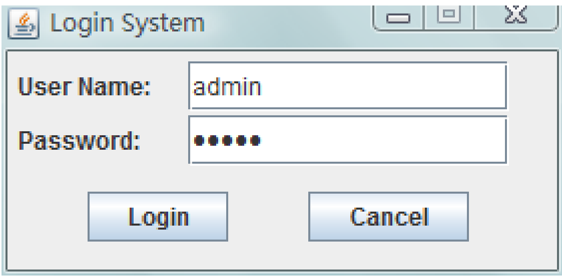


Figure 6.7: Login to the Book Management System

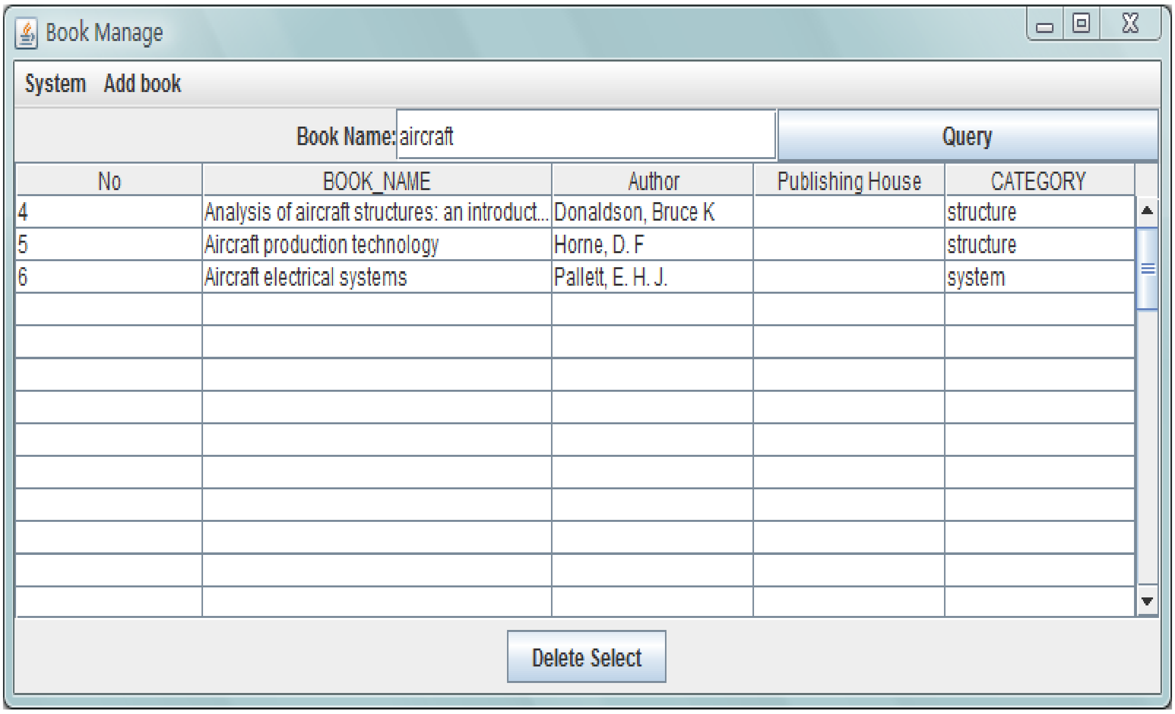


Figure 6.8: Search Result by Book Name

Shared files are required to be searched. Files could be searched by their name and content. Figure 6.9 is an example. It returns two files which contain the word of boeing in the content. Double clicking the result could open the file to read. Files could be searched as long as they are shared to the user. In this example, the user is root, and password is sa. 192.168.0.4 is the IP of this computer. Since Lucene's API is independent of the file format, text from PDF, HTML, Microsoft Word, and OpenDocument documents, as well as many others, could all be indexed as long as their textual information can be extracted.

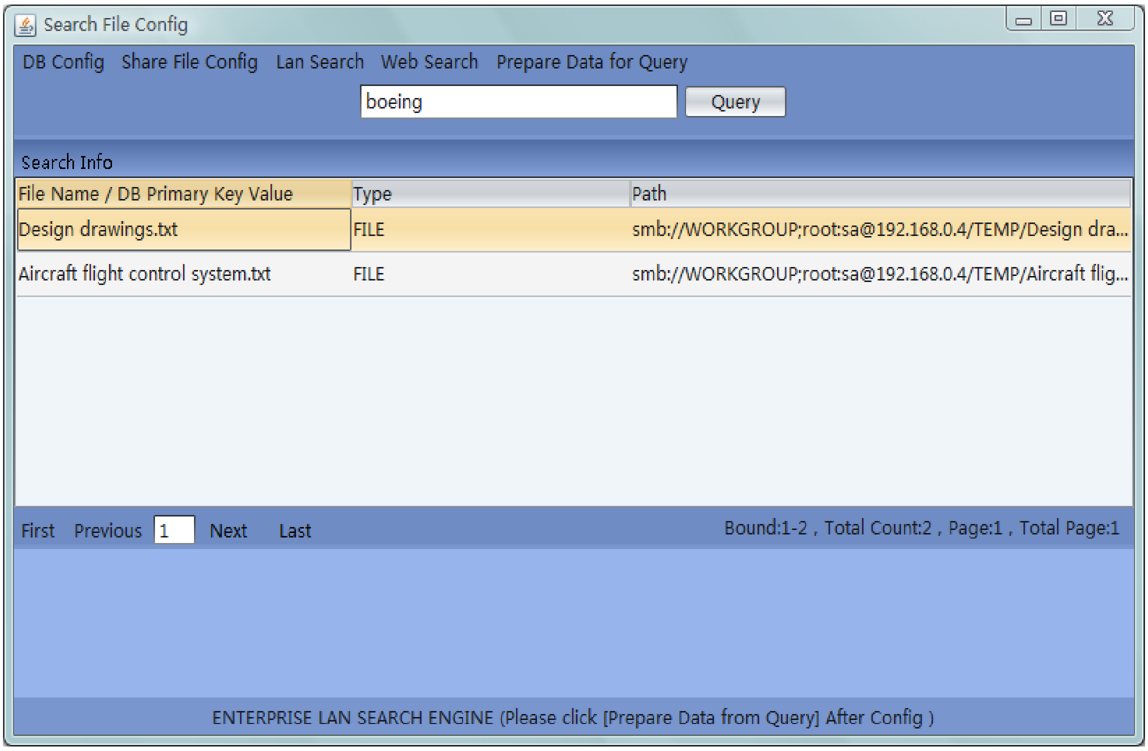


Figure 6.9: Example of Searching Shared Files

Figure 6.10 displays the search result of the information stored in a knowledge base. The search system must have the admission to the knowledge base to get the data.

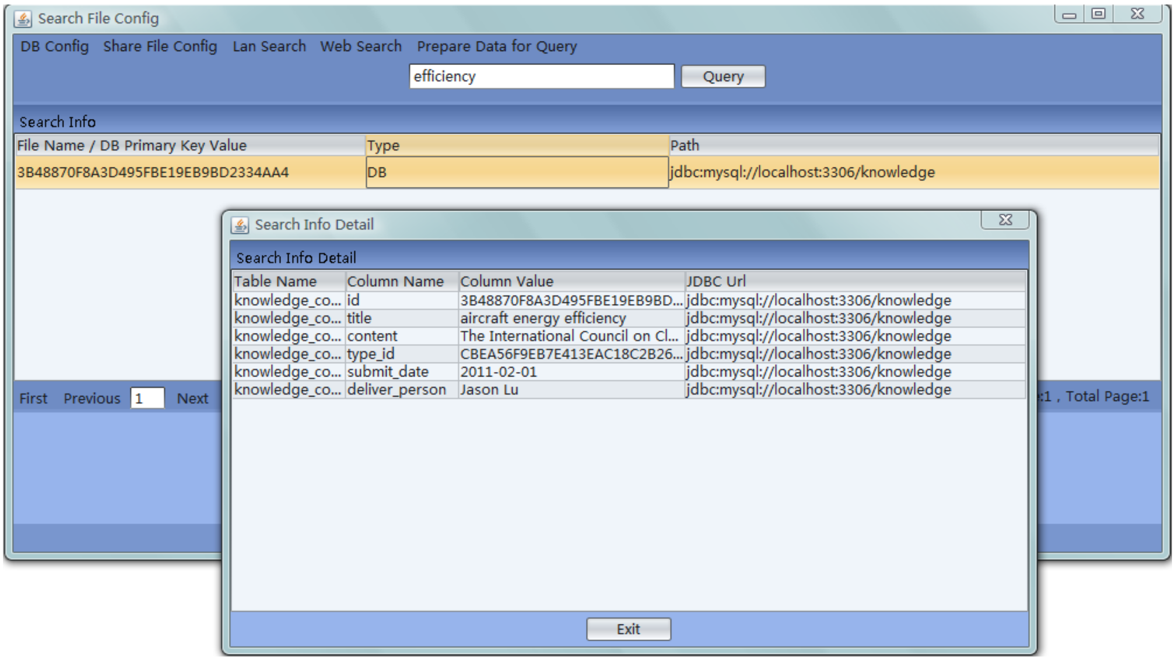


Figure 6.10: a Test of Knowledgebase Search

Users may could do web search to get knowledge and information outside the enterprise. Figure 6.11is an example of web search. The results are displayed in a popup window as Figure 6.12 shows, which employs bing as its web search engine.

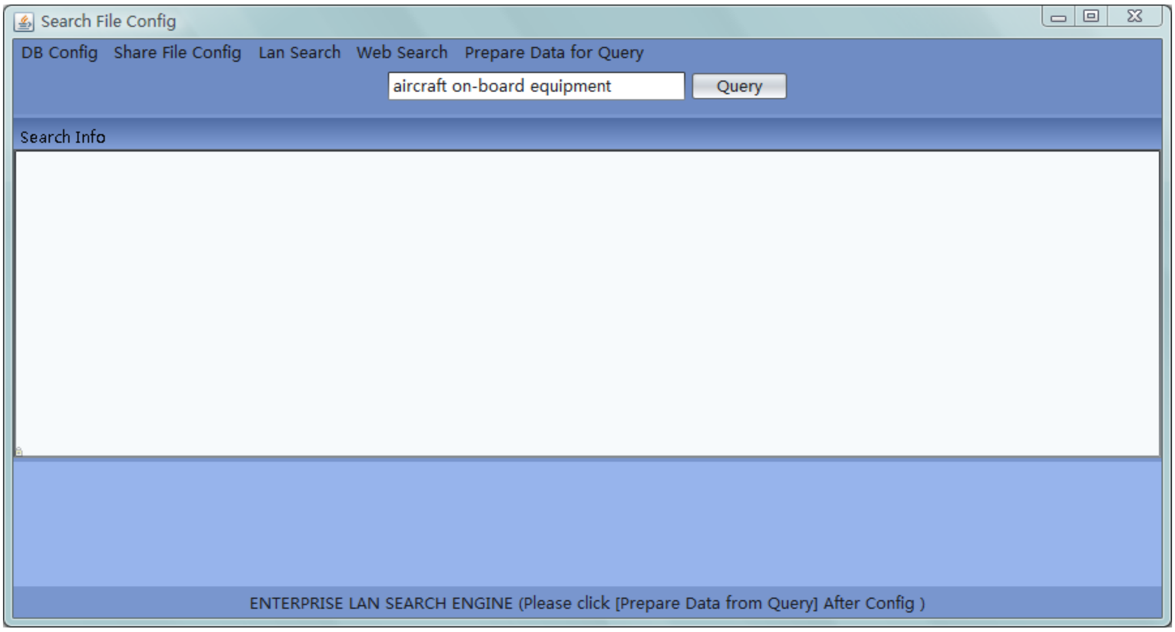


Figure 6.11: an Example of Web Search

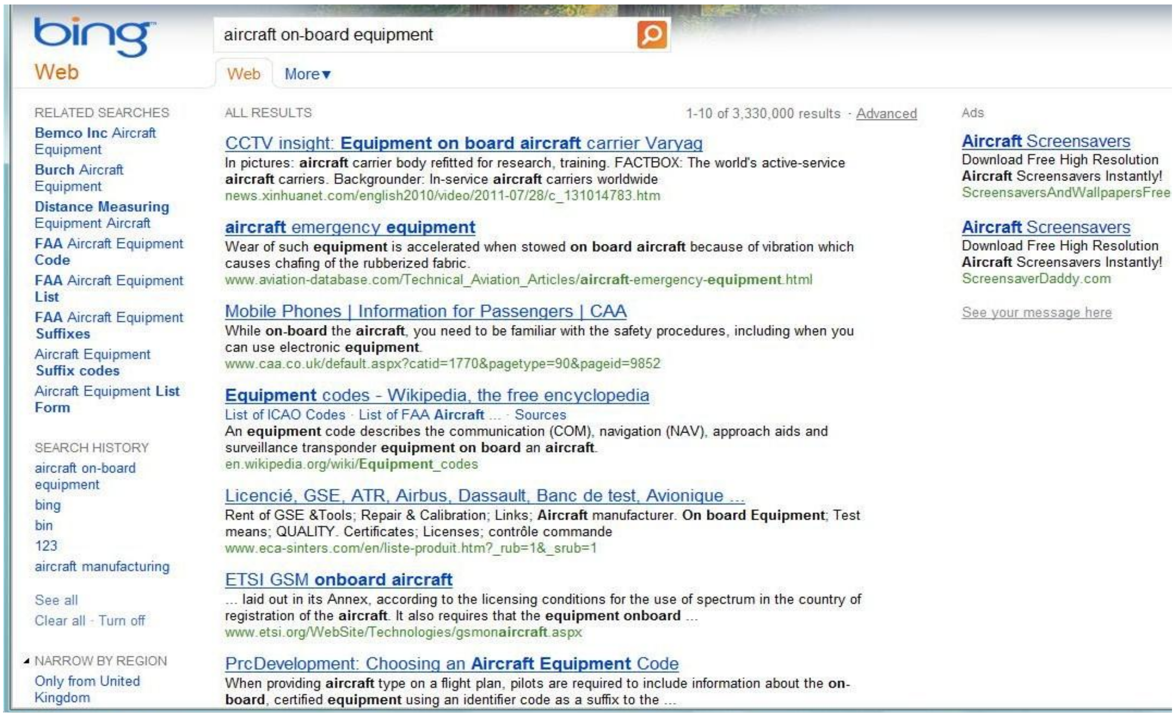


Figure 6.12: Example of the Results of Web Search

6.5 Structure of the Construction Classes

The complex classes and operations could be more clearly specified by the structure figures.

Figure 6.13 shows the three different data resources to be searched by the proposed search system. The bookmanager delegates the MIS since there is no reachable industrial MIS for the test.

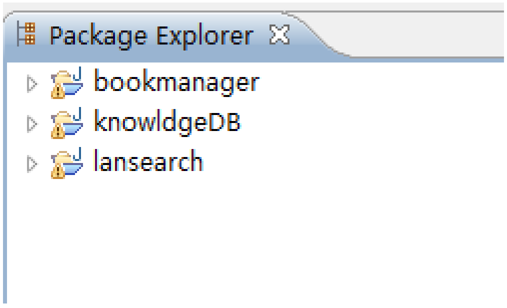
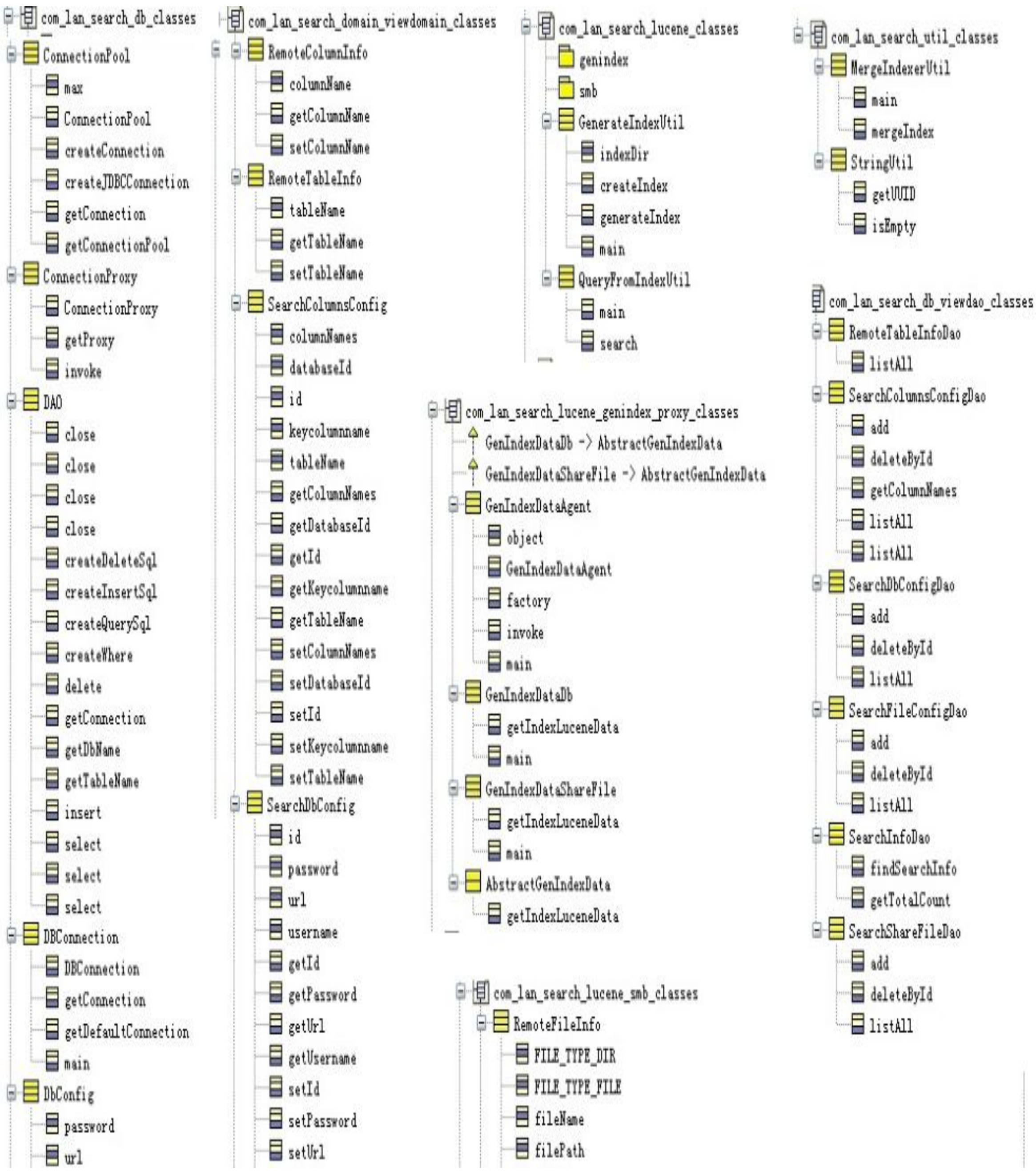


Figure 6.13: Structure of the Construction Packages

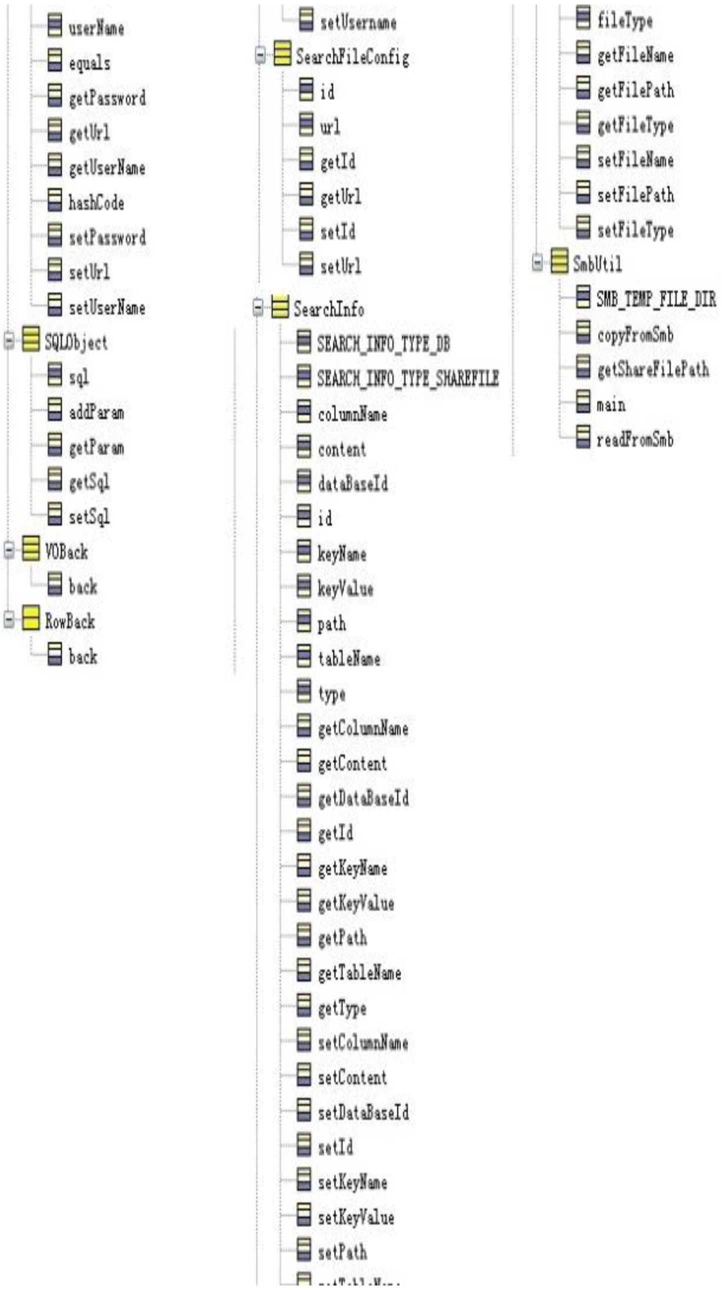
The lansearch delegates the shared database search. Figure 6.14 lays out the structure of the classes in lansearch. A detailed explanation is given for the lansearch. The rests are similar to it.



→ To be continued in the next page

Figure 6.14: the Structure of the Classes in lansearch

→ continued from the previous page



→ To be continued in the next page

Figure 6.14: the Structure of the Classes in lansearch

→ continued from the previous page

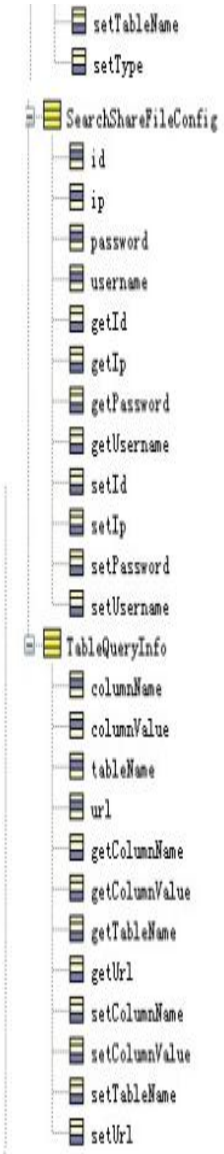


Figure 6.14: the Structure of the Classes in lansearch

The classes in Figure 6.15 are mainly served the operation of the database.

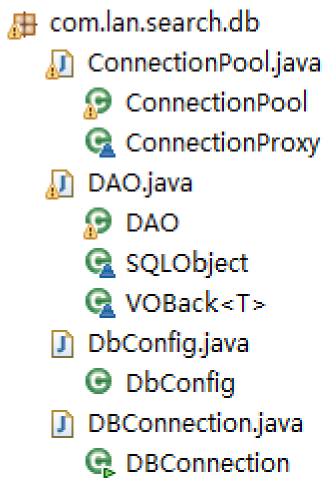


Figure 6.15: Classes Served the Operation of the Database

The classes in Figure 6.16 realise the function of reading data in database. Different DAO corresponds to different forms.

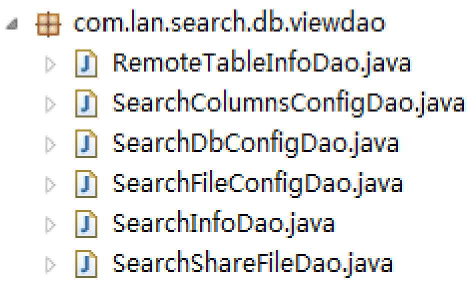


Figure 6.16: Classes for Reading Data in Database

The classes in Figure 6.17 define the display of the fetched information.

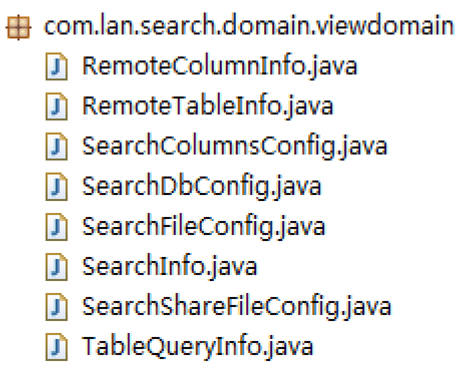


Figure 6.17: Plain Classes

The classes in Figure 6.18 are about the index generation and the index search by Lucene.

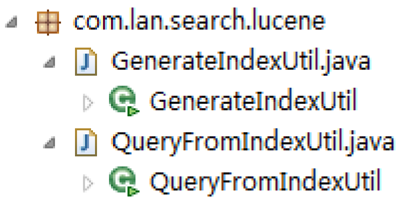


Figure 6.18: Classes of Generation and Searching Index by Lucene

The Classes in Figure 6.19 are the agents to generate index by Lucene.

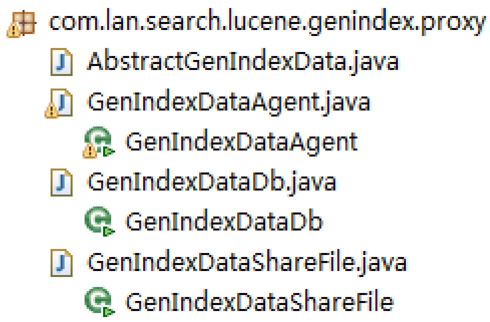


Figure 6.19: the Agent to Generate Index by Lucene

The classes in Figure 6.20 carry out the search in the Intranet.

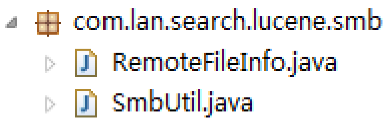


Figure 6.20: Classes Carry Out the Search in the Intranet

The classes in Figure 6.21 are used for display the swing UI. Swing is the primary Java GUI widget toolkit.

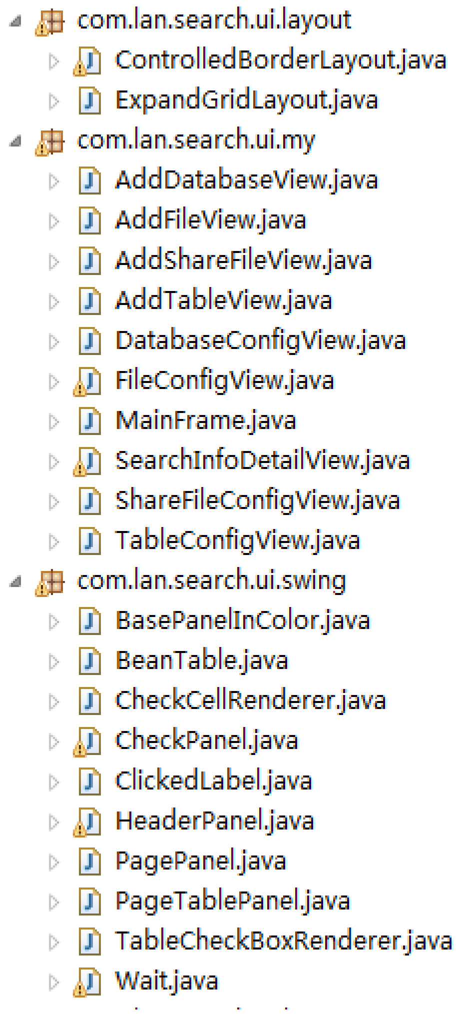


Figure 6.21: Classes for Swing UI Display

The classes in Figure 6.22 are tool classes.

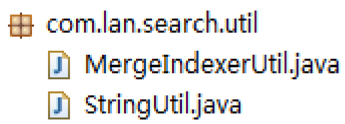


Figure 6.22: Tool Classes

6.6 The Classes for Lucene

Some of the classes for Lucene are explained here.

Figure 6.23 shows the structure of the AbstractGenIndexData. It has only one method:

```
public List<SearchInfo> getIndexLuceneData() throws Exception;
```

It is to obtain data for Lucene. The data returns in the pattern of SearchInfo set.

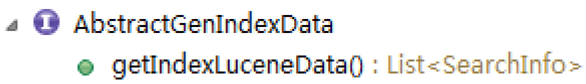


Figure 6.23: the Structure of AbstractGenIndexData

GenIndexDataDb realises the interface of AbstractGenIndexData. Its structure is displayed in Figure 6.24. This class is to obtain data from database for Lucene. The returned data is in the pattern of SearchInfor set. (refer to Appendix F)

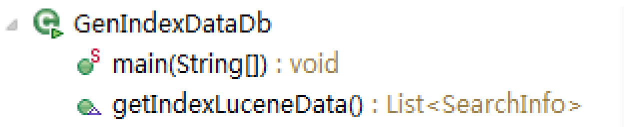


Figure 6.24: the Structure of GenIndexDataDb

GenIndexDataShareFile realises the interface of AbstractGenIndexData. Its structure is displayed in Figure 6.25. This class is to obtain data from shared files for Lucene. The returned data is in the pattern of SearchInfor set. (refer to Appendix G)

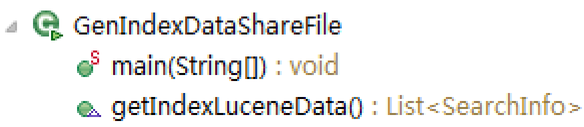


Figure 6.25: the Structure of GenIndexDataShareFile

In order to obtain data from other data resources besides database and shared files, the steps are realising the AbstractGenIndexData interface, the getIndexLuceneData method, and the SearchInfo set of data for generating index by Lucene.

GenIndexDataAgent is the Agent for generating index. It is employed to generate interface that inherits AbstractGenIndexData. (refer to Appendix H) Figure 6.26 shows its structure.

factory is to generate classes. Invoke method call in the getIndexLuceneData method. The approach is:

```
    AbstractGenIndexData abstractSubject = (AbstractGenIndexData)factory(new
GenIndexDataShareFile());
    abstractSubject.getIndexLuceneData();
```

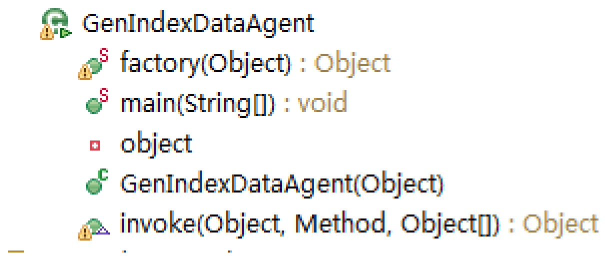


Figure 6.26: the Structure of GenIndexDataAgent

GenerateIndexUtil is employed to generate the index for Lucene. Its structure is displayed in Figure 6.27. GenerateIndex method obtains data from the classes that inherit AbstractGenIndexData, and generates index by createIndex. (refer to Appendix I)

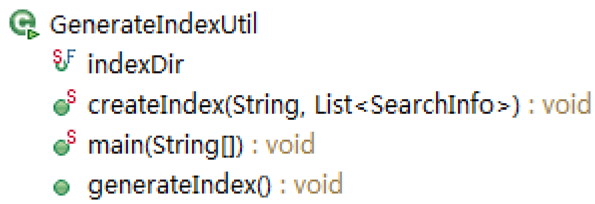


Figure 6.27: the Structure of GenerateIndexUtil

QueryFromIndexUtil is to search index of Lucene. (refer to Appendix J) Its structure is displayed in Figure 6.28.

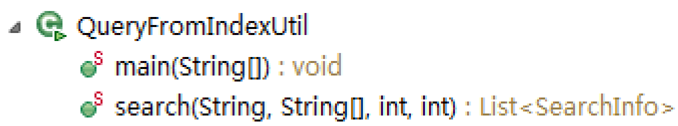


Figure 6.28: the Structure of QueryFromIndexUtil

SmbUtil is used to obtain data from the shared files of the Intranet. Figure 6.29 shows its structure. getShareFilePath method obtains the set of shared files from IP address. readFromSmb method obtains the content of a file. copyFromSmb method copies the shared file to local for swing use. (refer to Appendix K)

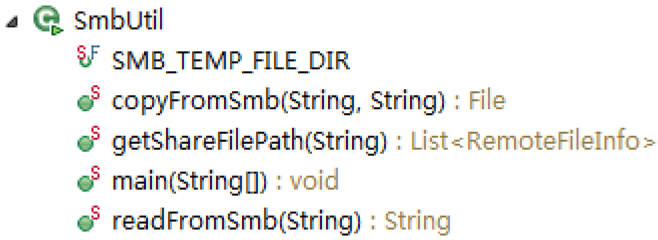


Figure 6.29: the Structure of SmbUtil

6.7 Summary

The testing of the methodology developed is conducted in this chapter. This chapter details the aim and scope of the testing, and the consideration of the testing design. Then it discussed the testing sessions using different searching requirements. It lays out the structure of the various classes in lansearch, and explains those served Lucene.

CHAPTER SEVEN: DISCUSSIONS, CONCLUSIONS AND FURTHER RESEARCH

This chapter presents the conclusions of the research project, and identified areas for further research.

7.1 Discussion of the Contributions and Limitations

The research question of this project is whether the Agent technology could help to improve knowledge search in new product development process in a manufacturing enterprise.

An industrial investigation was conducted in the aviation enterprises. Twelve people were invited to take the questionnaires and semi-structured interviews. For the R&D people, questions are mainly concerned about the required information and knowledge, knowledge sharing, and search tools. For the managers, the questions include the process information and responsibilities. For the archive people, the questions focus on the archive structure and documents transportation.

A review of literature was undertaken. It examined the current understanding of knowledge in the literature, models for knowledge management, and the knowledge-based activities of developing products in manufacturing. Then it considered the technologies for information and knowledge search, and the advantages and shortcomings of them. After that, it reviewed the agent technology for design and manufacturing, and for knowledge search especially.

After the literature review, it is found that Agent technology is used in computer support engineering design, and manufacturing. However, it is not fully developed in knowledge search area. Actually, Agent technology meets well the requirements of search. This project employs Agent for enterprise knowledge search. The Agent wraps legacy software systems, and encapsulates search activities. Therefore, search agent based on Multi-Agent technology utilises indexes built by other search engines instead of moving large amount of data to search engine o create index. And it enables the plug-and-play integration. In the developed system, autonomous Agents track user interests to help filter information. Special types of enterprise search like term definition, employee's personal information, expert on topics are searched.

The knowledge management within an enterprise is hindered by various barriers, like the identifying the knowledge assets, and encouraging knowledge sharing. Two main knowledge management problems emerged from the investigation are the inefficiency in information retrieval from the network, and the lack of knowledge sharing through the network.

In manufacturing enterprise knowledge search domain, a Multi-Agent system developed can contribute in different ways. Instead of moving large amounts of data to the search engine to create indexes, the Agents could get sorted URL information from collaborated search engines. This relieves stresses on web bandwidth. For enterprise search, some special types of search are common, like term definition, homepages of groups or topics, employees' personal information and expert on topics. All the types of searches will be combined together finally. This realises better sharing of tacit knowledge as well as explicit knowledge. In addition, the system adapt to user's individual interests, reduce the amount of information that are not required, and help users to access right knowledge with minimal effort.

Agents are designed and implemented separately, including the Dispatch Agent, the Learning Agent, and the Search Agent. The actions of the Dispatch Agent include creating basic information for user in the database, and noticing the Learning Agent to create new user interest model for new user; obtaining user login information, and sending user information to the Learning Agent; recording reflection information of user; recording user's query, obtaining user interest model from the Learning Agent, optimising user's query, and sending it to the Search Agent and the Learning Agent. The Learning Agent is responsible for creating user interest model by artificial modelling, creating user interest model by historical data, obtaining user interest model according to other Agent's requirement, and sending it to the Agent. The Search Agent is responsible for requesting the user interest model from the Learning Agent, obtaining primary search results by normal search techniques, filtering the results based on the contents, and filtering the results based on the cooperation with the Learning Agent. These three Agents construct the

Multi-Agent knowledge search system. The test of the system includes search of the database, search of shared files, search of the management information system, and search on the web.

The limitation of this project is the lack of evaluation of the system. Tests were conducted to prove that the proposed methodology and system works. However, further industrial investigation is needed to demonstrate that it retrieves the results as the NPD users expect.

7.2 Conclusions of the Research Project

This thesis presents a methodology and system for facilitating knowledge search in the new product development process in a manufacturing context.

A literature review conducted in the scoping phase of the research revealed that effective knowledge search for NPD project team members is important to the success of an NPD project. This assertion is supported by published empirical evidences. It was found that there are numerous barriers to knowledge search in the new product development environment. An exploratory case study was conducted in an aircraft manufacturing company in order to identify key knowledge search barriers and provide further focus for the remainder of the research project. In the early stage, the main purpose and effort devoted to is the understanding of the industrial collaborator's business processes, product range, organisational structure and its AS-IS situation of information and knowledge management. Then a more focused investigation into industrial problems and their requirements for knowledge management and search systems were carried out. In the later stage, examples of actual knowledge have been captured and used for case studies to test the proposed knowledge search methodology.

Indeed, there is a lack of research into the knowledge search in this NPD process in enterprise. To address this issue, a methodology and system was developed to reduce the knowledge search barriers, based on the findings of the exploratory case. The Multi-Agent knowledge search system was initially tested at the end.

In summary, the main achievements of this project have been:

- A further exploration into the nature of knowledge and available approaches of information and knowledge search in new product development process;

- The identification of key barriers to enterprise knowledge search, compared with general information search engines;
- The development of a Multi-Agent based knowledge search methodology which overcomes shortcomings of available search engines in enterprise applications such as limited search speed, unstructured knowledge, and types of knowledge more than documents.
- The implementation of the enterprise search engine to fulfil the methodology, and conduct an initial test of the methodology.

7.3 Further Work

The result of this research investigation has been the provision of a methodology and system for the facilitation of knowledge search in the new product development process in enterprise. The literature review revealed that there is a lack of research into various themes related to enterprise knowledge search. It is proposed that research is required in a number of areas to enrich and develop the research presented in this thesis. Additional testing of the Multi-Agent system is advocated, in order to lend verisimilitude to the findings already presented in this research. In summary, the following further work should be carried out:

- A further implementation is needed to complete the system. For example, the Multi-Agent search function on the web search outside the enterprise requires further work on it. The analysis and approach are developed in chapter five. The module is partly realised and tested in chapter six. But the realisation of collaboration with more search engines requires the analysis of more search server.
- A friendlier user interface needs further effort.
- Test the system with cases from other phases though, such as detail design and production, to provide evidence as to the applicability to the entire product development process.
- Test the system with other product types.
- A validation about to what extent the developed system improves the search performance of new product development teams.
- Whether the benefits of the developed system outweigh the cost of building and maintaining it.

REFERENCES

Ackoff, R.L. 1989. From data to wisdom. Journal of Applies Systems Analysis, 16.

Adams, M.E., Day, G.S. & Dougherty, D. 1998. Enhancing new product development performance: an organizational learning perspective. Jorunal of Product Innovation Management, 15, 403-422.

Alavi, M. & Leidner, D.E. 2001. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. Mis Quarterly, 25, 107-136.

Arbnor, I. & Bjerke, B. 2009. Methodology for creating business knowledge, Los Angeles ; London, SAGE.

Autonomy. 2009. Available: http://www.autonomy.com/. April 2009.

AVIC. 2008. Introduction of the Corporation. Available: http://www.avic.com.cn/index_jtfc.asp. December 2009.

AVIC. 2009. Aviation Industry Corporation of China. Available: http://www.avic.com.cn/. November 2009.

Badaracco, J. 1991. The knowledge link : how firms compete through strategic alliances, Boston, Mass., Harvard Business School Press.

Balasubramanian, S., Maturana, F.P. & Norrie, D.H. 1996. Multi-agent planning and coordination for distributed concurrent engineering. International Journal of Cooperative Information Systems, 5, 153-179.

BASIS. 2010. Available: http://www.basisplus.com/. October 2010.

Bellifemine, F., Poggi, A. & Rimassa, G. 2001. Developing multi-agent systems with a FIPA - compliant agent framework. Software Practice and Experience, 31, 103-128.

Bellinger, G., Castro, D. & Mills, A. 1997. Data, information, knowledge, and wisdom.

Bernon, C., Camps, V., Gleizes, M.-P. & Picard, G. 2005. Engineering adaptive multi-agent systems: the ADELFE methodology. Agent-oriented methodologies. London: Idea Group Publishing.

Blackler, F. 1995. Knowlege, knowledge work and organizations: an overview and interpretation. Organization Studies, 16, 1021-1046.

Blackler, F., Reed, M. & Whitaker, A. 1993. Editorial Introduction - Knowledge Workers and Contemporary Organizations. Journal of Management Studies, 30, 851-862.

Broens, R.C.J.A.M. & Vries, M.J.d. 2003. Classifying technological knowledge for presentation to mechanical engineering designers. Design Studies, 24, 457-471.

Bruccoleri, M., Nigro, G.L., Perrone, G., Renna, P. & Diega, S.N.L. 2005. Production planning in reconfigurable enterprises and reconfigurable production systems. CIRP Annals - Manufacturing Technology, 54, 433-436.

CACC. 2009. Commercial Aircraft Corporation of China. Available: http://www.comac.cc/. November 2009.

Caploe, D. 2009. UK Economy. Economy Watch. Available: http://www.economywatch.com/world_economy/united-kingdom/. June 2010.

Chauhan, D. & Baker, A.D.1998 JAFMAS: a multiagent application development system. In: AGENT '98 Proceedings of the second international conference on Autonomous agents.

CLARITECH. 2010. Available: http://www.claritech.com/. November 2010.

Collins, H.M. 1993. The Structure of Knowledge. Social Research, 60, 95-116.

Commercial Aircraft Corporation of China, L. 2009. Aviation Industry Corporation of China. Available: http://www.comac.cc/n16/n1154/n15744/n15986/index.html. October 2009.

Commercial Aircraft Corporation of China, L. 2009. Introduction of Corporation. Available: http://www.comac.cc/n16/n1154/n1244/index.html. October, 2009.

Cooper, R.G. 1994. Perspective third-generation new product processes. Journal of Product Innovation Management, 11, 3-14.

Crow, K. 1992. Computer-aided process planning. CA, USA: DRM Associates.

Cuadra. 2010. Available: http://www.cuadra.com/. November 2010.

Cutkosky, M.R., Engelmore, R.S., Fikes, R.E., Genesereth, M.R., Gruber, T.R., Mark, W.S., Tenenbaum, J.M. & Weber, J.C. 1993. PACT: an experiment in integrating concurrent engineering systems. IEEE Computer Society, 26, 28-37.

Darke, P., Shanks, G. & Broadbent, M. 1998. Successfully completing case study research: combining rigour, relevance and pragmatism. Information Systems Journal, 8, 273-289.

Dataflight. 2010. Dataflight Software, Inc. Available: http://law.lexisnexis.com/concordance. November2010.

Davenport, T.H. & Prusak, L. 1998. Working knowledge : how organizations manage what they know, Boston, Mass, Harvard Business School Press.

Drucker, P.F. 1993. Post capitalist society, New York, HarperBusiness.

Duffie, N.A. 1990. Synthesis of heterarchical manufacturing systems. Computers in Industry, 14, 167-174.

Edwards, P.K. 2003. Industrial relations : theory and practice, Malden, MA, Blackwell Publishing.

Excalibur. 2010. Available: http://excalibur.sourceforge.net/. August 2010.

Fahey, L. & Prusak, L. 1998. The eleven deadliest sins of knowledge management. California Management Review, 40, 265-276.

FAI. 2007. AVIC The First Aircraft Institute. Available: http://www.avic1-fai.com/Ywbm/aboutus.asp. October 2009.

Ferber, J. 1999. Multi-agent systems: an introduction of distributed artifical intelligent, Pearson Education Limited.

Franklin, S. & Graesser, A. 1997. Is it an agent, or just a program?: a taxonomy for autonomous agents. Lecture Notes in Computer Science, 21-36.

Fredericks, E. 2005. Cross-functional development in new product development: A resource dependency and human capital perspective. Qualitative Market Research: An International Journal, 8, 327-341.

Fulcrum 2000. Fulcrum SearchServer 4.0 Technical Overview. Hummingbird Ltd.

Gates, B. & Hemingway, C. 1999. Business @ the speed of thought : using a digital nervous system, New York, NY, Warner Books.

Grefenstette, G. 1997. Short query linguistic expansion techniques: Palliating one-word queries by providing intermediate structure to text. Computer Science, 1299/1997, 97-114.

Griffin, A. 1997. PDMA research on new product development practices: Updating trends and benchmarking best practices. Journal of Product Innovation Management, 14, 429-458.

Grundspenkis, J. 2007. Agent based approach for organization and personal knowledge modelling: knowledge management perspective. Jorunal of Intelligent Manufacturing, 18, 451-457.

Gürsel, A. & Sen, S.2009 Improving search in social networks by agent based mining. In: the Twenty-First International Joint Conference on Artificial Intelligence, Pasadena, California, USA. 2034-2039.

Halavais, A.M.C. 2009. Search engine society, Cambridge, U.K. ; Malden, Mass., Polity.

Hart, S.J. 1996. New product development : a reader, London, Dryden.

Hassard, J. & Kelemen, M. 2002. Production and consumption in organizational knowledge: the case of the 'paradigms debate'. Organization, 9, 331-355.

Hatvany, J. 1985. Intelligence and cooperation in heterarchic manufacturing systems. Robotics and Computer-Integrated Manufacturing, 2, 101-104.

Hutton, J. & Denham, J. 2008. Manufacturing: New Challenges, New Opportunities. In: Department for Business Enterprise & Regulatory Reform.

Ilgen, M.R. & Rushall, D.A.1996 Recent advances in HNC's context vector information retrieval technology. In: TIPSTER '96, Vienna, Virginia. Association for Computational Linguistics Stroudsburg, 149-158.

Introna, L.D. & Nissenbaum, H. 2000. Shaping the web: why the politics of search engines matters. Information Society, 16, 169-186.

Irani, Z., Ezingeard, J.-N., Grieve, R.J. & Race, P. 1999. A case study strategy as part of an information systems research methodology: a critique. International Journal of Computer Applications in Technology, 12, 190-198.

Jansen, B.J. 2005. Seeking and Implementing Automated Assistance During the Search Process. Information Processing & Management, 41, pp.909-928.

Jansen, J. 1997. Using an intelligent agent to enhance search engine performance. First Monday, 2.

Jashapara, A. 2004. Knowledge management: an integrated approach, Pearson Education Limited.

Jazayeri, M. & Luqmayr, W.A.2000 Gypsy: a component-based mobile agent system. In: Parallel and Distributed Processing, Rhodos, Greece. 126-134.

Jennings, N. & Wooldridge, M.J. 1998. Agent technology : foundations, applications, and markets, Berlin ; New York, Springer-Verlag.

Jini. 2010. Available: http://www.sun.com/jini/. December 2010.

Kalpakjian, S. 1991. Manufacturing Processes for Engineering Materials, Addison-Wesley.

Kannebley, S., Sekkel, J.V. & Araujo, B.C. 2010. Economic performance of Brazilian manufacturing firms: a counterfactual analysis of innovation impacts. Small Business Economics, 34, 339-353.

Kao, M., Cercone, N. & Luk, W.-S. 1998. Providing quality responses with natural language interfaces: the null value problem. Software Engineering, IEEE Transactions on, 14, 959-984.

Kashyap, V. & Sheth, A.1994 Semantics-based information brokering. In: ADAM, N. R., BHARGAVA, B. K. & YESHA, Y., eds. Conference on Information and Knowledge Management. ACM New York, 363-370.

Keane, B.T. & Mason, R.M.2006 On the nature of knowledge: rethinking popular assumptions In: Proceedings of the 39th annual Hawaii International Conference on System Sciences, Hawaii. 162b.

Khare, R., Cutting, D., Sitaker, K. & Rifkin, A. 2004. Nutch: a flexible and scalable open-source web search engine. Technical Report CN-TR-04-04. CommerceNet Labs.

Kleinberg, J.M. 1999. Hubs, authorities, and communities. ACM Computing Survey, 31.

Klusch, M. 2001. Information agent technology for the Internet: a survey. Data & Knowledge Engineering, 36, 337-372.

Kogut, B. & Zander, U. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. Organization Science, 3, 383-397.

Lawrence, S. & Giles, C.L. 1999. Accessibility of information on the web. NATURE, 107-110.

Lee, Y.-H., Kumara, S.R.T. & Chatterjee, K. 2003. Multiagent based dynamic resource scheduling for distributed multiple projects using a market mechanism Jorunal of Intelligent Manufacturing, 14, 471-484.

Luck, M. 2005. Agent technology : computing as interaction : a roadmap for agent based computing. Southampton: University of Southampton on behalf of AgentLink III.

Lynn, G.S., Reilly, R.R. & Akgun, A.E. 2000. Knowledge management in new product teams: practices and outcomes. IEEE Transactions on Engineering Management, 47, 221-231.

Lyotard, J.F. 1984. The postmodern condition : a report on knowledge, Minneapolis, University of Minnesota Press.

Maedche, A. & Staab, S. 2001. Ontology learning for the Semantic Web. Ieee Intelligent Systems & Their Applications, 16, 72-79.

Malhotra, Y. 2001. Knowledge management and business model innovation, Hershey PA, Idea Group Pub.

Marwick, A.D. 2001. Knowledge management technology. IBM Systems Journal, 40, 814-830.

Menczer, F., Degeratu, M. & Street, W.N. 2000. Efficient and scalable pareto optimization by evolutionary local selection algorithms. Evolutionary Computation, 8, 223-247.

Mertins, K., Heisig, P. & Vorbeck, J. 2001. Knowledge management : best practices in Europe, Berlin ; New York, Springer.

Monostori, L., Váncza, J. & Kumara, S.R.T. 2006. Agent-based systems for manufacturing. ANNALS - CIRP, 55, 697-720.

Moukas, A. & Maes, P. 1998. Amalthaea: an evolving multi-agent information filtering and discovery system for the WWW. Autonomous Agents and Multi-Agent Systems, 1.

Mukherjee, R. & Mao, J. 2004. Enterprise search: tough stuff. Queue, 2.

Nelson, P. 1994. The ConQuest system. NIST Special Publication Sp, 265.

Nelson, R.R. & Winter, S.G. 1982. An evolutionary theory of economic change, Cambridge, Mass., Belknap Press of Harvard University Press.

Newell, S., Maxine, R., Scarbrough, H. & Swan, J. 2002. Managing knowledge work, Houndsmillls, Basingstoke, Hampshire ; New York, Palgrave.

Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. Organization Science, 5, 14-37.

Nonaka, I. & Takeuchi, H. 1995. The knowledge-creating company : how Japanese companies create the dynamics of innovation, New York ; Oxford, Oxford University Press.

Nonaka, I. & Toyama, R. 2003. The knowledge-creating theory revisited: knowledge creation as a synthesizing process. Knowledge Management Research & Practice, 1, 2-10.

Nonaka, I., Toyama, R. & Konno, N. 2000. SECI, Ba and leadership: a unified model of dynamic knowledge creation. Long Range Planning, 33, 5-34.

Notess, G.R. 2006. Search engine showdown. Available: http://www.searchengineshowdown.com/features/google/. May 2010.

Odubiyi, J.B., Kocur, D.J., Weinstein, S.M., Wakim, N., Srivastava, S., Gokey, C. & Graham, J. 1997 SAIRE - a scalable agent-based information retrieval engine. In: the First International Conference on Autonomous agents, Marina del Rey, CA. 292-299.

Odyssey. 2010. Odyssey Development, Inc. Available: http://www.isys-search.com/. February 2011.

Park, H., Cutkosky, M.R., Conru, A.B. & Lee, S.-H. 1994. An agent-based approach to concurrent cable harness design. Artificial Intelligence for Engineering Design Analysis and Manufacturing, 8, 45-61.

Parunak, H.V.D., Sauter, J. & Fleischer, M. 1999. The RAPPID project: symbiosis between industrial requirements and MAS research. Autonomous Agents and Multi-Agent Systems, 2, 111-140.

Polanyi, M. 1967. The tacit dimension, London,, Routledge & K. Paul.

Pollard, D. 2008. PKM: A bottom-up approach to knowledge management. In: SRIKANTAIAH, T. K. & KOENIG, M. E. D. (eds.) Knowledge Management in Practice. Information Today, Inc.

Prusak, L. 2001. Where did knowledge management come from? IBM Systems Journal, 40.

Quigley, E.J. & Debons, A.1999 Interrogative theory of information and knowledge. In: 1999 ACM SIGCPR conference on Computer personnel research, New Orleans, LA.: ACM Press, 4-10.

Rhodes, B.J. 1997. The Wearable Remembrance Agent: a system for augmented memory. Personal Technologies: Special Issue on Wearable Computing, 1, 218-224.

Rhodes, B.J.2000 Margin notes: building a contextually aware associative memory. In: the 5th International Conference on Intelligent User Interfaces. 219-224.

Rhodes, B.J. & Starner, T.1996 Remembrance agent: a continuously running automated information retrieval system. In: the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, London. 487-496.

Říha, A., Pěchouček, M., Vokřínek, J. & Mařík, V. 2002. ExPlanTech: exploitation of agent-based technology in production planning Lecture Notes in Computer Science, 308-322.

Ryle, G. 1949. The concept of mind, London, Hutchinson & Co. Ltd.

Salomon, G. 1993. No distribution without individuals' cognition: a dynamic interactional view. In: SALOMON, G. (ed.) Distributed cognitions: psychological and educational considerations. Cambridge University Press.

Scarbrough, H. & Swan, J. 2001. Explaining the diffusion of knowledge management: The role of fashion. British Journal of Management, 12, 3-12.

Shen, W., Hao, Q., Yoon, H.J. & Norrie, D.H. 2006. Applications of agent-based systems in intelligent manufacturing: an updated review. Advanced Engineering Informatics, 20, 415-431.

Silberschatz, A., Stonebraker, M. & Ullman, J. 1996. Database research: achievements and opportunities into the 21st century. Sigmod Record, 25, 52-63.

Solskinnsbakk, G. & Gulla, J.A. 2008. Ontological profiles in enterprise search. Knowledge Engineering: Practice and Patterns, 5268/2008, 302-317.

Spek, R.v.d. & Spijkervet, A. 1999. Knowledge management: dealing intelligently with knowledge 3rd ed.: Utrecht: Kenniscentrum CIBIT.

Stata, R. 1989. Organizational learning - the key to management innovation Sloan Management Review, Spring, 63-74.

Sullivan, D. 2002. How Search Engines Work. Search Engine Watch. Available: http://searchenginewatch.com/2149051. March 2009.

Sun, J., Zhang, Y.F. & Nee, A.Y.C. 2001. A distributed multi-agent environment for product design and manufacturing planning. International Journal of Production Research, 39, 625-645.

Swan, J., Scarbrough, H. & Preston, J. 1999. Knowledge management - the next fad to forget people? 7th European Conference on Information Systems. Copenhagen.

Teevan, J., Alvarado, C., Ackerman, M.S. & Karger, D.R.2004 The perfect search engine is not enough: a study of orienteering behavior in directed search. In: Proceedings of ACM CHI Conference on Human Factors in Computing Systems. 415-422.

The-Manufacturing-Institute. 2009. The Facts about Modern Manufacturing. Available: http://www.nam.org/~/media/0F91A0FBEA1847D087E719EAAB4D4AD8.ashx. September 2008.

TheNielsenCompany. 2010. Nielsen reports March 2010 U.S. search rankings. Available: http://blog.nielsen.com/nielsenwire/online_mobile/nielsen-reports-march-2010-u-s-search-rankings/. March 2011.

Theobald, A. & Weikum, G. 2002. The Index-Based XXL Search Engine for Querying XML Data with Relevance Ranking. Lecture Notes in Computer Science. Springer Berlin / Heidelberg.

Toye, G., Cutkosky, M.R., Leifer, L.J., Tenenbaum, J.M. & Glicksman, J.1993 SHARE: a methodology and environment for collaborative product development. In: Enabling

Technologies: Infrastructure for Collaborative Enterprises, Morgantown, WV, USA. 33-47.

Trott, P. 2008. Innovation management and new product development, New York, Financial Times Prentice Hall.

Tsoukas, H. & Vladimirou, E. 2001. What is organizational knowledge? Journal of Management Studies, 38, 973-993.

Ughanwa, D.O. & Baker, M.J. 1989. The role of design in international competitiveness, London ; New York, Routledge.

Ulrich, K.T. & Eppinger, S.D. 1995. Product design and development, New York, McGraw-Hill.

Vise, D.A. & Malseed, M. 2005. The Google Story, Macmillan.

Vonkrogh, G. 1998. Care in Knowledge Creation. California Management Review, 40, 133-154.

Wang, W. 2007. Introduction of the Management Mode of a Large Aircraft Manufacturing Stock Company.

Ward, M. 2006. How the web went world wide. Technology Correspondent, BBC News.

Wick, C. 2000. Knowledge management and leadership opportunities for technical communicators. Technical Communication - Washington, 47, 15.

Wright, K. 2005. Personal knowledge management: supporting individual knowledge worker performance. Knowledge Management Research & Practice, 3, 156-165.

Yan, B. 2010. AVIC Annual Turnover is 191billion RMB. People's Daily. Available: http://finance.people.com.cn/GB/10809745.html. January 2010.

Yan, R., Hauptmann, A. & Jin, R. 2003. Multimedia search with pseudo-relevance feedback. Lecture Notes in Computer Science, 238-247.

Yin, R.K. 2003. Applications of case study research, Thousand Oaks ; London, Sage Publications.

Zahay, D., Griffin, A. & Fredericks, E. 2004. Sources, uses, and forms of data in the new product development process Industrial Marketing Management, 33, 657-666.

Zhao, F.L., Tso, S.K. & Wu, P.S.Y. 2000. A cooperative agent modelling approach for process planning. Computer in Industry, 41, 83-97.

APPENDICES

Appendix A: Questions for R&D Department in Vaillant Group

Knowledge Sharing Investigation

Interview Protocol

R&D Questions

This part is the questions in the questionnaire which is designed for R&D people in Vaillant Group.

Interviewee:

Role:

Questions

1. What is your understanding of the term ‘knowledge’?

Information or Knowledge Needs

Note to interviewer: Please inform the interviewee that any reference to “you” in isolation, means “you and your project team”, rather than “you personally”.

2. What kind of information/knowledge do you and your project team need in the course of a project?

(For example: Reports, expert advice etc.)

3. What is the format of this knowledge?

4. (a) Where do you and members of your project team look for knowledge?

(b) Do you and your project team search for knowledge just locally (at your site), or is the search extended to the whole Group’s knowledge, or even outside of the Vaillant Group?

5. (a) How do you and your project team search for the knowledge?

(b) Do you and your project team use tools to find knowledge? If so, what are the tools, and can you give real examples of tools and systems that support the search?

(c) Do you feel the tools could be improved? If so, how could the tools be improved?

(d) If no tools are used, what methods of searching for knowledge do you and your project team use and how could they be improved?

6. What (kind of) knowledge do you and your project team have problems finding?

7. What is the minimum knowledge that you and your project team expect to be able to find?

8. In the case of knowledge sought on the company's network (e.g. Project drive), is there any knowledge that you feel your project team needs, but has difficulty accessing?

9. What frustrates you and you project team about searching for knowledge?

10. Is there any knowledge that you consider is missing (e.g. information that would assist you in making a decision, but is unavailable)? If so, what is it?

11.Could you explain why this knowledge is unavailable?

Knowledge Sharing

12 (a) What kind of information or knowledge do you and your project team generate?

(b) Do you and your team currently collect and organize this information somewhere? If so, where?

13. Is this knowledge reusable in other projects or in other processes within the IP?

14. (a) What kind of information would your team be prepared to contribute to an shared R&D knowledge base:

- Now?
- In the future?

b) How and when could you and your project team prepare this shared information? What storage formats and what kind of support (e.g. tools, systems, methods) would you recommend?

c) Should this information or knowledge be shared among the different R&D sites?

d) How important is sharing this knowledge among the sites? Would you describe it as mandatory?

Use of the Intranet as a Search Tool

15. Do you or your project team use the Intranet to search for knowledge?

16. Is there any functionality that you feel could improve the process of searching for knowledge on the Intranet?

Appendix B: Answers from R&D Department in Vaillant Group

R&D Integration Study

R&D Questions

This part is the answers of the questionnaire which is designed for R&D people in Vaillant Group.

Interviewee: [deleted]

Role: NPD Project Manager

Questions

1. What is your understanding of the term 'knowledge'?

"Knowledge. I would say was the record of learnings or experiences. Whether that's mentally recording or physically recording, or…"

Information or Knowledge Needs

Note to interviewer: Please inform the interviewee that any reference to "you" in isolation, means "you and your project team", rather than "you personally".

2. What kind of information/knowledge do you and your project team need in the course of a project?

(For example: Reports, expert advice etc.)

"The biggest bit of knowledge that we would need is, a knowledge of, it sounds stupid, but a knowledge of what knowledge there already is. Because nine times out of ten, when we start something,, you get half way through it thinking that you're starting from scratch, and then you find out that somebody's already done it somewhere else six months ago, but nobody knows about it.

"It's what do we know that we know. It sounds stupid, but that is the one big thing that would help. A lot of the things, I mean you know about our [NPD Business Process]? A lot of the problems we have here certainly is knowing what's required for it. We've got this huge system … process that's all broken down, but when you read it, it still doesn't tell you what you need to do for it."

3. What is the format of this knowledge?

"Reports. Historically we are very bad at recording information. A lot of what we do is still based on experience. Properly formatted reports are getting more popular. Suppliers websites for information about materials. Test laboratory in Remscheid. Again, nobody knew it was there, the big materials testing laboratory in Remscheid. A lot of documented history... I've got two contacts now in the materials lab. And also with their life testing over there... But again it's just a case of knowing what's available. And that's the biggest thing I've come up with."

4. (a) Where do you and members of your project team look for knowledge?

"Up until recently it would have been local servers, internet, generally asking other people in the department. Probably in the last six months, certainly in the last three months, we've started using the Group servers. A materials database on the servers. We still use the Internet... Drawing Office for design and spec [specification] and materials specs and suitability. Lab [Laboratory] for test standards, test requirements and results from previous testing, just for comparison. Paper in filing cabinets, but this is no good if you don't know where to look."

(b) Do you and your project team search for knowledge just locally (at your site), or is the search extended to the whole Group’s knowledge, or even outside of the Vaillant Group?

"More and more it is the group, especially for myself."

5. (a) How do you and your project team search for the knowledge?

"There is a huge problem with translation for intergroup knowledge.
We use a lot of test specifications now for life testing that were written by Remscheid, We’ve adopted their test planning documents, we’ve adopted their specifications for this, but again, a lot of it is still in German and we’re not good at translating, there’s and awful lot to translate.

“I think knowledge sharing across the group is a brilliant idea because there is so much out there, but your biggest problem is gonna be language. Language was the biggest problem we had with IP [NPD business process] in the first place. It was all set up, it was all done in German. And the reason it didn’t get adopted properly in the UK and implemented and driven through the business is because bits were translated into Pidgin English, it just didn’t make sense and it was too difficult to work your way through.”

(b) Do you and your project team use tools to find knowledge? If so, what are the tools, and can you give real examples of tools and systems that support the search?

"Windows Explorer"

(c) Do you feel the tools could be improved? If so, how could the tools be improved?

"Retrieving knowledge is very slow. Unless you know exactly what you’re looking for and where it is, you’re gonna be looking for a while."

(d) If no tools are used, what methods of searching for knowledge do you and your project team use and how could they be improved?

Other methods:

6. What (kind of) knowledge do you and your project team have problems finding?

“The knowledge we have problems finding? It’s the same as I said at the start. It’s trying to find out what’s available. Once you’ve found out what’s available, you can generally find it. But it’s finding out that there is something there to find. It might sound silly, but…”

"The new format for specifications is dual language. It’s written in German and English. It’s written paragraph by paragraph in two languages. But the problem we have… it’s easier, the Germans will write that, and they’ll write it in German and English, but when it comes to us, we’ll ignore the German part and just write the English part because we don’t have the language skills to do it. We can’t translate back. That’s something that is a problem."

7. What is the minimum knowledge that you and your project team expect to be able to find?

"Test reports. That's field trials, combustion testing, validation testing, life tests, function and wear tests, endurance tests..."

8. In the case of knowledge sought on the company's network (e.g. Project drive), is there any knowledge that you feel your project team needs, but has difficulty accessing?

"Access rights... physically connecting to the server [in Germany]"

9. What frustrates you and you project team about searching for knowledge?

“What frustrates us? What certainly frustrates me is when you spend hours and hours trying to find the document that you want, and then you get it and it’s in German. It’s as good as not having it.”

10. Is there any knowledge that you consider is missing (e.g. information that would assist you in making a decision, but is unavailable)? If so, what is it?

"Not really. All the project managers in the UK are R&D "

11. Could you explain why this knowledge is unavailable?

See answer to question 10.

Knowledge Sharing

12 (a) What kind of information or knowledge do you and your project team generate?

"Drawings. Design side you’ve got drawings and specifications. Lab side you’ve got the whole host of all your approvals testing, all your development work, all your efficiencies, combustions, temperature. All this endless everything that you’ve got to develop to meet a CE approved appliance at the right level that you want it. Most things are electronic for things like that”

(b) Do you and your team currently collect and organize this information somewhere? If so, where?
"Radocs server local to UK, Group server – generally just used for IP and documentation for milestones. Drawings are from ProE and are in Intralink viewable via ProductView."

13. Is this knowledge reusable in other projects or in other processes within the IP?

No answer recorded to this question.

14. (a) What kind of information would your team be prepared to contribute to an shared R&D knowledge base:

No answer recorded to this question.

b) How and when could you and your project team prepare this shared information? What storage formats and what kind of support (e.g. tools, systems, methods) would you recommend?

No answer recorded to this question.

c) Should this information or knowledge be shared among the different R&D sites?

"I think we should share as much as we can."

d) How important is sharing this knowledge among the sites? Would you describe it as mandatory?

See answer to 14(b).

Use of the Intranet as a Search Tool

15. Do you or your project team use the Intranet to search for knowledge?

"Very rarely. For IP [NPD business process] stuff."

16. Is there any functionality that you feel could improve the process of searching for knowledge on the Intranet?

No response was recorded for this question.

R&D Integration Study

Interview Protocol

R&D Questions

Interviewee: [deleted]

Role: Project Manager

Date: February 2006

Questions

General

1. What is your understanding of the term ‘knowledge’?

“From my point of view, knowledge means: the skill of a person to be able to “do something”. The information you need in order to do your task.”

Information or Knowledge Needs

Note to interviewer: Please inform the interviewee that any reference to “you” in isolation, means “you and your project team”, rather than “you personally”.

2. What kind of information/knowledge do you and your project team need in the course of a project?
(For example: Reports, expert advice etc.)

“To be sure of the client needs (internal or external). When the needs change the project leader has to inform the team properly about the change, why the change, and what are the consequences. The needs and the technical requirements are described in:
- Marketing technical specifications
- Powerpoint files
- E- mails.”

3. What is the format of this knowledge?

The interviewee provided no response to this question.

(a) Where do you and members of your project team look for knowledge?

“- In the project folder.
In the project panel hanging in the project meeting room.”

(b) Do you and your project team search for knowledge just locally (at your site), or is the search extended to the whole Group’s knowledge, or even outside of the Vaillant Group?

“Depends on the projects. In Nantes, 75% (in accordance with my estimation) of the knowledge is in Nantes.”

(a) How do you and your project team search for the knowledge?

“In Nantes: We speak with the best person in the plant who (from our opinion) should have the best information or must know the person who knows. External to Nantes: Same, even it’s more difficult (our company is not too big).”

(b) Do you and your project team use tools to find knowledge? If so, what are the tools, and can you give real examples of tools and systems that support the search?

“No.”

(c) Do you feel the tools could be improved? If so, how could the tools be improved?

“Yes, a simple tool can be imagined. A real location for sharing data (like the common drive), but reliable, simple to use, with a standard skeleton for all the project in order to be sure that a team member will find the information in another project folder.
The current common drive is not a good solution because is not user-friendly enough.
For instance I (with the help of the IT support) have tried to copy my project folder since December on the drive without success!!!”

(d) If no tools are used, what methods of searching for knowledge do you and your project team use and how could they be improved?

“See above.”

6. What (kind of) knowledge do you and your project team have problems finding?

“All kinds of information, technical, financial etc.”

7. What is the minimum knowledge that you and your project team expect to be able to find?

“It depends…”

8. In the case of knowledge sought on the company's network (e.g. Project drive), is there any knowledge that you feel your project team needs, but has difficulty accessing?

“In the case of our company size, very often men will never replace the IT tools. We can of course, improve the tool (see above) but it’s so simple to phone one of our colleagues (in Nantes, UK or Germany). If the colleague doesn’t have the information, may be he can help me and direct me to the right person. Next time it will be his turn to help one of our colleagues. More and more the European team builds itself.”

9. What frustrates you and your project team about searching for knowledge?

“Sometime when the IT tool already exist he doesn’t work and when the tool exists is too complicated and nobody wants to use it. (See MS project server implementation 3 years ago)
Simple is beautiful and useful!”

10. Is there any knowledge that you consider is missing (e.g. information that would assist you in making a decision, but is unavailable)? If so, what is it?

“All kind of information can be missing.”

11.Could you explain why this knowledge is unavailable?

“Very often we haven’t enough standard documents and template. Each project or site reinvents the wheel! Defining standard templates for each task is also a priority for us. When the common templates are defined and used it will be easier to find them on the IT drives.
With a software research tool or without.”

Knowledge Sharing

12 (a) What kind of information or knowledge do you and your project team generate?

“Status reports, Financial reports, pictures etc.”

(b) Do you and your team currently collect and organise this information somewhere? If so, where?

“Yes of course, each team member uses the project folder. The knowledge is shared through the folder and the weekly team meeting.”

13. Is this knowledge reusable in other projects or in other processes within the IP?

“Yes.”

14. (a) What kind of information would your team be prepared to contribute to an shared R&D knowledge base:
Now?

“We have defined a common template for MSP used by the 3 project managers in Nantes. One of us started to build it, a second improved it.”

In the future?

“I would like to improve the project WEB base for the project development. It’s a kind of workflow for the projects which are not on SAP. Currently this tool works at 50% of is capacity. We are improving it. This tool can be used in each plant and can help people to share information of the parts “under development”. I invite you to Nantes, if your are interested in testing it!”

b) How and when could you and your project team prepare this shared information? What storage formats and what kind of support (e.g. tools, systems, methods) would you recommend?

“See above.”

c) Should this information or knowledge be shared among the different R&D sites?

“Of course.”

d) How important is sharing this knowledge among the sites? Would you describe it as mandatory?

The interviewee provided no response to this question.

Use of the Intranet as a Search Tool

15. Do you or your project team use the Intranet to search for knowledge?

“Yes.”

16. Is there any functionality that you feel could improve the process of searching for knowledge on the Intranet?

“- A common template for project management.
- A common skeleton for the project organisation in the folders.
- Real common drive for sharing project folders.
- A simple tool for information searching on the intranet and the common drive (like Google)
- Simple workflow for parts “under development”. The need is to be able to share information between the project team (In Nantes or between two sites) before the parts are added to SAP.”

R&D Integration Study

Interview Protocol

R&D Questions

Interviewee: [deleted]

Role: Project Manager, Controls Expertise

Questions

1. What is your understanding of the term 'knowledge'?

"Knowledge is mainly high value information combined especially with experience, context, reflection, based on 5 would call it on the one hand trial and error, on the other hand learning from tests experiences and so on; Scientific methods."

Information or Knowledge Needs

Note to interviewer: Please inform the interviewee that any reference to "you" in isolation, means "you and your project team", rather than "you personally".

2. What kind of information/knowledge do you and your project team need in the course of a project?

(For example: Reports, expert advice etc.)

"Future trends, for example, future technologies, test reports, test plannings, erm, test specifications, FMEA, patents, I think these are the most important ones."

3. What is the format of this knowledge?

"So in general we start every-time with advice from colleagues, so we start to organise meetings, the colleagues are defined only by, ja I think the, know the people very well, you know the background, you know their performance, their knowledge performance, so we organise meetings. The next point is the Internet, it's clear, or through Intranet, Patent gate, where you can make patent research, for example, literature , this is also clear, and don't forget the external suppliers."

4. (a) Where do you and members of your project team look for knowledge?

"Mainly visiting fairs or something else. We take our existing systems supplier or preferred supplier and on top we try to find new suppliers or new ideas mainly on fairs or on the Internet or something else and then we call down the supplier and make a meeting together."

(b) Do you and your project team search for knowledge just locally (at your site), or is the search extended to the whole Group's knowledge, or even outside of the

Vaillant Group?

"To be honest, from group point of view, from the internal view it's mostly locally. Ok, suppliers and so on, this is defined. There is also Universities, but it's a very small part."

5. (a) How do you and your project team search for the knowledge?

"Meetings, personal contact with colleagues visits, the Internet."

(b) Do you and your project team use tools to find knowledge? If so, what are the tools, and can you give real examples of tools and systems that support the search?

"Mainly the Internet. Really only a small part [comes from project drives]. Very often you ask the colleagues directly to meet each other and to discuss about the information or the knowledge of the project."

(c) Do you feel the tools could be improved? If so, how could the tools be improved?

"What really should be improved is the searching for patents. Because I had some software here to search in patents to different objects, but this is really not very comfortable."

(d) If no tools are used, what methods of searching for knowledge do you and your project team use and how could they be improved?

"Meetings... Custom search tools. Very difficult to search for a special object."

6. What (kind of) knowledge do you and your project team have problems finding?

"Patents, but also the FMEAs. To find the relevant person, to be sure that they have the right knowledge for your project, your question, so this is also a big problem here. But this is only based on your network. You have to know the people. If you don't know the people..."

7. What is the minimum knowledge that you and your project team expect to be able to find?

"All FMEAs and all relevant patents. "

8. In the case of knowledge sought on the company's network (e.g. Project drive), is there any knowledge that you feel your project team needs, but has difficulty accessing?

"And also access for my team members placed in Nantes to this R&D drive, because the performance is very, very bad, and that's the main reason why they don't use this common R&D drive."

9. What frustrates you and you project team about searching for knowledge?

"APIS! Slow drive, again Patents."

10. Is there any knowledge that you consider is missing (e.g. information that would assist you in making a decision, but is unavailable)? If so, what is it?

"One of the problems, is if we start to discuss, for example, new technologies, or something else, the next question is always what are the performance costs behind of this future technology and there we have a big problem to make the calculation based on our heating business for such new technology. For example, wireless LAN or something else. The manufacturing costs and so on. Not the cost to develop it."

11. Could you explain why this knowledge is unavailable?

"If I take my example of wireless LAN, so it's our business totally different and we cannot use wireless LAN with our past protocols and so on. So, the changes we have to do in front to be able to use wireless LAN are very big and also the quantities behind are very low."

"The luck [indistinguishable] is where you can feel free to develop without looking on the cost, but normally it's in parallel to an integration project, So you start with a pre-developed [unit] at the beginning of the innovation project and they ask directly due to the business plan what are the costs and this is every-time the problem and this is a big risk to take the wrong decision."

Knowledge Sharing

12 (a) What kind of information or knowledge do you and your project team generate?

"Mainly the FMEA from our side. The specification for standard modules by communication or something else, which is used all over the group. We update special milestone checklists which are part of the IP [NPD business process]."

(b) Do you and your team currently collect and organize this information somewhere? If so, where?

"The R&D drive. Mainly in the project folder."

13. Is this knowledge reusable in other projects or in other processes within the IP?

"Yes, if they find this information, yes."

14. (a) What kind of information would your team be prepared to contribute to an shared R&D knowledge base:

- Now?
- In the future?

"So what we do now is to store our FMEAs to define checklists for the IP, to define test rules for hardware components , to define test rules for software components, and in the future, I don't know. Difficult."

b) How and when could you and your project team prepare this shared information? What storage formats and what kind of support (e.g. tools, systems, methods) would you recommend?

No answer recorded for this question.

c) Should this information or knowledge be shared among the different R&D sites?

"Yes. Yes!"

d) How important is sharing this knowledge among the sites? Would you describe it as mandatory?

"It's a must. I see no difference between the sites."

Use of the Intranet as a Search Tool

15. Do you or your project team use the Intranet to search for knowledge?

"Yes, especially for the IP. Also methods like House of Quality, FMEA and so on."

16. Is there any functionality that you feel could improve the process of searching for knowledge on the Intranet?

"One common Intranet for all the brands, English language, because sometimes we are searching also for very easy information about functionalities of applications which are only sold by Saunier Duval, Glow Worm and something else. Ok and it's very difficult to go then on a SD [Saunier Duval] Intranet page and to search inside this page, it's any unpossible."

R&D Integration Study

Interview Protocol

R&D Questions

Interviewee: [name deleted]

Role: Programme Manager, Hydraulics and Accessories

Questions

1. What is your understanding of the term 'knowledge'?

"Knowledge is at the end all informations [sic] I could need for doing the job and it could be test reports, it could be documentations [sic] about positions, it could be business plans, business figures, it could be ideas for new developments to every kind of knowledge what I need in the end for the job."

Information or Knowledge Needs

Note to interviewer: Please inform the interviewee that any reference to "you" in isolation, means "you and your project team", rather than "you personally".

2. What kind of information/knowledge do you and your project team need in the course of a project?

(For example: Reports, expert advice etc.)

"It general you could say, all the documents what we have in the IP [NPD business process], yeah, more or less. But some of them are more important and some are less important. I would say the main important things, data from the markets, for example a system product running in the market, how many parts do we sell, what would was in the past, what will be in the future so we have a tracking curve, or could make a tracking curve about this. Prices and all stuff, so that they could evaluate the business very clear. Competitor informations [sic] are useful, so, especially main competitors, what are the prices, about price positioning, markets, what are the advantages and disadvantages of the product, all the things what we get from competitor analysis.

"To evaluate a new product against a competitor's, for example. What we make as well and what we have available in the Intranet is competitor analysis. In general, we should have access to all kind of specification and to all kinds of contracts and test reports. So, these are the main things that you get during development and during negotiations. Prices should be clearly available, so prices from parts what you buy, calculations, this is done by SAP. I'm not so really interested in general about decision documentation. This could change from time to time. And, anyway, two years later all the parameters has [sic] changed. And on the other side I would also say, we have to balance the things what [sic] we have to document because it's necessary. And do we really need it, do we have really to sit people down and to spend the time making these things.

“In general, due to the fact that we could have sooner or later problems with suppliers, let's say quality issues for example, then we need to document which [sic] we had exchange with the suppliers. So we should have a link where we could place these documents, only the e-mails, and store it. In case, if there is a problem in the future, we could have access, but this is enough at the end. It's useful when you have a special hierarchy in the document store, but it's not in any way necessary. Principal is that we can get access five to ten years later if we have a problem.”

3. What is the format of this knowledge?

“Well, normally, I think the most common way to send out informations [sic] is e-mail, in the company. If you open the e-mails you have there Excel, PowerPoint, Win-Word. I think it's nothing particular, it should only be a storage where you can find all the e-mails inside. Business figures are mostly done as you know in Excel report or in SAP, for example, in SAP you also use the history at the end, because you could also look for prices which are two or three years made ago for a project. Test reports mainly done as a Win-Word document. And we have for Remscheid, we have a common storage. This is in the test department under the direction of [name deleted]. I'm not sure if the test reports from Belper and Nantes are also collected there; this could be a weak point.

“Mostly difficult is not during a running project to find informations [sic], because there you have networks and know who is storing what and so on. It's more or less an informal base. The main problem is if you have two or three years later any problems to find it again, because there is no rule how to store such informations [sic], especially e-mails.”

4. (a) Where do you and members of your project team look for knowledge?

“There are some standard systems we have, I mentioned the test report database, ok Quality data you could get from the Quality department. The link to SAP or link to Internet where you have a lot of information. Another point is that if you are looking for new information which are not present in the company, then well normally we go to Internet, make some researches, or we have some networks to some Universities or suppliers, from other companies and so on. But then it's not really a structured base, it's more or less you find it [by chance] at the end. For some things, for example [you use a network of experts you know]. If you step in something completely new, it's difficult.”

(b) Do you and your project team search for knowledge just locally (at your site), or is the search extended to the whole Group’s knowledge, or even outside of the Vaillant Group?

“What we also have is an informal network. This Hydraulics There we had each four weeks a meeting together for two days and discussing topics and there at the end there is an informal network and you could ask for informations [sic] there, or if you have a problem you could ask Nantes, Belper and so on, to give you some informations [sic]. This will also work.

“If you have some new technologies or you are looking for a new supplier then you are looking inside or you go to a [technology] fair. There is a tradition. We have [a brand name] and there it is very difficult for me to get information. To ask and get several people to get information and so on. This is a little bit more difficult."

5. (a) How do you and your project team search for the knowledge?
See above. Also "Personal contacts, e-mails, documents."

(b) Do you and your project team use tools to find knowledge? If so, what are the tools, and can you give real examples of tools and systems that support the search?

"SAP, Internet, Test database is our main things, now we have also introduced APIS software where we have the FMEAs. At the end they are also linked afterwards to the Intranet. There is also project files. We have one project file [folder on R&D server] and each project stores its information there, what's not stored are the e-mails."

(c) Do you feel the tools could be improved? If so, how could the tools be improved?

"There was a presentation from [name of a major CAD software company]. They showed a tool that could be linked to SAP or Interlink, what we use for data storing, where we could add also some files to that. For example, if you have a drawing or specification or so on, this linked today and this is stored there. We could add also some other things, we could prepare a complete catalogue. What was not included was, for example was how to store the e-mails and to 'Interlink' it with that."

And on the other side we have the common group share, where we store all the files, like Excel, Win-Word, ProE data maybe are stored in these files. During a project you can find it. The problem is always afterwards."

(d) If no tools are used, what methods of searching for knowledge do you and your project team use and how could they be improved?

"Networks of colleagues."

6. What (kind of) knowledge do you and your project team have problems finding?

"Well, at the moment for example, I am looking for contracts, I am looking for contracts with customers, because we have some OEM products, I am looking for business figures to evaluate the business what we are making. Which is important if you make for example for a new project the business plan . It's very good to look only at the old business plan to look for the other changes behind it. You have a guideline at the end. If you don't have such things, you have to build up everything from scratch, from new."

7. What is the minimum knowledge that you and your project team expect to be able to find?

"I've pointed out some issues. Business figures is important, technical data is important, test reports are important, drawings and specifications are important. Contracts or important agreements with clients, not all agreements with external suppliers are written down and are easy to find. Sometimes it's only done in an e-mail, but there's not really a written contract that you can find in a lawyer department.

“I forgot one thing, quality data. Quality data from existing products and also from former products or from comparable products, to be able to understand what are the failures, what is the failure rate and to analyse and to make improvements at the end, forecasts for the new project, as well. There we had a very good system in the past, but in the moment, I think we are missing some data due to changing of the computer system and all this stuff. And FMEA is also important, there we have started to make all FMEAs with APIS, and if everybody will do that, if we have a link from Internet to APIS, it's normally easy to find out the things and then you could make your new FMEAs and build up knowledge based on this.. And then we document potential risks for products, for example.”

8. In the case of knowledge sought on the company's network (e.g. Project drive), is there any knowledge that you feel your project team needs, but has difficulty accessing?

“To make it short, e-mails are not linked to this project drive.”

9. What frustrates you and you project team about searching for knowledge?

“In general I think each site, so Nantes, Remscheid, Belper and so on has made decisions in the past. And each company or each brand has its own history. And it's very difficult to understand the history and to get the information out of this history, because at the end you have to stay several years in a company and have to feel the spirit of the company or brand. And it's different and it's ok. But this is the point that. For example, some informal things which are clear for everybody in Remscheid and clear for everybody in Nantes, but the link is difficult. And the other thing is that the project drive is always fully loaded with information. As I mentioned and then you get e-mails and you have to take information out. The question is why we collect the information when we then have to take it out, away. Sure then we make some hard copies, but what about the hard copies?”

10. Is there any knowledge that you consider is missing (e.g. information that would assist you in making a decision, but is unavailable)? If so, what is it?

“I think it's a repetition of the points which been made. Another point is more for the future. If you're looking for new developments, I think the network what we have to Universities for example, to experts, this is not so really common and not divided to the whole group. Everybody has a network, but it's not linked together, Big important point for example is approvals and standards. So we have in each location some people who has a good contact to an approval association in France, or to Germany or wherever. But it's not necessarily that if somebody in Nantes has access to the information, that everybody has the information and vice versa.

“Another thing is, for example in approvals and standards, it's not really only knowledge but there's in principle the possibility to drive. We have this strategy and I want to drive approval standards in this direction. So knowledge could not only be what we can get, knowledge could be what we make.”

11. Could you explain why this knowledge is unavailable?

See response to question 10.

Knowledge Sharing

12 (a) What kind of information or knowledge do you and your project team generate?

“If you look on the IP, are you familiar with the IP? So in each phase of the IP we are putting data together. Normally, it's not completely that we are going along this line and collecting all this data, but it's a guideline. And along this guideline we put all these things. So beginning from the strategy, calculating the business, making analyses, testing things, document our tests and our results and drawings and so on. Specifications, all this stuff, workshops with customers. So along this line all of these things will be documented.”

(b) Do you and your team currently collect and organise this information somewhere? If so, where?

“Yes. Project drive and tools like SAP.”

13. Is this knowledge reusable in other projects or in other processes within the IP?

“In principal, yes. So not everything, but especially business figures, strategies, tests FMEA, specifications. Those are the main items… Contracts.”

14. (a) What kind of information would your team be prepared to contribute to an shared R&D knowledge base:

- Now?
- In the future?

Interviewee indicated their agreement with the statement made by the interviewer: "Anything that's on the R&D drive."

b) How and when could you and your project team prepare this shared information? What storage formats and what kind of support (e.g. tools, systems, methods) would you recommend?

“What I would like to have is a kind of Google and that's it. And I would be very interested as I have mentioned how do it other companies.”

c) Should this information or knowledge be shared among the different R&D sites?

“Yes, it's clear. I would also be very interested in Marketing information.”

d) How important is sharing this knowledge among the sites? Would you describe it as mandatory?

See response to question 14(c).

Use of the Intranet as a Search Tool

15. Do you or your project team use the Intranet to search for knowledge?

“Intranet clear, we do. But Intranet is mostly at the moment linked side by side. Vaillant in

Remscheid, Saunier Duval in Nantes. We have a common Intranet, but okay sometimes you find the information is French and so it's not always translated for example. The Standard in English for example. There is also another point, which is at the end difficult. If a project team is mostly for example located in Germany, then it's clear that the language is German. To transfer everything in English... then it's difficult to get all the things together and find it out. The question is, what is the right balance?"

16. Is there any functionality that you feel could improve the process of searching for knowledge on the Intranet?

"We have also a kind register at the moment. But this means that you have to place it there. It's not like Google."

R&D Integration Study

Interview Protocol

R&D Questions

Interviewee: [deleted]

Role: Project Manager

Questions

1. What is your understanding of the term ‘knowledge’?

- Experience
- Rules
- Standards
- What you know

Information or Knowledge Needs

Note to interviewer: Please inform the interviewee that any reference to “you” in isolation, means “you and your project team”, rather than “you personally”.

2. What kind of information/knowledge do you and your project team need in the course of a project?

(For example: Reports, expert advice etc.)

- Basis of customer requirements from marketing departments in countries
- Regulations and standards
- Experience from former projects – to avoid (“reinventing the wheel”)
- Responsibilities in former projects – who was the project leader? This is important in order to exchange experience

3. What is the format of this knowledge?

- Old test reports (.xls/.doc)
- Regulations and standards – Vaillant Intranet updated regularly
- Personal contact e.g. designers

4. (a) Where do you and members of your project team look for knowledge?

- Intranet – Regulations etc
- SAP – old and new drawings
- Personal contacts

(b) Do you and your project team search for knowledge just locally (at your site), or is the search extended to the whole Group's knowledge, or even outside of the Vaillant Group?

Yes.

Group knowledge:

- Search for knowledge in complete group, since the CoCs are located on all sites. Regular meetings with colleagues in Nantes. The knowledge searched for is very dependent on the topic.

- External knowledge sources:
- Search Internet for competitors. Normally this is to look for Boilers, but it is useful to obtain manuals.
- This is not really a requirement for a knowledge database, since there would be too many competitors and appliances.

5. (a) How do you and your project team search for the knowledge?

See answer to Question 4(a).

- Personal contact with experts
- Search for drawings for old products
- Intranet for regulations and standards
- Must navigate to right place
- No direct link to homepage

(b) Do you and your project team use tools to find knowledge? If so, what are the tools, and can you give real examples of tools and systems that support the search?

"Yes."

- Intranet
- SAP – Drawings, part lists for products still in production.

(c) Do you feel the tools could be improved? If so, how could the tools be improved?

- Intranet not user friendly
- Especially for regulations and standards - could be improved

(d) If no tools are used, what methods of searching for knowledge do you and your project team use and how could they be improved?

Does not like SAP.

- A search engine (like Google) would be good) and get relevant
- Search results must be up-to-date and relevant
- Does not have to be fast (a wait time of 1 to 2 seconds for results would be acceptable)

- Would be good to have a “search within a search feature”

6. What (kind of) knowledge do you and your project team have problems finding?

- “It is always difficult to find information that has been purely documented."
- Locating the information is difficult, even if it is on the project drive
- A project leader must look for files created in his or her own project folder
- Files must be stored in a simple structure e.g. in line with the functional roles of the user, Testing, R&D, Purchasing. The simpler, the better.

7. What is the minimum knowledge that you and your project team expect to be able to find?

- The experience of the people working on parallel projects or former projects
- It would be great to have testing protocols in order to avoid repetition if testing. It is mandatory that everyone involved in the development of an appliance can have access to these testing protocols.
- A problem is that the company loses experience of people when they leave the company. For example, one must currently ask people in the testing department for information about a particular test, which becomes impossible if the relevant person has left.

8. In the case of knowledge sought on the company's network (e.g. Project drive), is there any knowledge that you feel your project team needs, but has difficulty accessing?

- On Intranet site for regulations and standards
- A search engine
- Currently must link many links to find huge table of results, some of which may not be there.

9. What frustrates you and you project team about searching for knowledge?

- Time spent searching for knowledge
- Big space for search
- Multiple search terms (synonyms) e.g. Stroemungsversicherung, draft diverter, hood, draft collector
- Languages are a big problem

10. Is there any knowledge that you consider is missing (e.g. information that would assist you in making a decision, but is unavailable)? If so, what is it?

(i) "Most important data has never been written down. This is the experience of people, which is so detailed and difficult to write down. Consider which role someone played in a project. For new people or could be for people on another site. It is important to know whom to contact. Can be found on the Intranet phone book, but need the name first. The roles of people change very quickly and people leave."

(ii) "A link to the IP must be included. There should be one point in each phase in the IP where knowledge must be stored in a database."

(iii) Document management system. Already put knowledge in database a few years ago.

11.Could you explain why this knowledge is unavailable?

"People leave the company."

Knowledge Sharing

12 (a) What kind of information or knowledge do you and your project team generate?

Create knowledge related to:

(i) Realisation of customer requirements
e.g. 3 star efficiency rating for Italy
(ii) Internal topics peculiar to post merger scenarios
Cross country activities
(iii) Experience about a lot of topics e.g. use of a special material that cannot be handled easily

(b) Do you and your team currently collect and organize this information somewhere? If so, where?

“Yes.”

(i) R&D drive. Each project has a special number and special folder. Subfolders are used for special items such as Phase In / Phase Out information.
(ii) Drawings database. Intralink. This system is still not connected to all the sites and people , so one cannot easily exchange drawings. The drawings must be e-mailed to colleagues. In Nantes, this is done via an Intranet tool (hgl or CAD files).
(iii) Normal paper folders in cupboards.

13. Is this knowledge reusable in other projects or in other processes within the IP?

“Yes it is.”

14. (a) What kind of information would your team be prepared to contribute to an shared R&D knowledge base:

- Now?
- In the future?

- Decisions taken during project that will be valid as guidelines for subsequent projects.
- Why decisions were taken.
- Some parts of the IP we did not use.
- Realisation of special topics e.g. relating to special requirements of markets

b) How and when could you and your project team prepare this shared information? What storage formats and what kind of support (e.g. tools, systems, methods) would you recommend?

To motivate people, formats from known tools must be used e.g. Microsoft Word, Microsoft Excel and Microsoft PowerPoint. Some may use PDFs. This makes sense in certain cases, since a PDF cannot be easily changed e.g. test reports, single decisions. However, some documents may need to be amended later on (e.g. by original author). Should be a supervisor who can delete files. The author should be able to delete old files.

c) Should this information or knowledge be shared among the different R&D sites?

- Yes.
- More of a “must” than a “can”.

d) How important is sharing this knowledge among the sites? Would you describe it as mandatory?

"Absolutely mandatory. We do the same jobs on each site for similar products in the same markets. It is a big help to share experience or at least to know who has this experience."

Use of the Intranet as a Search Tool

15. Do you or your project team use the Intranet to search for knowledge?

"Yes, to find out more about products."

Download manuals, pictures

“Document Server” on Intranet is very useful. Can download manuals with a different release version.

16. Is there any functionality that you feel could improve the process of searching for knowledge on the Intranet?

PDF Archive on Intranet

- Must login and interface is unwieldy
- Must know exactly what one is searching for
- The names of the documents vary from project to project
- Easier to do a keyword search – This is possible, but currently it is only possible to search with a single word. Also it is not possible to perform a search within a search
- This would be a brilliant start

Appendix C: Questions for Edwards Vacuum Limited

Questionnaire for designers and project managers

Name:
Date:
Position:

This part is the questions in the questionnaire which is designed for Edwards Vacuum people.

1. General Questions
    a. What type of product do you develop? Do you have incremental multi-generation or completely new generation of product?

    b. Could you describe your job title and responsibility?

    c. How many people are there in your department, company, and typically in a project team?

2. What are the business objectives of current product development in your company? How do you transfer the business objectives to actual project objectives?

3. In your company, what is the scope or boundary of new product development, especially product concept development? What is the normal development time?

4. How often are project reviews held? What are the criteria used to measure the performance of the team?

5. Expertise location: How is an employee's expertise defined e.g. based on his/her involvement in projects, role of the employee in the project, or what information requested by the employee…)?

6. Does your company have distributed project teams? If so, do the teams have national or international scope? How do you manage the communication and cooperation between them in terms of:

    a. Documentation (networks for data sharing, replicated databases for speed increase etc.)

    b. Support tools for processes (workflow management systems, allocation of data, engineering change management, Product Data Management systems, compatibility between PDM and CAD System)

c. Communication (videoconferences, regular international team meetings)

d. Cultural barriers (implementation of standard processes and tools in different cultures, acceptance rates, language etc.)

7. Are all the IT-systems integrated and accessible from a single enterprise portal (Product Data Management / Product Lifecycle Management, Enterprise Resource Planning, databases, information residing in local computers etc.)?

8. How do you and your project team search for the knowledge?

    a. Do you and your project team use tools to find knowledge? If so, what are the tools, and can you give real examples of tools and systems that support the search?

    b. Do you feel the tools could be improved? If so, how could the tools be improved?

    c. If no tools are used, what methods of searching for knowledge do you and your project team use and how could they be improved?

9. How do you structure your documents? Based on your daily work, do you have any recommendations or comments related to current documents structure and the knowledge tool?

10. What kind of problems do you and your project team have in finding available knowledge?

11. What kind of information or knowledge do you and your team generate? Could you give some examples?

12. Do you and your team currently collect and organize this information and knowledge somewhere? If so, how?

13. Can you tell what the composition of your project team is, especially during the product concept design stage (including roles, expertise background, responsibilities)? Is your team cross-functional and/or multi-disciplined? Is your team fixed or dynamic? Do team members regularly contact with customers and other stakeholders?

14. What kind of techniques do you use in the concept design stage (from rough idea to final product concept/specification) of new product development process in your company? (For example: brainstorming, QFD, Axiomatic design, Theory of inventive problem solving (TRIZ), FMEA, Taguchi methods, Six sigma theory, Knowledge management, etc.) In your opinion, what is the importance of these techniques, and rank them?

15. Are there any tools you are using to facilitate quality assurance in product concept design stage? Can you describe the details of your work about quality assurance in this stage?

16. If you are using anyone of FMEA, TRIZ and Axiomatic design, can you tell what the problem you use them to tackle is? Is there any obstacle to use them? And in you opinion, what is the reason for these obstacles?

17. What kind of software do you use in your team to support product concept generation? Can you describe functions of that kind of software? Is there any deficiency about these software? If yes, can you list these deficiencies that need to be improved?

18. Do you use any knowledge management software tools in you team in the concept design stage? If yes, can you describe functions of these tools that you are using in your team? What are the benefits of using them?

19. Is your new product development requirements driven? If yes, please describe how requirement drive the development process.

20. How does your company capture, specify and use the various requirements in new product development? Such as customer requirements, business requirements, design requirements and engineering requirements. How is your organisation getting feedback on the market acceptance of your products?

21. Is there any knowledge based engineering in your company? If yes, please describe how to capture, store, share and reuse knowledge in the engineering department.

22. Are there any enterprise guidelines or processes as a standard to guarantee the company achieve the business goals and objectives? If yes, please describe it with its architecture. E.g. TOGAF

23. How does your company keep product data? Are there any descriptions of each data in database? E.g. drawing, CAD Data and etc.

24. What are the challenges and most critical issues you presently face when aligning your products and services with customer needs?

25. How does your company keep the leading position in the competition?

Appendix D: Answers from Edwards Vacuum Limited

Questionnaire for designers and project managers

Name: [deleted]
Position: Project manager and the assistant

This part is the answers to the questionnaire which is designed for Edwards Vacuum people.

1. General Questions
    a. What type of product do you develop? Do you have incremental multi-generation or completely new generation of product?

Several different types of vacuum pumping;
Pumping vacuum systems;
Vacuum abatement systems which take the gas out in customer process.

We have done new technology, designs, developed some top pumping malicame
e.g. you get some vacuum reused, you imagine that your pumping is not liquid, so you get chose around basically to get high level pumping. So it's the roughing pump staff to the high level vacuum technology. So you kept the whole range including the … the vacant which is be cleaned in the process.

    b. Could you describe your job title and responsibility?

At the moment, for the particular project G4 pump deciding, Ross responsible for the system organisation, design aspects of the product. And Mark is also subs to those, responsible for design the layout and deciding how to bring all the difference to the aspects of the product all together into a unified design. We also responsible for creating the documenting of design within the CAD system and also managing the team of designers that basically looking after the product's different aspects of design. Try to coordinate the designers and try to make sure that some of the …is kind of split over the designers we used is across different sort of barriers that we dealing with.

    c. How many people are there in your department, company, and typically in a project team?

We were in the technical department in Unit 2, about 80 people.
For the company, there are about 4000 people world-wide.
For the current project team, 5 people, but 30-35 people in large project.

2. What are the business objectives of current product development in your company? How do you transfer the business objectives to actual project objectives?

(published condition of business targets, costs, cash flow…turn over per head…profit…different ways of measuring from other different ways… individual labour with the company. Qualified into tools).
GV base more on trying to get involved with other company. Pump is not get exactly regular framework in another machine, so that we have driven towards giving the pump qualified as a sample to manufacturer and processes.
Get into large phone works to design to, and technique related partner, TV speaker is so big. The relation that they would have to pumping is like a suc(?) sort solving. We could…a smallest building. It could be that kind of size. Huge frameworks, lots of pumps that stacked, which is called stackers. We do the solution to provide the involvement of massive vacuation of huge campus to create liquid crystal, which is a kind of screen. A lot of pumps put in basic materials where they want. It is an example these pumps into those involvements manufacturers.
Top manager give some specific…for each new product/project from technical specification.

In process generally they look at what the key processes and education that customers are currently running. There is a case transfer customers and educations that going to change within the next five years. So they will come back saying that" give us particular sort of chemical and education", and they working at really…product range to develop a key, to……working back from that all the time.
That comes down into when you look at the rest marketing specification, and that within there, would be target to sell X many per year, that sort of thing. Marketing is got to connect inlet. The outlet has been decided, in order to generally describe sort of performance and size primaries.
The product is in order to address the marketing requirement, so that could help business objectives.
Pump is going to be customer's pump, and the vacuum that customer needs. So you won't design one pump to do one job. You are on the way, but you got some level to change. Commercialization managers of each project need to understand beyond the design, how each unit is a contributor, what contribution level should the product makes, to forecast what we get. So there always is a link between technical and the commercial marketing. Commercialization managers are the key person. They look at the product and contribution, and costs, and make sure that in design they said very strict cost primaries to the product, so that even can satisfy those sort of business objectives, it makes business sense to develop the product.
To some extend, business objectives are already converted into a product design aspect. For that we don't really the linking between. It is already done. As we know a pump cost and sample cost so much. And therefore there're something works I make most of. I have worked with other scenarios, which is especially when you doing business with customers want integrate pumps. They don't want separate unit. They want build pump into a unit. They require something a bit more than just the standard design of pump.

3. In your company, what is the scope or boundary of new product development, especially product concept development? What is the normal development time?

If you take technical mental project, e.g. customer direction exercise to analysis…you'll take an existing pump country, a pump and booster. You don't want to touch those. You would like to say we leave those alone. What we do is that we look at external part of it, the cost analysis, if break down how much the machine's component, how much is part work. You will look and see where to do operate analysis, where is the cost lie. It helps you evaluate money. With those kind of incremental works, you generally have a very quick process, maybe a month, maybe a couple of months.
You real appreciation product we aspect to, again real application what do customers really want, and what's the history, since you got feedbacks from the field so you could upon that and say what is the really important things, the list. This kind of thing may cost couple of months.
For big project, if you start again with a different pump co…, then you may spend maybe a year to look at the different options. For development, a couple of years could be spent just to look at different technologies, and decide which is best suit to the arrange of applications that available. After that you could go through embodiment. We have reached something last three years. And we are trying to get down to two years for the whole including the concept.
The recent product is also a platform abroad arrange product. The raw…has the same basic architecture. That's why we could test all in such a broad application range. You test rather all the technology would satisfy with those different ranges so within there you have product development evaluation. You do some embodiment within it as well.

Incremental design – mainly do a cost analysis and takes 1-2 months.
Main design of new products – 12-24 months for concepts, 3 years for the whole product (aim to reduce it to 2 years).

4. How often are project reviews held? What are the criteria used to measure the performance of the team?

We have formal reviews and informal reviews.
Informal reviews are weekly based. The main one that have weekly based is called PCM, Project Control Meeting. The technical staff will attend those.
We did sort of project development process, which is called PCP, Product Commercial Organisation Process. And that's something which is a stage gating formal process in order to release funding in the company for the next stage of development. You need to demonstrate that …be accomplished. And there are 4 or 5 stages in that of things. That's the design of you at the gating processes.
When they feel that the designer can make sure they can hold it. Normally you have a designer with you. It will be safety. So the safety aspect is slipped in parallel.
People know what involved in the design, which may measure one way. Some is demonstrated much better is I have joined to produce. Individual members of a team have a yearly appraisal. This sort of top level business requirements in terms of, the project is part of your objectives in there. So we always help to solve these objectives but it's no so easy.
The other objective would be for example finish concept stage by e.g. March, or particular year. You can't get timing deadlines during stage.

Informal reviews – weekly
Formal reviews – product control meetings, product commercial process, safety review, design review, etc.

5. Expertise location: How is an employee's expertise defined e.g. based on his/her involvement in projects, role of the employee in the project, or what information requested by the employee…)?

Based on: Assume we know our team. What have been done before and who were involved, appraisals, recruitment, sometimes develop skills needed.
This is the area that the company will benefit a lot if staff expertise is available and visible.

6. Does your company have distributed project teams? If so, do the teams have national or international scope? How do you manage the communication and cooperation between them in terms of:

Yes. Yes.

   a. Documentation (networks for data sharing, replicated databases for speed increase etc.)

Email pass the documents.
Drawings, shared folders, video links. SmarTeam PLM system.

b. Support tools for processes (workflow management systems, allocation of data, engineering change management, Product Data Management systems, compatibility between PDM and CAD System)

PDM
CAD
Catia, which is more successful used here. But for other groups it may not be used because of cost of licence, people may not like to change, or local managers opinion. Senior management may not be interested in the workflow.

c. Communication (videoconferences, regular international team meetings)

Skype instead of big screen video conferences. From big to small. Not much international communication except some telephone calls.

d. Cultural barriers (implementation of standard processes and tools in different cultures, acceptance rates, language etc.)

Yes. Not from individual but from company point of view
e.g. in Japan, no graphics by designers. Different ways of working. But Korea is more same to western.

Japan: normally do what is told by superior, not much individual ideas (but when asked individuals will bring up ideas)

EU: more individual ideas

USA: do not listen to others. You tell them to done some thing, they come up with different ideas.

7. Are all the IT-systems integrated and accessible from a single enterprise portal (Product Data Management / Product Lifecycle Management, Enterprise Resource Planning, databases, information residing in local computers etc.)?

No.
MRP system as backbone
SAP for commercial
SCP links these two above.
It may not need really to put everything in a single portal.
But there are some benefits of a single one, like time for data transferring, common feel for workers. And it might be useful for people management, track information, not for data.
Mentioned SOPHIN software, which was used in a company to track system used by staff.

People focus on their area, without considering other areas

8. How do you and your project team search for the knowledge?

a. Do you and your project team use tools to find knowledge? If so, what are the tools, and can you give real examples of tools and systems that support the search?

No.

Intranet for reference MRP system sort information. Some decision could be driven from this.
Collection of paper work.

People put 'folders' in their way. No structure for the management of the folders.

b. Do you feel the tools could be improved? If so, how could the tools be improved?

Some search tool to find information in the 'folders' put by different people would be very useful.

Alan is doing some work in this area

c. If no tools are used, what methods of searching for knowledge do you and your project team use and how could they be improved?

Design a component.
Potential dangerous.
Talk to people what we would like to know and they may know.
Each project get a structure file all documents could be find.
Database for searching for.
Product menus for search in Intranet, but not brilliantly sorted.

9. How do you structure your documents? Based on your daily work, do you have any recommendations or comments related to current documents structure and the knowledge tool?

Smarteam takes all document PDM system.
Individual put folder in the system everywhere.
For drawings and PDM are OK. But for individual it is hard to find their folders.
No strategy on how to handle causes the hazard.
It should be official way to tell where these folders should be put.
Project related documents are difficult to find than department documents.

10. What kind of problems do you and your project team have in finding available knowledge?

11. What kind of information or knowledge do you and your team generate? Could you give some examples?

12. Do you and your team currently collect and organize this information and knowledge somewhere? If so, how?

13. Can you tell what the composition of your project team is, especially during the product concept design stage (including roles, expertise background, responsibilities)? Is your team cross-functional and/or multi-disciplined? Is your team fixed or dynamic? Do team members regularly contact with customers and other stakeholders?

14. What kind of techniques do you use in the concept design stage (from rough idea to final product concept/specification) of new product development process in your company? (For example: brainstorming, QFD, Axiomatic design, Theory of inventive problem solving (TRIZ), FMEA, Taguchi methods, Six sigma theory, Knowledge management, etc.) In your opinion, what is the importance of these techniques, and rank them?

FMEA use them.
But not so much in concept design stage.
TRIZ is used when have problems.
Taguchi is not used.
Six sigma is not used in design.
Brainstorm …
QFD is not used because it is a bit fuzzy.
Customer may drive design, and design may find customer. But all are not from new technologies.

15. Are there any tools you are using to facilitate quality assurance in product concept design stage? Can you describe the details of your work about quality assurance in this stage?

Project Knowledge Reuse

Project 1 copies similar project in the past. Project 2 copies similar project in the past. But there is no link between project 1 and 2 about shared information and knowledge.

16. If you are using anyone of FMEA, TRIZ and Axiomatic design, can you tell what the problem you use them to tackle is? Is there any obstacle to use them? And in you opinion, what is the reason for these obstacles?

Obstacle: They won't interest in techniques unless show they the savings.

17. What kind of software do you use in your team to support product concept generation? Can you describe functions of that kind of software? Is there any deficiency about these software? If yes, can you list these deficiencies that need to be improved?

CAD Three forms is useless.
Sketchers
10 unique designs in 2 weeks
Pen and paper is better to do the sketch.

18. Do you use any knowledge management software tools in you team in the concept design stage? If yes, can you describe functions of these tools that you are using in your team? What are the benefits of using them?

No.

19. Is your new product development requirements driven? If yes, please describe how requirement drive the development process.

Sometimes Customer gives design, and sometimes design find customer.
Concept design
Predict the market

20. How does your company capture, specify and use the various requirements in new product development? Such as customer requirements, business requirements, design requirements and engineering requirements. How is your organisation getting feedback on the market acceptance of your products?

21. Is there any knowledge based engineering in your company? If yes, please describe how to capture, store, share and reuse knowledge in the engineering department.

Refer to Question 15.
Design menu in Intranet

22. Are there any enterprise guidelines or processes as a standard to guarantee the company achieve the business goals and objectives? If yes, please describe it with its architecture. E.g. TOGAF

23. How does your company keep product data? Are there any descriptions of each data in database? E.g. drawing, CAD Data and etc.

PDM system
Project folder electronic
Documents change not free through

24. What are the challenges and most critical issues you presently face when aligning your products and services with customer needs?

Make sure what customer really want
Formalized and agreed
Sub control because things change. Keep chasing.

25. How does your company keep the leading position in the competition?

Cost issue
Reliability
Good service
Company's revenue
Quickly turn around, design specific for customer response
React, produce the product for customer and cost standard
Understanding the market
Copy exactly, which means that to keep the old products and keep making for customer, totally the same.

Appendix E: Answers from Archive Department in CACC

Knowledge Search Investigation

Interview Protocol

Archive Management Questions

Name: [deleted]
Position: Archive assistant

This part is the answers to the questionnaire which is designed for Archive people in CACC.

1. Could you describe your work scope and responsibility?

2. Expertise location: How is an employee's expertise defined?

Employees profile, involvement in projects, personal decision
Where they could find this information
Analysis of requested information by each employee

3. Does the company have distributed project teams? If so, do the teams have national or international scope? How do you manage the communication and cooperation between them in terms of:

If YES and International team scope

a. Documentation

Enterprise intranet for data sharing
May have duplicated databases in international site, and data update every week
e.g. CPC flat form for both design site and several manufacturing sites. Data transferred from the design site FAI to the main manufacturing site SAMF, and SAMF do planning work and transfer the date to other sites from CPC flat form as well. This CPC flat form is developed by PTC company. Database in SAMF is updated once data is sent from FAI.

b. Support tools for processes

Workflow management systems, allocation of data, engineering change management, Product Data Management systems, compatibility between PDM and CAD System
e.g. PDM system on CPC flat form for SAMF

c. Communication

Video conferences, telephone conferences, regular international team meetings

d. Cultural barriers

Implementation of standard processes and tools in different cultures
Acceptance rates
Language problems
e.g. special translation team to translate documents from English to Chinese
same tool with the same version are used on all sites to avoid compatibility problem

4. Are all the IT-systems integrated and accessible from a single enterprise portal?

PDM/PLM (Product Data Management / Product Lifecycle Management)
ERP (Enterprise Resource Planning)
Databases
Information residing in local computers, etc

5. How do you and your project team search for the knowledge?

a. Do you and your project team use tools to find knowledge? If so, what are the tools, and can you give real examples of tools and systems that support the search?

Microsoft search tool to search for documents in shared company document structure
Special search tool developed by expert team to search for documents by name, file number, department, author, project, key word, generated period, etc

b. Do you feel the tools could be improved? If so, how could the tools be improved?

Reduce search result
Precision of the retrieval information

c. If no tools are used, what methods of searching for knowledge do you and your project team use and how could they be improved?

According to document structure
According to document serial number

6. Based on your daily work, do you have any recommendations or comments related to current documents structure and the knowledge tool?

7. What kind of knowledge does you and your project team have problem finding?

Knowledge mastered by someone who is retired or left
Unstructured information exists in proprietary formats like paper, media objects

8. What kind of information or knowledge do you and your team generate?

Weekly conference report
CAD model
Report of failure
e.g. EBOM which is the list of material
PD which is the bottom layer design model of the empennage in Catia V5
EO (Engineering Order for detailed description of Catia model in form of Excel
)
FO (Failure and Reject Order)

9. Do you and your team currently collect and organize this information somewhere? If so, where?

PDM (Product Development Management) system on CPC flat form
VPM system to manage data of design

Appendix F: Codes for Obtaining Data from Database

GenIndexDataDb realises the interface of AbstractGenIndexData. It is to obtain data from database for Lucene. The returned data is in the pattern of SearchInfor set.

```
package com.lan.search.lucene.genindex.proxy;

import java.sql.Connection;
import java.sql.PreparedStatement;
import java.sql.ResultSet;
import java.sql.SQLException;
import java.util.ArrayList;
import java.util.List;
import com.lan.search.db.DAO;
import com.lan.search.db.DBConnection;
import com.lan.search.db.DbConfig;
import com.lan.search.db.viewdao.SearchColumnsConfigDao;
import com.lan.search.domain.viewdomain.SearchColumnsConfig;
import com.lan.search.domain.viewdomain.SearchDbConfig;
import com.lan.search.domain.viewdomain.SearchInfo;

public class GenIndexDataDb implements AbstractGenIndexData{

    public List<SearchInfo> getIndexLuceneData() throws Exception {
        List<SearchInfo> result = new ArrayList<SearchInfo>();
        List<SearchColumnsConfig>        tableList        =        new
SearchColumnsConfigDao().listAll();
        Connection conn = null;
        PreparedStatement pstm = null;
        ResultSet rs = null;
        for(SearchColumnsConfig config : tableList){
            String databaseId = config.getDatabaseId();
            String tableName = config.getTableName();
            String columnName = config.getColumnNames();
            String keycolumnName = config.getKeycolumnname();

            SearchDbConfig queryDbOb = new SearchDbConfig();
            queryDbOb.setId(databaseId);
```

```
        List<SearchDbConfig>        dbconfigList        =        new
DAO().select(queryDbOb,DBConnection.MAIN_DATABASE);
        if(dbconfigList.size() > 0){
            SearchDbConfig schDbConfig = dbconfigList.get(0);
            String url = schDbConfig.getUrl();
            String userName = schDbConfig.getUsername();
            String password = schDbConfig.getPassword();

            DbConfig dbConfig = new DbConfig();
            dbConfig.setUrl(url);
            dbConfig.setUserName(userName);
            dbConfig.setPassword(password);

            conn = DAO.getConnection(dbConfig);
            try {
                pstm = conn.prepareStatement(" select * from " + tableName);
                rs = pstm.executeQuery();
                while (rs.next()) {
                    SearchInfo queryInfo = new SearchInfo();
                    queryInfo.setDataBaseId(databaseId);
                    queryInfo.setPath(url);
                    queryInfo.setTableName(tableName);
                    queryInfo.setColumnName(config.getColumnNames());
                    queryInfo.setKeyName(keycolumnName);
                    queryInfo.setKeyValue(rs.getString(keycolumnName));
                    queryInfo.setType(SearchInfo.SEARCH_INFO_TYPE_DB);

                    StringBuffer searchInfoValue = new StringBuffer();
                    String[] keyColumns = columnName.split(",");
                    for(int i=0;i<keyColumns.length;i++){

    searchInfoValue.append(rs.getString(keyColumns[i].trim()));
                        searchInfoValue.append(" ");
                    }
```

```
                    queryInfo.setContent(searchInfoValue.toString());
                    result.add(queryInfo);
                }
            } catch (SQLException e) {
                e.printStackTrace();
            }finally{
                if(rs!=null){
                    rs.close();
                }
                if(pstm!=null){
                    pstm.close();
                }
                if(conn!=null){
                    conn.close();
                }
            }
        }
    }
    return result;
}
public static void main(String[] args) throws Exception {
    GenIndexDataDb db = new GenIndexDataDb();
    List<SearchInfo> queryInfos = db.getIndexLuceneData();
    for(SearchInfo info : queryInfos){
        System.out.println(info.getContent());
    }
}

}
```

Appendix G: Codes for Obtaining Data from Shared Files

GenIndexDataShareFile realises the interface of AbstractGenIndexData. It is to obtain data from shared files for Lucene. The returned data is in the pattern of SearchInfor set.

```
package com.lan.search.lucene.genindex.proxy;

import java.util.ArrayList;
import java.util.List;

import com.lan.search.db.viewdao.SearchShareFileDao;
import com.lan.search.domain.viewdomain.SearchInfo;
import com.lan.search.domain.viewdomain.SearchShareFileConfig;
import com.lan.search.lucene.smb.RemoteFileInfo;
import com.lan.search.lucene.smb.SmbUtil;

public class GenIndexDataShareFile implements AbstractGenIndexData{

	public List<SearchInfo> getIndexLuceneData() throws Exception {
		List<SearchInfo> result = new ArrayList<SearchInfo>();
		List<SearchShareFileConfig>	configList = new SearchShareFileDao().listAll();
		for(SearchShareFileConfig config : configList){
			String ip = config.getIp();
			String userName = config.getUsername();
			String password = config.getPassword();
			String remoteUrl = 
"smb://WORKGROUP;"+userName+":"+password+"@"+ip;
			List<RemoteFileInfo> fileInfoList = SmbUtil.getShareFilePath(remoteUrl);
			for(RemoteFileInfo info : fileInfoList){
				String fileName = info.getFileName();
				String suffix = "";
				if(fileName.indexOf(".")!=-1) {
					suffix = fileName.substring(fileName.indexOf(".")+1);
				}
				SearchInfo queryInfo = new SearchInfo();
				queryInfo.setDataBaseId("");
				queryInfo.setPath(info.getFilePath());
				queryInfo.setTableName("");
```

```
                queryInfo.setColumnName("");
                queryInfo.setKeyName("");
                queryInfo.setKeyValue(fileName);
                queryInfo.setType(SearchInfo.SEARCH_INFO_TYPE_SHAREFILE);
                queryInfo.setContent(fileName+"                    "+suffix+"
"+SmbUtil.readFromSmb(info.getFilePath()));
                result.add(queryInfo);
            }
        }
        return result;
    }
    public static void main(String[] args) throws Exception {
        GenIndexDataShareFile file = new GenIndexDataShareFile();
        List<SearchInfo> infoList = file.getIndexLuceneData();
        for(SearchInfo info : infoList){
            System.out.println(info.getContent());
        }
    }
}
```

Appendix H: Codes for Generating Interface

GenIndexDataAgent is the Agent for generating index. It is employed to generate interface that inherits AbstractGenIndexData.

```
package com.lan.search.lucene.genindex.proxy;

import java.lang.reflect.InvocationHandler;
import java.lang.reflect.Method;
import java.lang.reflect.Proxy;
import java.util.List;
import com.lan.search.domain.viewdomain.SearchInfo;

public class GenIndexDataAgent implements InvocationHandler{

	private Object object;

	public GenIndexDataAgent(Object prObject){
		this.object = prObject;
	}
	public static Object factory(Object object){
		Class cls = object.getClass();
		return Proxy.newProxyInstance(cls.getClassLoader(), cls.getInterfaces(), new
GenIndexDataAgent(object));
	}

	public Object invoke(Object proxy, Method method, Object[] args) throws Throwable
{
	    List<SearchInfo> queryInfoList = (List<SearchInfo>)method.invoke(object,
args);
		return queryInfoList;
	}

	/**
	 * @param args
	 * @throws Exception
	 */
	public static void main(String[] args) throws Exception {
```

```
        AbstractGenIndexData abstractSubject = (AbstractGenIndexData)factory(new GenIndexDataShareFile());
        abstractSubject.getIndexLuceneData();
    }
}
```

Appendix I: Codes for Generating Index for Lucene

GenerateIndexUtil is employed to generate the index for Lucene. GenerateIndexUtil method obtains data from the classes that inherit AbstractGenIndexData, and generates index by createIndex.

```
package com.lan.search.lucene;

import org.apache.lucene.index.IndexWriter;
import jeasy.analysis.MMAnalyzer;
import java.util.ArrayList;
import java.util.List;
import org.apache.lucene.document.Document;
import org.apache.lucene.document.Field;
import com.lan.search.domain.viewdomain.SearchInfo;
import com.lan.search.lucene.genindex.proxy.AbstractGenIndexData;
import com.lan.search.lucene.genindex.proxy.GenIndexDataAgent;
import com.lan.search.lucene.genindex.proxy.GenIndexDataDb;
import com.lan.search.lucene.genindex.proxy.GenIndexDataShareFile;

public class GenerateIndexUtil {
	public final static String indexDir = 
Class.class.getClass().getResource("/").getPath()+"luceneDir";

	public static void createIndex(String indexDir, List<SearchInfo> searchInfos) throws 
Exception {
		if(searchInfos==null || searchInfos.size() == 0){
			return;
		}
		IndexWriter writer = new IndexWriter(indexDir, new MMAnalyzer(),true);
		writer.setUseCompoundFile(true);
		for(SearchInfo searchInfo : searchInfos){
			Document doc = new Document();
			doc.add(new Field("dataBaseId", 
String.valueOf(searchInfo.getDataBaseId()),Field.Store.YES, Field.Index.NO));
			doc.add(new Field("searchInfoType", 
String.valueOf(searchInfo.getType()),Field.Store.YES, Field.Index.NO));
			doc.add(new Field("path", 
String.valueOf(searchInfo.getPath()),Field.Store.YES, Field.Index.NO));
```

```
            doc.add(new Field("tableName", String.valueOf(searchInfo.getTableName()),Field.Store.YES, Field.Index.NO));
            doc.add(new Field("columnName", String.valueOf(searchInfo.getColumnName()),Field.Store.YES, Field.Index.NO));
            doc.add(new Field("keyName", String.valueOf(searchInfo.getKeyName()),Field.Store.YES, Field.Index.NO));
            doc.add(new Field("keyValue", String.valueOf(searchInfo.getKeyValue()),Field.Store.YES, Field.Index.NO));
            if(searchInfo.getContent()!=null && !"".equals(searchInfo.getContent())){
                doc.add(new Field("searchInfoValue", searchInfo.getContent(), Field.Store.YES,Field.Index.TOKENIZED));
            }
            writer.addDocument(doc);
        }
        writer.optimize();
        writer.close();
    }

    public void generateIndex() throws Exception{
        List<SearchInfo> queryInfos = new ArrayList<SearchInfo>();
        SearchInfo queryInfo = new SearchInfo();
        queryInfo.setDataBaseId("");
        queryInfo.setTableName("");
        queryInfo.setColumnName("");
        queryInfo.setKeyValue("");
        queryInfo.setKeyName("");
        queryInfo.setType(SearchInfo.SEARCH_INFO_TYPE_DB);
        queryInfo.setContent("");
        queryInfos.add(queryInfo);
        createIndex(indexDir,queryInfos);

        List<SearchInfo> queryInfoList = new ArrayList<SearchInfo>();
        AbstractGenIndexData abstractDBSubject = (AbstractGenIndexData)GenIndexDataAgent.factory(new GenIndexDataDb());
```

```
            queryInfoList.addAll(abstractDBSubject.getIndexLuceneData());

            AbstractGenIndexData                    abstractFileSubject                    =
(AbstractGenIndexData)GenIndexDataAgent.factory(new GenIndexDataShareFile());
            queryInfoList.addAll(abstractFileSubject.getIndexLuceneData());

            createIndex(indexDir,queryInfoList);
      }
      public static void main(String[] args) throws Exception {
            GenerateIndexUtil util = new GenerateIndexUtil();
            util.generateIndex();
      }
}
```

Appendix J: Codes for Searching Index of Lucene

QueryFromIndexUtil is to search index of Lucene.

```
package com.lan.search.lucene;

import java.util.*;
import jeasy.analysis.MMAnalyzer;
import org.apache.lucene.analysis.Analyzer;
import org.apache.lucene.document.Document;
import org.apache.lucene.queryParser.MultiFieldQueryParser;
import org.apache.lucene.search.Hits;
import org.apache.lucene.search.IndexSearcher;
import org.apache.lucene.search.Query;
import org.apache.lucene.search.Searcher;
import com.lan.search.domain.viewdomain.SearchInfo;

public class QueryFromIndexUtil {
    public static List<SearchInfo> search(String indexDir,String[] condition, int startIndex,int endIndex) throws Exception{
        Searcher searcher = null;
        List<SearchInfo> result = null;
        try {
            searcher = new IndexSearcher(indexDir);
            Analyzer analyzer = new MMAnalyzer();
            String[] fields = { "searchInfoValue"};
            Query query = MultiFieldQueryParser.parse(condition, fields, analyzer);
            Hits hits = searcher.search(query);

            result = new ArrayList<SearchInfo>();
            if (endIndex >= hits.length()){
                endIndex = hits.length() - 1;
            }
            for (int i = startIndex; i <= endIndex; i++) {
                Document doc = hits.doc(i);
                SearchInfo searchInfo = new SearchInfo();
                searchInfo.setDataBaseId(doc.get("dataBaseId"));
```

```
				searchInfo.setType(doc.get("searchInfoType"));
				searchInfo.setPath(doc.get("path"));
				searchInfo.setTableName(doc.get("tableName"));
				searchInfo.setColumnName(doc.get("columnName"));
				searchInfo.setKeyName(doc.get("keyName"));
				searchInfo.setKeyValue(doc.get("keyValue"));
				searchInfo.setContent(doc.get("searchInfoValue"));
				result.add(searchInfo);
			}
		} catch (Exception e) {
			throw e;
		} finally {
			if (searcher != null){
				searcher.close();
			}
		}
		return result;
	}
	public static void main(String[] args) throws Exception {
		try {
			SearchInfo searchInfo = null;
			String[] condition = { "moto"};
			List<SearchInfo>  result  =  search(GenerateIndexUtil.indexDir,condition,  0,
100);
			for (Iterator<SearchInfo> iter = result.iterator(); iter.hasNext();) {
				searchInfo = iter.next();

	System.out.println("==========================================");
				System.out.println(searchInfo.getDataBaseId());
				System.out.println(searchInfo.getType());
				System.out.println(searchInfo.getPath());
				System.out.println(searchInfo.getTableName());
				System.out.println(searchInfo.getColumnName());
				System.out.println(searchInfo.getKeyValue());
```

```
                System.out.println(searchInfo.getContent());

    System.out.println("================================================");
            }
        } catch (Exception e) {
            e.printStackTrace();
        }
    }
}
```

Appendix K: Codes for Obtaining Data from Shared Files of the Intranet

SmbUtil is used to obtain data from the shared files of the Intranet. getShareFilePath method obtains the set of shared files from IP address. readFromSmb method obtains the content of a file. copyFromSmb method copies the shared file to local for swing.

```
package com.lan.search.lucene.smb;

import java.io.BufferedInputStream;
import java.io.BufferedOutputStream;
import jcifs.smb.SmbFileInputStream;
import java.io.FileOutputStream;
import java.io.IOException;
import java.io.InputStream;
import java.io.OutputStream;
import java.util.ArrayList;
import jcifs.smb.SmbFile;
import java.util.List;
import java.io.File;

public class SmbUtil {
    public final static String SMB_TEMP_FILE_DIR = Class.class.getClass().getResource("/").getPath()+"SmbTempFileDir";

    public static List<RemoteFileInfo> getShareFilePath(String remoteSmbUrl) throws Exception{
        List<RemoteFileInfo> result = new ArrayList<RemoteFileInfo>();
        SmbFile remoteFile = new SmbFile(remoteSmbUrl);
        SmbFile[] files = remoteFile.listFiles();
        for(int i=0;i<files.length;i++){
            SmbFile file = files[i];
            String fileName = file.getName();
            String filePath = file.getCanonicalPath();
            if(filePath.indexOf("$/")==-1){
                RemoteFileInfo fileInfo = new RemoteFileInfo();
                fileInfo.setFileName(fileName);
                fileInfo.setFilePath(filePath);
                if(file.isDirectory()){
                    result.addAll(SmbUtil.getShareFilePath(fileInfo.getFilePath()));
```

```java
                }else{
                    fileInfo.setFileType(RemoteFileInfo.FILE_TYPE_FILE);
                    result.add(fileInfo);
                }
            }

        }
        return result;
    }
    public static File copyFromSmb(String smbMachine, String localpath) {
        File localfile = null;
        InputStream bis = null;
        OutputStream bos = null;
        try {
            SmbFile rmifile = new SmbFile(smbMachine);
            String filename = rmifile.getName();
            rmifile.getDfsPath();
            bis = new BufferedInputStream(new SmbFileInputStream(rmifile));
            localfile = new File(localpath);
            localfile.mkdirs();
            localfile = new File(localpath + File.separator + filename);
            bos = new BufferedOutputStream(new FileOutputStream(localfile));
            int length = rmifile.getContentLength();
            byte[] buffer = new byte[length];
            bis.read(buffer);
            bos.write(buffer);
        } catch (Exception e) {
            System.out.println(e.getMessage());
        } finally {
            try {
                bos.close();
                bis.close();
            } catch (IOException e) {
                e.printStackTrace();
```

```
            }
        }
        return localfile;
    }
    public static String readFromSmb(String smbMachine) {
        InputStream bis = null;
        byte[] buffer = new byte[0];
        try {
            SmbFile rmifile = new SmbFile(smbMachine);
            bis = new BufferedInputStream(new SmbFileInputStream(rmifile));
            int length = rmifile.getContentLength();
            buffer = new byte[length];
            bis.read(buffer);
        } catch (Exception e) {
            System.out.println(e.getMessage());
        } finally {
            try {
                if(bis!=null){
                    bis.close();
                }
            } catch (IOException e) {
                e.printStackTrace();
            }
        }
        return new String(buffer);
    }
    public static void main(String[] args) throws Exception {
        String remoteUrl = "smb://WORKGROUP;Guest:helloworld@10.180.3.201";
        List<RemoteFileInfo> fileInfoList = SmbUtil.getShareFilePath(remoteUrl);
        for(RemoteFileInfo fileInfo : fileInfoList){
            System.out.println("==========================");
            System.out.println(readFromSmb(fileInfo.getFilePath()));
            System.out.println("==========================");
        }
```

```
    }
}
```