An Analysis of the Current Nature, Status and Relevance of Data Mining tools to enable Organizational Learning

Martin Hattingh

Assignment presented in partial fulfillment of the requirements for the degree of Master of Commerce at the University of Stellenbosch.

Supervisor : Prof M Leibold

December 2002

Declaration

I, the undersigned, hereby declare that the work contained in this assignment is my own original work and has not been previously in its entirety or in part been submitted at any university for a degree.

Abstract

The use of technological tools has developed rapidly over the past decade or two. As one of the areas of business technology, Data Mining has been receiving substantial attention, and thus the study defined the scope and framework for the application of data mining in the first place.

Because of the wide area of application of data mining, an overview and comparative analysis was given of the specific data mining tools available to the knowledge worker.

For the purposes of the study, and because the goals of data mining involve knowledge extraction, the concept of organizational learning was analysed. The factors needed to facilitate this learning process were also taken into consideration, with a view towards enabling the process through their improved availability.

Actual enablement of the learning process, through the improved factor availability described above, was analysed through the use of each specific tool reviewed.

The salient conclusion of this study was that data mining tools, applied correctly, and within the correct framework and infrastructure, can enable the organizational learning process on several levels. Because of the complexity of the learning process, it was found that there are several factors to consider when implementing a data mining strategy.

Recommendations were offered for the improved enablement of the organizational learning process, through establishing more comprehensive technology plans, creating learning environments, and promoting transparency and accountability. Finally, suggestions were made for further research on the competitive application of data mining strategies.

Abstrak

Die gebruik van tegnologiese hulpmiddels het gedurende die afgelope dekade of twee snel toegeneem. As een afdeling van ondernemings tegnologie, is daar aansienlike belangstelling in 'Data Mining' (die myn van data), en dus het die studie eers die omvang en raamwerk van 'Data Mining' gedefinieer.

As gevolg van die wye toepassingsveld van 'Data Mining', is daar 'n oorsig en vergelykende analise gegee van die spesifieke 'Data Mining' hulpmiddels tot beskikking van die kennis werker.

Vir die doel van die studie, en omdat die doelwitte van 'Data Mining' kennis-onttrekking behels, is die konsep van organisatoriese leer geanaliseer. Die faktore benodig om hierdie leerproses te fasiliteer is ook in berekening gebring, met die mikpunt om die proses in staat te stel deur verbeterde beskikbaarheid van hierdie faktore.

Werklike instaatstelling van die leerproses, deur die verbeterde faktor beskikbaarheid hierbo beskryf, is geanaliseer deur 'n oorsig van die gebruik van elke spesifieke hulpmiddel.

Die gevolgtrekking van hierdie studie was dat 'Data Mining' hulpmiddels, indien korrek toegepas, binne die korrekte raamwerk en infrastruktuur, die organisatoriese leerproses op verskeie vlakke in staat kan stel. As gevolg van die ingewikkeldheid van die leerproses, is gevind dat daar verskeie faktore is wat in ag geneem moet word wanneer 'n 'Data Mining' strategie geïmplementeer word.

Aanbevelings is gemaak vir die verbeterde instaatstelling van die organisatoriese leerproses, deur die daarstelling van meer omvattende tegnologie planne, die skep van leer-vriendelike omgewings, en die bevordering van deursigtigheid en rekenskap. In die laaste plek is daar voorstelle gemaak vir verdere navorsing oor die kompeterende toepassing van 'Data Mining' strategieë.

Acknowledgements

This thesis is dedicated to two men who embody the spirit of business, and who have taught me the value of learning. Without their support and encouragement, this thesis would not have been possible. To my father Henri and uncle John.

Also, to the spirit of innovation. This thesis was researched and compiled in its entirety through the use of electronic media – Specifically, not a single piece of paper was used as a reference. Rather, several internet based data mining technologies were employed in gathering *relevant* published information in electronic format.

Contents :

Chapter **Page**

CHAPTER 1 : INTRODUCTION……………………………………………..…..1

1.1 Background to the study………………………………………………….…..1
1.2 Problem Statement……………………………………………………..…..4
1.3 Objective of the study ………………………………………………..……9
1.4 Scope of the study………………………………………………………..…...8
1.5 Methodology……………………………………………………………..10
1.6 Structure of presentation……………………………………………………10

CHAPTER 2 : CURRENT APPROACHES TO DATA MINING……….…….13

2.1 Introduction…………………………………………………………….……13
2.2 The emergence of Data Mining………………………………………..…...13
2.3 Factors Driving the Emergence of Data Mining……………………………14
 2.3.1 Increase in Data quantity…………………………………….………14
 2.3.2 Accelerated pace of change…………………………………….………16
 2.3.3 The move from "crunching" to communicating……………………18
 2.3.4 Global Relocation…………………………………………………...19
 2.3.5 The changing nature of the organization……………………………19

2.4 Data Mining approaches……………………………………………………21
 2.4.1 Categories of approaches……………………………………………22
 2.4.2 Particular Approaches to Data Mining……………………………...23
  2.4.2.1 Statistics……………………………………………………23
  2.4.2.2 Artificial Intelligence (AI)…………………………………24
  2.4.2.3 Decision Trees……………………………………………..25
  2.4.2.4 Visualisation………………………………………………26
 2.4.3 Goals of approaches to Data Mining………………………………..27

2.5 Falconview case study..28
2.6 Analysis and Synthesis..29
2.7 Summary..30

CHAPTER 3 : DATA MINING TOOLS..31

3.1 Introduction..31
3.2 Specific Data Mining Tools..32
3.2.1 Expert Systems..32
3.2.2 Genetic Algorithms..33
3.2.3 Clustering..34
3.2.4 Neural Networks..35

3.3 Comparative analysis of Data Mining tools..36
3.3.1 Application areas of Data Mining Tools..37
3.3.1.a Expert Systems..37
3.3.1.b Genetic Algorithms..38
3.3.1.c Clustering..39
3.3.1.d Neural Networks..40
3.3.2 Implementability of Data Mining Tools..41
3.3.2.a Expert Systems..42
3.3.2.b Genetic Algorithms..43
3.3.2.c Clustering programs..43
3.3.2.d Neural Networks..44
3.3.3 Interoperability of Data Mining Tools..44
3.3.3.a Expert Systems..45
3.3.3.b Genetic Algorithms..45
3.3.3.c Clustering programs..46
3.3.3.d Neural Networks..46
3.3.4 Interoperability case study : K-12 School databases..47

3.4 Analysis and Synthesis..49
3.5 Summary..49

CHAPTER 4 : ORGANIZATIONAL LEARNING – THE FACILITATING FACTORS……………………………………………………………………………50

4.1 Introduction……………………………………………………………...…..50

4.1.1 Purpose of Organizational Learning…………………………...……....50

4.1.2 Organizational Learning defined………………………………….....50

4.2 Organizational Learning's facilitating factors…………………………….....53

4.2.1 Scanning imperative and performance gap……………………..…53

4.2.2 Concern for measurement…………………………………………..…..54

4.2.3 Experimental mindset…………………………………………………..…56

4.2.4 Climate of openness……………………………………………………....58

4.2.5 Continuous education..……………………………………………….....59

4.2.6 Operational variety…………………………………………………..….59

4.2.7 Involved leadership……………………………………………………..…60

4.2.8 Systems perspective………………………………………………..….61

4.3 Conclusions on facilitating factors…………………………………..….63

4.4 Infrastructure for improving learning……………………………….……64

4.4.1 The information system infrastructure………………………….……67

4.5 Analysis & Synthesis……………………………………………………..…..69

4.6 Summary…………………………………………………………………..….70

CHAPTER 5 : ENABLING ORGANIZATIONAL LEARNING THROUGH DATA MINING…………………………………………………………………………71

5.1 Introduction……………………………………………………………………..71

5.2 The relevance of applying Data Mining in Organizational Learning………..72

5.3 Data Mining Tools – Improving the availability of Organizational Learning's facilitating factors ………………………………………………76

5.3.1 Factor 1 – Scanning imperative and performance gap....................76

a. Expert Systems....................77

b. Genetic Algorithms....................77

c. Clustering....................78

d. Neural Networks....................78

5.3.2 Factor 2 – Concern for measurement....................79

5.3.3 Factor 3 – Experimental Mindset....................80

a. Expert Systems....................80

b. Genetic Algorithms....................81

c. Clustering....................81

d. Neural Networks....................81

5.3.4 Factor 4 – Climate of openness....................82

5.3.5 Factor 5 – Continuous Education....................82

5.3.6 Factor 6 – Operational Variety....................83

a. Expert Systems....................83

b. Neural Networks....................84

5.3.7 Factor 7 – Involved Leadership....................84

5.3.8 Factor 8 – Systems Perspective....................85

a. Expert Systems....................85

b. Genetic Algorithms....................85

c. Neural Networks....................86

5.4 The direct effects of Data Mining tools on Organizational Learning....................86

5.5 Effects of a Data Mining orientation on Organizational Learning....................88

5.5.1 Internal Benchmarking...89
5.6 Analysis & Synthesis...90

5.7 Summary...91

CHAPTER 6 – SUMMARY, CONCLUSIONS AND RECOMMENDATIONS...92

6.1 Summary...92
6.1.1 Defining Data Mining...92
6.1.2 The need for Data Mining...93
6.1.3 Approaches to Data Mining...93
6.1.4 Comparative analysis of Data Mining tools...94
6.1.5 Enablement of Organizational Learning...95
6.1.6 Enablement infrastructure...96
6.1.7 Using Data Mining tools to enable Organizational Learning..........96

6.2 Synthesis and Conclusions...98
6.3 Recommendations for Business Applications...100
6.4 Recommendations for further research...101

Bibliography...102

List of figures

Figures **Page**

Figure 1.1: How knowledge relates to organizational learning and sustainable competitive advantage..3

Figure 1.2: The Knowledge Warehouse..7

Figure 2.1: Principle modules of "documentation" in information space.............21

Figure 2.2: Role of knowledge-based systems in the knowledge management chain...23

Figure 4.1: Dimensions of organizational learning....................................52

Figure 4.2: The balanced scorecard..56

Figure 4.3: Relating corporate vision to core business activities.......................61

Figure 4.4: Elements of the learning process..63

Figure 4.5: The Knowledge Cycle...65

Figure 4.6: Data, Information and Knowledge..66

Figure 4.7: Separation of organisational and technological issues.....................69

Figure 5.1: Organizational knowledge infrastructure and its relation to sustainable competitive advantage..74

Figure 5.5 : The model of innovative management.....................................89

List of Tables

Tables **Page**

Table 4.1: Individual and Organizational Learning…………………………………...53

Table 5.1: Distinctions between data, information, knowledge and wisdom……...73

Chapter 1 : Introduction

1.1 Background :

Throughout the advancement of technology in the last 15 years, the focus has been on improving efficiency in all facets of work and life in general, with efficiency being one of the highest rated perceived benefits of knowledge management (McAdam & Reid, 2000). This drive towards efficiency is in part facilitated by an ever increasing technology adoption rate, and the rapid increase in microcomputer processor speed, which effectively doubles every 18 months according to Moore's law (Anderson et al, 2002).

Above and beyond improving efficiency, technology is also used to improve effectiveness. Organizational effectiveness has served as a unifying theme for more than a century of research on the management and design of organizations, yet no universal theory has developed (Lewin & Minton, 1986).

However, in one instance, organizational effectiveness is defined as the extent to which an organization, by the use of certain resources, fulfils its objectives without depleting its resources and without placing undue strain on its members and/or society (Thibodeaux & Favilla, 1996). In this case, the correct activities have to be performed, which implies that the individual or organization concerned has to know what these activities are.

In this study, the "activities" refer to organizational activities, those activities performed to directly increase profitability, since profit is a mainstream goal of every business organisation (Morris, 1996). However, these activities can also be aimed at reaching other organizational goals such as growth, market share, customer service, value delivery etc. In general, these activities are performed with relatively concrete goals, set by the organization's management, shareholders, or both.

As a relatively vague definition (for the purpose of this study) it is assumed that the primary organizational goal is that of creating value for customers. An Organizational

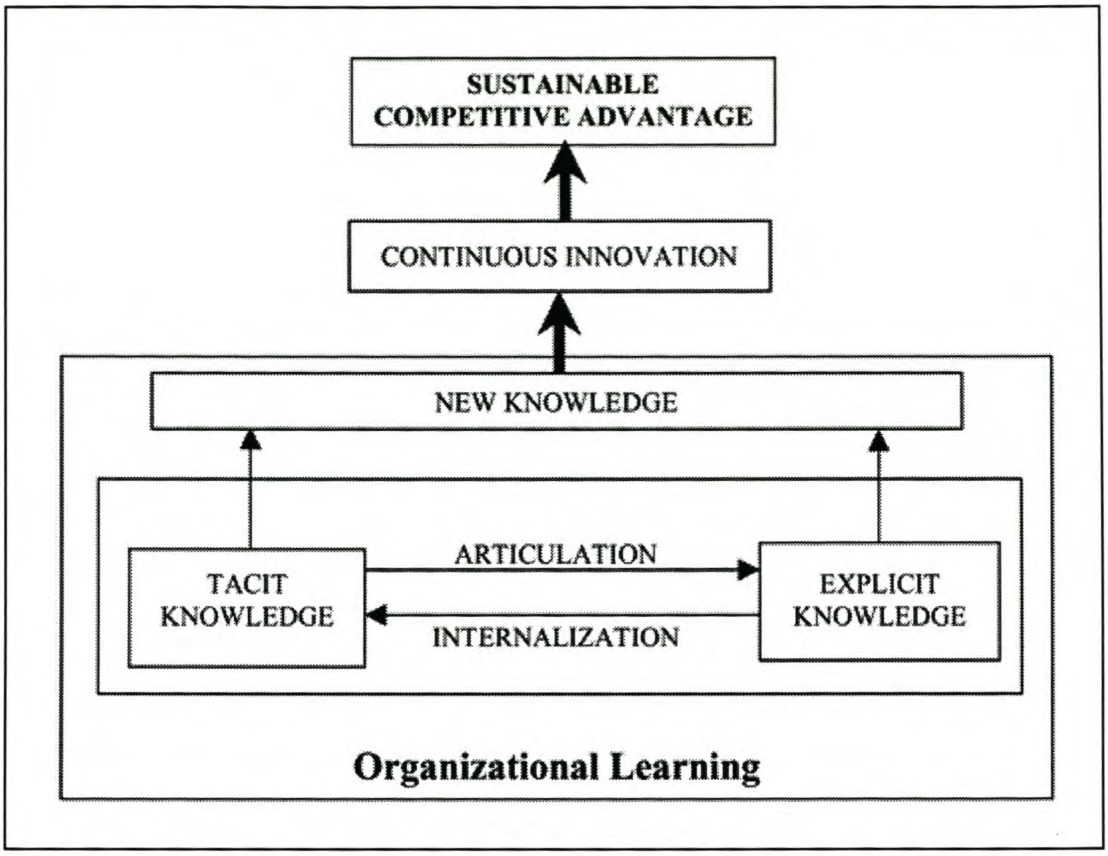
Knowledge Management System (OKMS) enables a firm to leverage the strategic value of its knowledge (Peteraf, 1993). For the purpose of this study it is not necessary to go into the specifics of this goal, as it is simply used to illustrate organizational success. Organizational success as such is achieved through performing organizational activities.

Without performing organizational activities *effectively* the organization can thus not reach its goal of creating value (or will take longer to reach this goal). It is this effectiveness which forms the core motivation for this study, more specifically *how* the organization knows *what* to do, in order to do it efficiently and ultimately create and add value.

Even though the definition of *what to do* is an extremely broad one, the aim is to touch on this subject through an analysis of *how* an organization knows what to do - through utilizing databases (facilitated through technology). More specifically, utilization of these databases through an interactive and intelligent process – Data Mining, which can by applied across a variety of industries.

For example, by effectively harnessing the capabilities provided by information technologies, Ford is attempting to create a sustainable competitive advantage in the auto industry (Kerwin, 2000). Meso & Smith (2000) use the following figure to conceptualize how knowledge is used to create a sustainable competitive advantage.

Figure 1.1 : How knowledge relates to organizational learning and sustainable competitive advantage.



Source : Meso & Smith (2000)

Data Mining in itself is can also be a relatively wide field of exploration, thus the aim will become even more focused on specific activities needed to facilitate the process of Data Mining. These activities (as facilitators) will be classified as "tools" in order to create the structure needed to examine the process as a whole. Inspired by different paradigms, there is a wide spectrum of data mining tools available (Firebaugh, 1998).

The "tools" referred to above are methods, techniques, technologies and processes used to make possible the goal of data mining, and shall be explored as the main subject of this study.

1.2 Problem Statement :

As was mentioned in the introduction, without the organization performing its activities effectively, it cannot really add value - thus effectiveness is a strong requirement for organizational success. One of the requirements of performing activities effectively is knowing *what* to do – This is accomplished through having knowledge of which activities add the most value to the organization. These activities are performed through utilising the organization's resources.

An organization's tangible and intangible resources may be grouped into two main categories: firm resources and firm capabilities (Grant, 1991). According to Grant, this designation implies that resources are inputs into the production process and the capability of a firm is the capacity, what it can do, as a result of teams of resources working together.

One of the prerequisites to knowing *what to do* with the abovementioned resources is for employees to have information relevant to them available in a form applicable to them, and useable by them. Most organizations routinely (either with explicit purpose, or in the course of daily activities) collect and maintain large databases covering a range of information types.

Examples of data routinely collected are :

- Customer / Client contact details
- Transaction records
- Customer interaction records
- Movement patterns

These data types are usually collected either manually, or through automated processes. Manual collection of data is extremely time consuming, and does not lend itself to large databases, thus the focus will fall mainly on data collected through automated processes.

The automated processes referred to are mainly electronic-capture based, and run through Information Technology Networks which automatically manage both the capture and storage phases of the process.

Even though this capture and storage process is often very well planned, and executed with a large degree of efficiency, it usually only delivers a large amount of *data,* and not *knowledge*. Although data is a prerequisite in the process of data mining, it cannot aid decision-making in its raw form. For data to become useful, it has to be converted to knowledge. This is where the problem arises :

Knowledge cannot easily be stored (Gopal and Gagnon, 1995). Unlike raw material, knowledge usually is not coded, audited, inventoried, and stacked in a warehouse for employees to use as needed. It is scattered, messy, and easy to lose (Galagan, 1997).

This knowledge (as described above) is usually the result of a conversion process, of which data is the main input. Conversion is needed to create knowledge out of large sources of data. The conversion process may take place from two directions :

- Explicit Knowledge
- Tacit Knowledge

Knowledge may be tangible or intangible in nature. Know-what, know-how, and know-why, when articulated into the organization's database and operating technologies, are tangible. Similarly, explicit knowledge is tangible because it has been encoded into documents, databases, or some other permanent medium. (Quinn *et al.*, 1996; Michalisn *et al.*, 1997; Davenport *et al.*, 1998).

Tacit knowledge, on the other hand, is intangible. Because tacit knowledge resides in an individual, its benefits are not long-term. It is lost when the individual leaves the organization. Because tacit knowledge can be captured as explicit knowledge through the process of articulation, it becomes possible to acquire. (Michalisn *et al.*, 1997; Peteraf, 1993; Wernerfelt, 1984).

When the organization captures data into a database, and then stores and processes it, it becomes possible for the organization to acquire the knowledge held within the data – However, this knowledge is not available for consumption before it has been properly converted and made available in the correct form to the correct people.
It is this conversion process which is critical in making knowledge useable and useful to individuals within, and the organization as a whole.

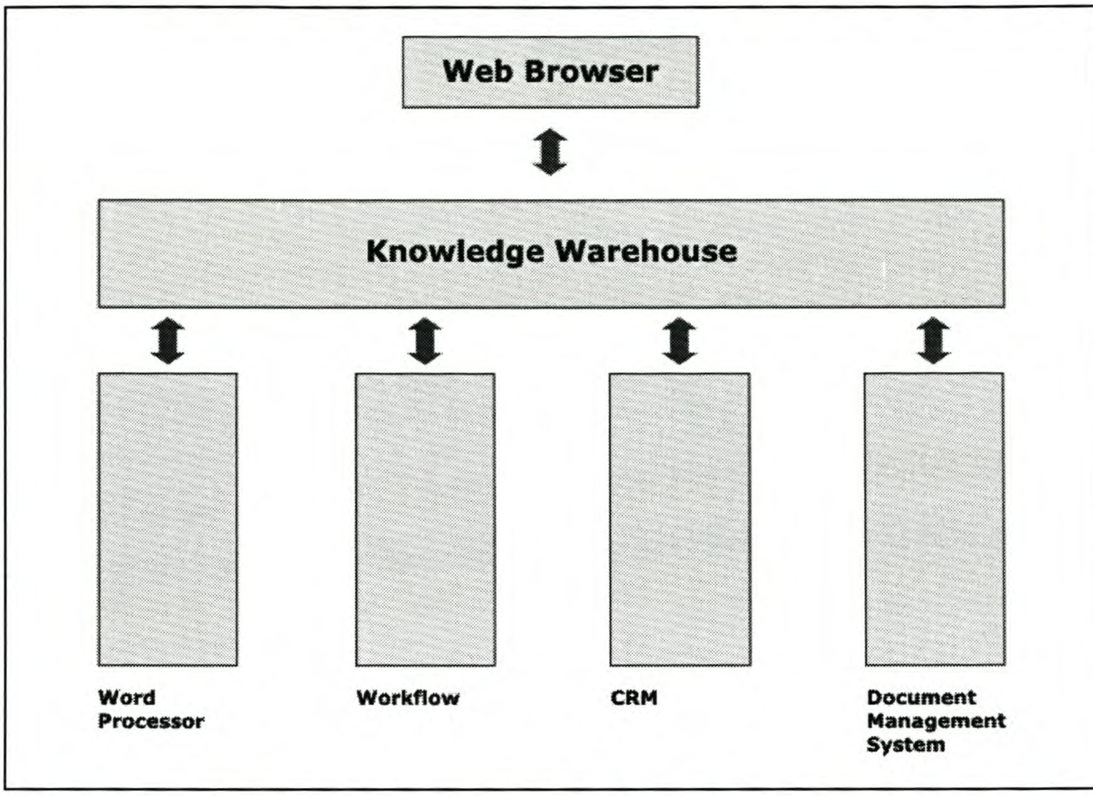
The core components needed in the conversion process are :

- The Data Warehouse
- Data Mining

<u>The Data Warehouse (DW):</u>
The information (or knowledge) is often stored in an organizational Data Warehouse. The data warehouse architecture consists of a consolidated, consistent relational database and an information management system server which is the focal point of all end user queries, as well as, the data access mechanism for analytical and quantitative studies. (Gargano, 1999) Figure 1.1 illustrates the Data (or Knowledge in more advanced form) warehouse.

<u>Data Mining (DM):</u>
Data mining is concerned with discovering new, meaningful information, so that decision makers can learn as much as they can from their valuable data assets. Using advanced information technologies, knowledge discovery in databases (KDD) can uncover veins of surprising and golden insights in a mountain of factual data.
(Gargano, 1999)

Figure 1.2 : The Knowledge Warehouse



The synergy created between the holistic paradigms of data warehousing (DW) and data mining (DM) allows goal oriented decision makers to leverage their massive data assets, thereby improving the effectiveness and quality of their decisions (Berson and Smith, 1997).

Making information available to the correct people in the correct form is the main goal of Data Mining.

Once the correct information has been converted, and is in the hands of the appropriate people, the organization can start applying this information as knowledge.

The application of this knowledge forms part of the process of organizational learning. Organizational learning is defined by one author as being :

*A process by which a firm and its people develop their capabilities to create a desired future (*Senge, 1990)

Another definition :

Organizational learning is the process of continued innovation through the creation of new knowledge (Quinn et al., 1996; Nonaka, 1991).

Critical facilitators which make possible the application and successful implementation of a culture of organizational learning include data mining, but data mining is often seen as a technical tool in the domain of the technically inclined. A clearer understanding of the benefits of applying a data mining strategy is needed to enable management to make informed decisions about it. This understanding needs to be rooted in a defined explanation of its indirect (and also direct) benefits to business, through the organizational learning (and knowledge distribution) process.

Data mining as such is not a practical "tool" that can be purchased, but rather a synergy of different techniques and methods of using information to maximum advantage in the organizational context.

1.3 Objective :

The objective of the study is **an analysis of the current nature, status and relevance of Data Mining tools to enable Organizational Learning.**

In this regard, the ultimate goal is a "learning organization". A learning organization is one that creates, acquires, and communicates information and ***knowledge***, behaves differently because of this, and produces improved organizational results from doing so (Huber, 1991).

1.4 Scope of the Study :

The scope of the study includes the current approaches and techniques used in the mining of data from large databases (as described in the objectives), and thus deals with the extraction and conversion process. Conversion in this case refers to the classification of data in order to facilitate context-related retrieval by the relevant parties.

With the above in mind, it would be unnecessary to deal with the general approaches and theories regarding database management, although these are used to illustrate points and bring into context issues in the study.

More specifically then, the scope of the study is as follows :

- To examine the current theories and approaches regarding the extraction of specific information from databases.
- To examine the tools applied in various data mining approaches
- To examine the technological advances in the field of data mining
- To explore the effects of data mining on the dissemination of information within the organization.
- The use of data mining as a tool in assisting Organizational Learning.

1.5 Methodology :

As stated in the objective (see page 8), the study aims to conduct an analysis of the *use of* data-mining tools in enabling organizational learning.

For the purpose of this study, only secondary sources were necessary, with no primary source research.

The scope defined in point 1.4 includes both organizational learning and database management techniques, but precludes probing these subjects in-depth. Thus, the methodology is as follows :

- A review of the tools currently available for data-mining purposes.
- Brief analysis of how each tool is applied.
- An overview of what is needed to enable Organizational learning.
- Analysis of the application of each tool in organizational learning.
- Comparison between the effectiveness of each tool and an analysis of general tool use in enabling organizational learning.

1.6 Structure of the presentation :

The study is presented in 6 chapters, with an overview as follows :

Chapter 1:

The first chapter provides a background to the study, with preliminary definitions of the various areas of the study. It also provides concise definitions of the most important concepts and terms. The problem statement and objective of the study is described, with descriptions of the scope and methodology employed following as natural progressions.

Chapter 2:

The second chapter provides an overview of the current approaches to data mining, including a focus on the emergence of data mining, and the factors driving this emergence. A few approaches to data mining are reviewed, with descriptions of the basic thinking behind each approach. Finally, conclusions are formulated on the apparent goals of the various approaches.

Chapter 3:

This chapter provides a review of specific data mining tools available to the knowledge professional. A comparative analysis of these tools is done, with focus on their application areas, implementability, and interoperability. In this analysis, detail is provided of the impact of these tools on the organization's information technology infrastructure.

Chapter 4:

The fourth chapter moves the focus towards organizational learning, with specific reference to what factors are needed to facilitate the learning process. These factors are also analysed according to availability, and the subsequent common reasons for shortage.

Chapter 5:

This chapter probes the effect that data mining tools have on organizational learning's facilitating factors (as discussed in chapter 4). It provides an analysis of the relevance of applying data mining in organizational learning context, and focuses on the application of data mining tools in improving the availability of the facilitating factors. Also, an analysis is done of how data mining tools directly affect the construct of organizational learning.

Chapter 6:

The last chapter provides a summary of the study, and brings to light conclusions on the current state of data mining, more specifically the effect it has on enabling organizational learning. This chapter also provides reccomendations on an improved understanding of the area of study, and states the possible future research challenges.

Chapter 2 : Current Approaches to Data Mining

2.1 Introduction

In order to analyse the effects that specific Data Mining tools have on Organizational Learning, the focus first has to turn towards how Data Mining is applied in the organization. Specific Data Mining tools are widely used, but it is prudent to first take into consideration different approaches an organization can take in adopting a mining strategy.

This chapter first reviews the emergence of the Data Mining trend, and looks at why it has become necessary to mine information to gain from it. Different factors influencing the trend are reviewed.

Next, the study turns towards categories of approaches, more specific approaches, and the goals of these approaches. It culminates with a case study on the application of Data Mining in a military environment.

2.2 The Emergence of Data Mining :

Data mining as a field is one of the areas of management which relates very closely to (and is ultimately very dependent on) the rapid progress of technology. Technology plays a critical facilitation role in providing management with the means to disseminate ever increasing amounts of data. The storage of this data is also directly linked, as the availability of economically viable mass storage has increased as the development cycle continues.

Information Technology (IT) specifically provides a high performance and relatively stable platform for management on which to base a strategy of utilising data and information in all its forms to improve the capabilities of the organization. This platform is again dependent for its viability and cost-effectiveness on the large amounts of data it supports. It is thus a self-sustaining system which provides all parties involved with increased resources in their respective business areas.

As information volume has increased as rapidly as IT has developed and evolved, many organisations have been caught lacking in the ability to use this information to attain their organisational goals. Increasing the ability to store information does not necessarily mean that the ability to *utilise* this information is also increased. Many organisations have encountered endless problems related to large investments in IT without an implementable strategy of using this information to increase management effectivity (note effectivity as defined in Chapter 1, *doing the right things*). These problems have often been rooted in a misunderstanding of the capabilities of a properly managed and implemented IT infrastructure.

The capabilities mentioned above include not only the large-scale storage and processing capabilities of an efficient IT infrastructure, but also the information retrieval capabilities, and through this the knowledge transfer facilitation on offer. It is this facilitation which is critical in making possible the true utilisation of IT resources. Because of the need to enable transfer of knowledge, there have been several factors driving data mining / information dissemination.

2.3 Factors driving the emergence of Data Mining :

As mentioned above, it is possible to broadly list the factors either directly or indirectly responsible for the increased interest and investment in Data Mining techniques and technologies. Some of these factors are :

2.3.1 Increase in Data Quantity :

As mentioned previously, the increasing use of mass storage devices has driven down prices in this market, and thus increased the mass usage as an effect. In the 1980's, storage of 500 Megabytes (500 Million characters) was considered mass storage, today the largest storage devices in SAN (Storage Area Network, a networked array of optical storage devices) configuration can reach up to 25 or more Terabytes (25

000 Gigabytes). Since 1975, the information-carrying capacity of the global telecommunications networks has increased by over a million-fold. New optic fibre networks - each with a wire smaller than the size of a human hair - are being installed in cities, suburbs and countrysides all around the world, alongside power cables and rail lines. Each fibre has the capacity to transmit the data equivalent of the entire *Encyclopaedia Britannica* in less than five seconds (*The Economist*, 1996a).

As the economical price of mass storage capability has decreased, adoption of this medium among individual users has also increased. This individual storage phenomenon, together with the increased internet connection rate, has also served to enlarge the global electronic data storage capacity.

Businesses are becoming more distributed. Technology is a major factor. Prices are falling, performance is increasing, and ease and speed of deployment (especially for Internet- and intranet-based services) are increasing. Technologies that have been seen as addressing different requirements (such as workflow, groupware, database, Internet, desktop products) are converging and overlapping in function. There are three consequences:

1. The options for providing ICT services keep widening, and the choices are harder to make.

2. It is becoming impractical to isolate facets of ICT services, such as workflow and database, and design them separately.

3. It is wise to specify ICT services in such a way that they can be mapped to newer, cheaper, easier-to-use technologies when the time is right to do so.

Also, technologies for managing non-database data (documents, product data, images, sound, video) are rapidly becoming cheaper, increasing in capability, and being integrated with workflow and intranet products. Adopting more of this technology will cause major changes to many organizations and will dramatically change the ratio of their database and document-based IT support - it is generally accepted that less than 10 per cent of business information is stored in databases. (Franckson et al, 1998)

The problem that has surfaced with mass data proliferation is a finding a meaningful way to process all this data, and thus technology has been developing in this direction. Data Mining is one of these technologies, and the interest in its development has served to increase the interest in understanding relationships between large amounts of information better.

2.3.2 Accelerated pace of change :

The management of change is increasingly becoming a factor regularly discussed at the top level of management. A few decades ago, it was relatively common to see organizations and individuals being relatively complacent about the future, and striving towards the ultimate status quo. Several academic teachers encountered by the author in his undergraduate studies have listed "the quiet life" as one of the ultimate organizational goals. This has changed.

In this new world, entire industries may spring up, thrive and be eliminated in a decade, as knowledge-based growth continues to shorten product life cycles, compress development times, drive new product prices downward, and increase the competition for better technical standards (*The Economist*, 1996a, p. 10; Howitt, 1996; Utterback, 1994). The computer industry itself provides a typical example, where some 70 per cent of revenue today comes from products which did not even exist two years ago.

This change means a rapid increase in innovation diffusion, implying that companies cannot rely on existing knowledge for nearly as long as was previously possible. It has become possible for individuals to quickly acquire skills and knowledge previously only available to the organizational elite – Increasing the potential for innovation. An individual or small business can thus quickly, through rapid innovation, become a competitor to a large corporation (which might be unaware of the competition until it is too late.)

To keep up with the vast amounts of information they have acquired, and to manage the changes inherent in it, organizations have sought to obtain tools to manage the knowledge contained within this information. More importantly, the changes related through the inherent knowledge have to be managed, and this is where Data Mining comes in. Data Mining acts as a tool in change management, providing management with timely information on patterns of change within the organization and its competitors.

2.3.3 The move from "crunching" to communicating :

Although (as mentioned in the previous points) the quantity of information has increased radically, the major effect on the way business is done is not directly quantity related. Rather, technology has moved from serving as a major source of data processing – "crunching" – to a tool making possible communication in ways never thought possible before. It is impossible to quantify the effect that a simple "killer app" such as e-mail has had on the way that most organisations (and individuals for that matter) conduct their day-to-day activities.

It was this shift from "crunching to communicating" that allowed electronic systems to revolutionize financial markets and which meant that organizations for the first time were able to capture quickly, "codify" and disseminate huge amounts of information and distribute it among branch offices, suppliers and customers anywhere in the world (Tapscott, 1996, pp. 95-121).

As communication in all its forms increases in quantity and speed, information is distributed over many more nodes than previously. Access to this information is also made easier through the technology, with it being technically possible for anyone within the organisation and outside it to access any information if needed (and if allowed).

The effect rapid communication has had on the change of business processes should once again be emphasized, as it has made speed of reaction one of the key factors contributing to competitive advantage. The organisation wanting to exploit this advantage to its full effect needs to be able to attain specific extracts of knowledge from large sources of data quickly to react quickly (and thus communicate effectively). Data Mining serves as the tool with which this quick reaction and communication is made possible.

2.3.4 Global Relocation :

The global location trend has recently shifted with traditional "third world" countries having rapidly adapted to the advances in technology. These countries have managed to develop manufacturing systems which are able to compete with the best in the world at a fraction of the cost.

It is still difficult to gauge the ultimate effect of this new phase in the evolution of global economics, but it is fairly certain that, in advanced economies, low and medium-skill production will increasingly be either moved away to low-cost labour markets globally, or abandoned altogether, forcing a further shift toward the high-skill, high-technology, "knowledge-based" industries where advanced economies still retain a comparative advantage.

Data Mining is one of the important facilitators which will contribute to the "knowledge-based" economies retaining a competitive advantage. Even though the low-cost-producing economies can adapt very quickly, they do not necessarily have the insights gained from utilising advanced technologies in the Data Mining field, being applied by those focused purely on the intellectual and knowledge domain.

2.3.5 The Changing nature of the Organization :

With the changes in demographics mentioned in the previous point taking place so quickly, it is no surprise that the structure of the organization itself is also changing.

Organizations have to adapt to changing environments, and these dictate two options :

- Continue with the current "base" organisational structure, and adapt to changes as they occur. This approach is often found in organizations with a local / domestic market, which rely on standardised production techniques. The reaction to environmental change is often to lower costs in all possible areas, of which human resources is usually the most visible. This leads to "broken" relationships between key knowledge players within the organization, and to stagnation of knowledge resources. There is no proactive management of change. In this regard, it has been stated that "Organisations have to be prepared to abandon knowledge that has become obsolete" (Drucker, 1993).

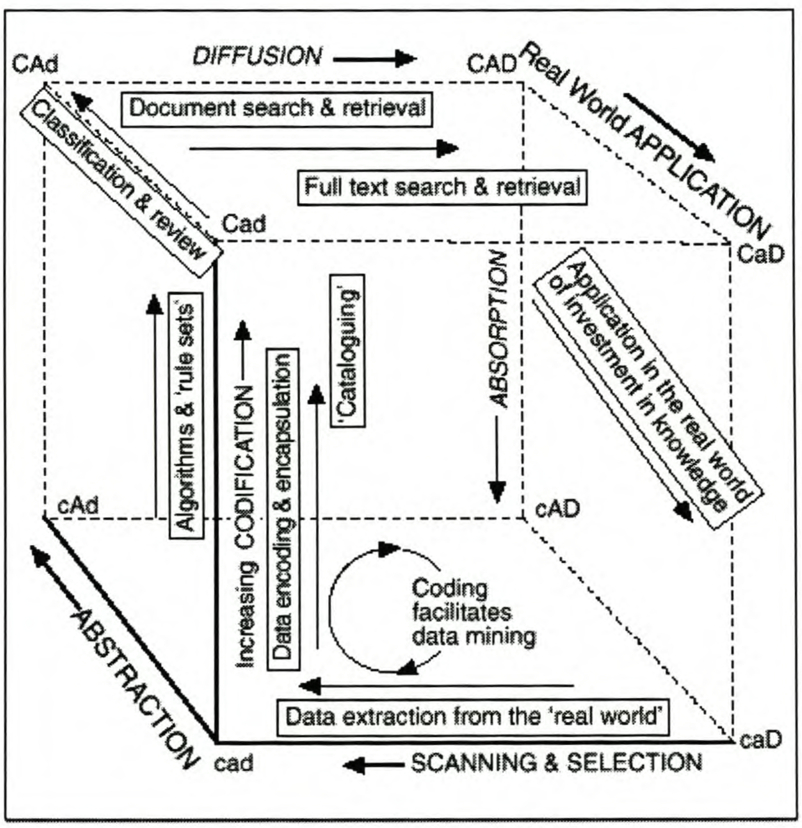
- Transform the organization into one which proactively manages change in the environment, as well as within itself. Management of Human Resources (with its intrinsic knowledge pool) as a key asset becomes very important. Relationships and the sharing of knowledge that coincides with them cannot be "broken" in this approach. According to some of the studies on learning organization, the individuals are connected to the organization by a shared vision and by a perspective of wholeness (e.g. Senge, 1990a).

With a clearer understanding of some of the reasons Data Mining has emerged and become popular, an overview of a few models on which to base a Data Mining process follows.

2.4 Data Mining Approaches :

Before a Data Mining implementation strategy can be formulated, the organization in question needs to carefully consider the different models available. In this case, models refer to general approaches to Data Mining, and not to technical specifics. When looking at the possible complexity of data input / output within the organizational context as in figure 2.1 (Ashford, 1997), it is easy to realise that the choice of a strategic approach to follow with regards to Data Mining is very important.

Figure 2.1 : Principle modules of "documentation" in information space



Source : Ashford (1996)

2.4.1 Categories of approaches :

According to one author (Chen *et al.*, 1996), approaches to a Data Mining strategy can be broadly defined as falling into three categories :

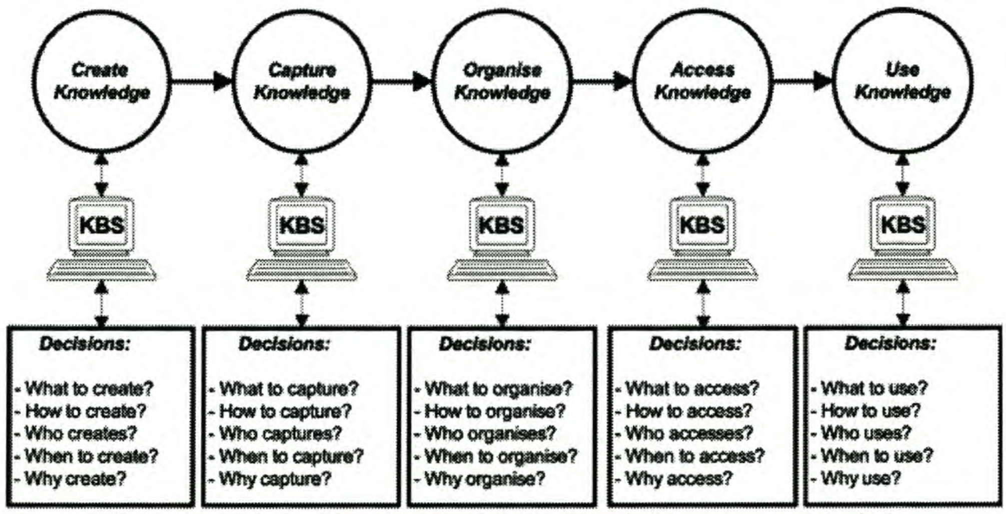
- **Based on the Database** - There are many database systems that are used in organizations, such as : relational databases, transaction databases, object-oriented databases, spatial databases, multimedia databases, legacy databases, and web databases. A Data Mining system can be classified based on the type of database it is designed for. For example, it is a relational DM system if the system discovers knowledge from relational databases and it is an object-oriented DM system if the system finds knowledge from object-oriented databases.

- **Based on the Knowledge** – Data Mining systems can discover various types of knowledge, including association rules, characteristic rules, classification rules, clustering, evolution, and deviation analysis. DM systems can also be classified according to the abstraction level of the discovered knowledge. The knowledge may be classified into general knowledge, primitive-level knowledge, and multiple-level knowledge.

- **Based on the Techniques** – Data Mining systems can also be categorized by DM techniques. For example, a DM system can be categorized according to the driven method, such as autonomous knowledge mining, data-driven mining, query-driven mining, and interactive DM techniques. Alternatively, it can be classified according to its underlying mining approach, such as generalization-based mining, pattern-based mining, statistical- or mathematical-based mining and integrated approaches. (Lee & Siau, 2001)

The categories of approaches are important considerations because they influence the area of the business within which the Data Mining strategy will be applied. If one looks at figure 2.2 (Soliman, 1999), which relates to the creation of a Knowledge-based system, the different phases in the knowledge management chain could apply to different areas of the organization, thus the approach chosen has a direct effect on the way implementation takes place.

Figure 2.2 : Role of knowledge-based systems in the knowledge management chain



Role of knowledge-based systems in the knowledge management chain

Source : Soliman et al (1999)

2.4.2 Particular approaches to Data Mining

2.4.2.1 Statistics

Statistics is an indispensable component in data selection, sampling, data mining, and extracted knowledge evaluation. It is used to evaluate the results of data mining to separate the good from the bad. In a data cleaning process, statistics offer the techniques to detect "outliers", to smooth data when necessary, and to estimate noise. Statistics can also deal with missing data using estimation techniques. Techniques in clustering and designing of experiments come into play for exploratory data analysis. Work in statistics, however, has emphasized generally the theoretical aspects of techniques and models. As a result, search, which is crucial in data mining, has received little attention. In addition, interface to database, techniques to deal with massive data sets, and techniques for efficient data management are very important issues in data mining. These issues, however, have only begun to receive attention in statistics (Kettenring and Pregibon, 1996).

The use of statistics is a very important approach in Data Mining, however, it has (according to the author above) not been applied as proactively as it could be.

2.4.2.2 Artificial Intelligence (AI)

AI has received widespread attention in IT circles during the last decade, and it features much discussed technical subsets such as :

- Pattern Recognition
- Machine Learning
- Neural Networks

Other techniques in AI such as knowledge acquisition, knowledge representation, and search, are relevant to the various process steps in Data Mining.

One of the most important problems in applying Data Mining is *classification*. Classification is the process of dividing a data set into mutually exclusive groups such that the members of each group are as "close" as possible to one another, and the members of different groups are as "far" as possible from one another. For example, a typical classification problem is to divide a database of customers into groups that are as homogeneous as possible with respect to a variable such as creditworthiness. (Lee & Siau, 2001)

One solution to the classification problem is to use neural networking. According to Lu *et al.* (1996), a neural network-based Data Mining approach consists of three major phases:

1 *Network construction and training*: in this phase, a layered neural network based on the number of attributes, number of classes, and chosen input coding method are trained and constructed.
2 *Network pruning*: in this phase, redundant links and units are removed without increasing the classification error rate of the network.
3 *Rule extraction*: classification rules are extracted in this phase.

Other AI techniques applicable include :

- Case-based reasoning : This technique relies on previous cases to determine the best solution according to patterns.
- Intelligent agent : This technique uses a computer program (called an agent) to process and retrieve large amounts of data. An example of an agent is an internet search engine "crawler" or "spider" which sifts through thousands of web pages and indexes and finds relevant information.

2.4.2.3 Decision Trees

Decision trees are based on simple tree models where at each branch during tree growth the data set is strategically partitioned into different classes and subclasses. At each split, the most effective way of partitioning and classifying the data set is accomplished by using the most distinguishing feature encountered at that step of the algorithm. Decision trees are used to predict and/or classify. They can also be effective when the knowledge seeker desires a quick and dirty exploratory partitioning of the data set in order to get a gut feel for the data. (Gargano & Raggad, 1999)

Decision trees are constructed by a two-step process:
1 forward pass; followed by
2 backward pass.

The forward pass involves the decision maker in identifying the decisions to be made, the events that might occur and the sequence of decisions and events. The forward pass reveals the structure of the problem. During the forward pass, conditional profits must be calculated (i.e. the profits that will be achieved if certain strategies are adopted and certain events occur) and the probability of events must be assessed. The backward pass is concerned with analyzing the decision problem. Expected values are estimated by working backwards through the tree, as follows:

- For each set of event branches the expected value is calculated.
- For each set of decision branches the one with the highest expected value is selected.
- The strategy with the highest expected value is selected and \\ is drawn through lines representing the other options.

The final analysis should lead to a decision. Once this decision has been identified, a preferred strategy will emerge. (Coles & Rowley, 1995)

2.4.2.4 Visualisation :

Visualisation techniques are used to display data in a graphical / visual format to enable managers to interactively analyse datasets and draw conclusions from this data.

Tufte (1983) provided many examples of visualization techniques that have been extended to work on large data sets and produce interactive displays. There are several well-known techniques for visualizing multidimensional data sets: scatter plot matrices, co plots, prosection matrices, parallel coordinates, projection matrices, and other geometric projection techniques such as icon-based techniques, hierarchical techniques, graph-based techniques, and dynamic techniques.

2.4.3 Goals of approaches

The descriptions above are meant to provide an outline of some of the current approaches used in managing Data Mining activities. The list is by no means complete, and its purpose is simply to provide an overview as a precursor to the rest of the study.

Bearing the above in mind, it seems as if most of the approaches mentioned provide knowledge professionals with a way to extract patterns from large amounts of data, convert these patterns into trends, from which conclusions can be drawn about future trends to be fed to management – Data Mining is the search for valuable information in large volumes of data (Weiss and Indurkhya, 1998). It is the process of nontrivial extraction of implicit, previously unknown and potentially useful information such as knowledge rules, constraints, and regularities from data stored in repositories using pattern recognition technologies as well as statistical and mathematical techniques (*Technology Forecast*, 1997; Piatetsky-Shapiro and Frawley, 1991).

The explosive growth in data and databases results in the need to develop new technologies and tools to process data into useful information and knowledge intelligently and automatically. Data mining (DM), therefore, has become a research area with increasing importance (Weiss and Indurkhya, 1998; *Technology Forecast*, 1997; Fayyad *et al.*, 1996; Piatetsky-Shapiro and Frawley, 1991).

A significant advantage of data mining is the ability to seamlessly automate and embed some of the mundane, repetitive, or tedious decision steps not requiring continuous human intervention. While simultaneously reducing the costs and potential errors encountered in the decision-making process, such embeddings also allow for the automated preprocessing and massaging of the data to be used in the model. This implies that the practice of one time model development to find the solution of an *ad hoc* problem should be avoided if possible (Berson and Smith, 1997).

Data Mining can thus be understood to apply largely to processes involving knowledge. This knowledge may (as has been explained) be found in different forms. Humans inherently possess knowledge (Malhotra, 1998). We define knowledge as the understanding, awareness, or familiarity acquired through study, investigation, observation, or experience over the course of time. It is an individual's interpretation of information based on personal experiences, skills, and competencies. (Bollinger & Smith, 2001)

Different approaches to knowledge management, and the accompanying different approaches to data mining can provide the organization with particular challenges in the integration field.

2.5 FalconView Case Study :

Adapted from source : Gates (1999)

During the Gulf War of 1991, the United States was hailed as winning through the use of technology, employed through high-tech cruise missiles and radar-evading stealth bombers. The planning for these “technological” attacks on the other hand, was done without the large-scale use of technology. Realising this after the war, the US Air Force held conferences to determine how to better automate the strike planning process.

Rather than deciding to use proprietary technology, the Air Force decided on a PC-based solution, employing modular software architecture. The resulting system, FalconView, was developed in 18 Months for about $ 2.5 million.

FalconView reduces the mission planning process from 7 hours to less than 20 minutes on average, and increases accuracy immensely. The system uses digital data and aeronautical mapping technology to streamline the planning process.

Because FalconView uses advanced Data Mining technology, pilots can almost instantaneously locate targets and landmarks, speeding up the learning curve and accuracy of missions flown.

After the missions have been flown, the real advantages of FalconView come to light. Using Data Mining, the Air Force analyses patterns in missions through the software, and determines better options for future missions. This information is invaluable in helping the Air Force increase pilot safety and success rates.

Data Mining in its various forms is thus not only applicable in business situations, but can also mean the critical difference between life and death in the military environment.

2.6 Analysis and Synthesis :

Organizations have come to realise that Data Mining is an important facet of knowledge management, and that it provides management with a tool with which to make sense of the sometimes impossibly large amounts of data that they gather.

As part of the objective of the study relates to the current nature and status of data mining tools, it is prudent to note the classification of these tools into generalised management approaches.

Without classification, the technical nature of data mining tools would cause them to be seen as just that – Technical tools for the technically inclined. With classification, management can begin to understand the tangible benefits that these tools can offer.

Understanding the categories of tools available is thus the first step in the analysis of the nature and status of data mining tools.

2.7 Summary :

As a precursor to the analysis of specific data mining tools, this chapter provided an overview of the possible approaches to data mining available to the organization, of which four were overviewed.

These approaches are not necessarily mutually exclusive, as they imply utilization of different techniques and specific tools, which may or may not be interoperable.

The study now turns toward the specific Data Mining tools applicable in the approaches described, and an overview of their functionality and implentability.

Chapter 3 : Data Mining Tools

3.1 Introduction

Approaches to Data Mining (as described in chapter 2) mostly imply an outlined strategic approach towards the process, and do not necessarily specify the concrete tools used in the process. These tools are specific techniques used by the knowledge professional in enabling the Data Mining process, and are listed with an overview of their main characteristics. A brief analysis is then done according to their:

- Application areas
- Implementability
- Interoperability

The Data Mining process is usually a fusion of several different approaches and specific techniques, as data is rarely uniform in nature. Data differs in its suitability to be analysed with an approach or specific tool, and it is the knowledge professional's responsibility to decide which mining tools are best suited to specific data (and specific management needs). In this regard, the characteristics of the tools mentioned above are used to aid the decision making process.

In this chapter, a description of each specific tool is provided, followed by a comparative analysis on the basis of the characteristics bulleted above.

3.2 Specific Data Mining tools:

This section provides a description of the specific data mining tools available to the knowledge professional in enabling the data mining process. The logic and functioning of each tool is described

3.2.1 Expert Systems

Expert Systems are a form of artificial intelligence used to make decisions about specific subject matter. Expert systems can be divided into 3 main categories :

- "Standard" Expert Systems
- Fuzzy Expert Systems
- Real-Time Expert Systems

An Expert System usually consists of the following core components :

- A knowledge base – Facts concerning objects in the specific domain and their relationships.
- Facts – Data on the specific field.
- Inference engine – Processing ability which allows the system to process queries according to the rules and facts it has.

The system uses the rules programmed into it to make decisions about queries / information it is fed. These queries are usually handled by way of if / then / or / else statements, much the same as most computer programming operates.

Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth -- truth values between "completely true" and "completely false". Many expert systems incorporate fuzzy logic to enable the system to cope with gray / insufficiently descriptive data. Fuzzy logic also provides the system with intelligence to more effectively handle even complete data.

In 1951 Lotfi Zadeh (Zadeh, 1962), a professor of electrical engineering at Columbia University, wrote a paper called "System theory", which founded a new discipline of measuring hard data. However, Zadeh, in his studies of complex systems, found that human interactions in organizations defy predictable behaviours and, therefore, hard data are not always possible or accurate in tracking processes.

Expert systems can mimic, to some extent, the reasoning of experts whose knowledge of a narrow domain is deep. In this way, human experts and expert systems can arrive at similar conclusions. Since a Fuzzy Expert System is not as brittle as normal expert system, it can easily handle illogical complexities, poor clarity (in the facts and/or rules), or internal inconsistencies. Fuzzy expert systems are highly supervised, since like expert systems the knowledge engineer elicits fuzzy expert knowledge from either human experts through interviews or from textbook procedures. Knowledge elicitation is the most difficult and time-consuming aspect of developing fuzzy expert systems.

There has been substantial research in real-time expert systems (Laffey *et al.*, 1988). Real-time expert systems, as the name suggests, react to inputs and process them to provide feedback almost instantaneously. Real-time expert systems must often work in limited time and adapt to a changing environment, their input usually being sensor data.

3.2.2 Genetic Algorithms:

Genetic Algorithms are search algorithms which use principles inspired by natural genetics to evolve solutions to problems (Holland, 1975). The basic idea is to maintain a population of chromosomes, which represents candidate solutions to the concrete problem being solved, which evolves over time through a process of competition and controlled variation. Genetic Algorithms have got a great measure of success in search and optimisation problems.

A Genetic Algorithm starts off with a population of randomly generated chromosomes (solutions), and advances toward better chromosomes by applying genetic operators modelled on the genetic processes occurring in nature. During successive iterations, called generations, chromosomes in the population are rated for their adaptation as solutions and, on the basis of these evaluations, a new population of chromosomes is formed using a selection mechanism and specific genetic operators such as crossover and mutation. An evaluation or fitness function must be devised for each problem to be solved. Given a particular chromosome, a possible solution, the fitness function returns a single numerical fitness, which is supposed to be proportional to the utility or adaptation of the solution represented by that chromosome.

Genetic Algorithms are powerful in difficult environments where the space is usually large, discontinuous, complex and poorly understood. They are not guaranteed to find the global optimum solution to a problem, but they are generally good at finding acceptably good solutions to problems quickly. These reasons have been behind the fact that, during the last few years, genetic algorithm applications have grown enormously in many fields. (González & Fernández, 2000)

Genetic algorithms are thus a form of algorithm which is utilised in the solving of complex problems, such as extracting complex patterns from very large databases. Large and complex problems require a fast computer in order to obtain appropriate solutions in a reasonable amount of time. Mining large data sets by genetic algorithms has become practical only recently due to the availability of affordable high-speed computers. (Lee & Siau, 2001)

3.2.3 Clustering :

Clustering methods group together closely related data records in the database, thereby classifying them based on their similarities. Clustering usually involves data pre-processing (e.g. normalization) that insures every reduced data record (i.e. containing only those attributes to be analyzed) has an unbiased influence in the classification process. Clustering systems are highly unsupervised. (Gargano & Raggad, 1999)

Clustering is well suited to handling very noisy datasets, where it would be too time-consuming for a human to filter and classify all the data. Because of the inherently large amount of processing involved, clustering tools have relatively long training times.

One specific form of clustering is median-clustering, as in the Lance-Williams flexible method (Everitt, 1980) to fuse together into a cluster objects which exhibit the greatest similarity (within their condition attributes).

Clustering is one of the Data Mining tools which allow the knowledge professional to sift through extremely large amounts of data to find similar objects / items.

3.2.4 Neural Networks:

During the last decade, neural networks have gained a lot of popularity. Although the use of neural networks is new to a lot of application areas, the first neural network was developed in 1958 by Frank Rosenblatt. This network was called the *perceptron*. Rosenblatt was able to prove the convergence of the *perceptron learning rule* in 1962 (Rosenblatt, 1962).

A neural network is a collection of interconnected homogeneous processing units, called neurons (Caudill, 1991). Neuron activity is controlled by a continuous and differentiable mathematical function that aggregates input signals received from other neurons and produces output signals transmitted to other neurons. A neural network consists of an input layer, an output layer and many hidden layers.

The neural model is capable of understanding situations unknown to various knowledge-based systems of the organization's functional areas (Raggad, 1996). Neural Networks also help the organization to more easily access information stored in its knowledge base.

The lifecycle of a Neural Network consists of the following phases :

- Training
- Design
- Testing

The system is trained by feeding it with repeated sets of induced rules, which gives it its "learning ability". Training is the most important part of the implementation of a Neural Network, as it has the largest effect on the effectivity or "ability" of the system.

The use of a Neural Network can have profound implications for Data Mining. Neural Networks are essentially a form of AI (Artificial Intelligence), in that they become "smarter" every time they encounter new and more challenging problems. A neural network is only limited in its capacity to solve problems by the amount of problems it is presented with – As it solves more problems (or mines more data), it becomes even more adept at the task, thus is very well suited to extremely complicated Data Mining applications where a degree of adaptability and intelligence is constantly needed.

3.3 Comparative analysis of Data Mining tools:

This section of the study provides a comparative analysis of the tools so far described. Comparisons are provided in :

- Application areas
- Implementability
- Interoperability

3.3.1. Application areas

As is the nature of most tools, specific data mining tools are suited to specific applications. This section analyses and compares the different areas in which the listed tools can be applied.

3.3.1.a) Application areas of Expert Systems

Expert systems are tools which are often assumed to be a substitute for human interaction and expertise. Management may try to replace certain specialist knowledge workers with expert systems, taking into account only the superior processing ability that expert systems have. This is however a mistake. Gargano and Raggad (1999), list the following limitations :

- Although the most difficult and time-consuming aspect of developing a good expert system is attaining knowledge from the experts, Expert Systems are only as good as the knowledge captured through them.

- Expert systems are not very robust since they are notoriously brittle and cannot easily support illogical complexities, poor clarity (in the facts and/or the rules), or internal inconsistencies.

- Expert Systems are not easy to scale, that is, if more rules are added to the knowledge base (or even if the rules were rearranged) unforeseen results may occur.

These limitations are inherent in the way Expert Systems operate, and do not necessarily mean that they are not applicable in specific applications. Often management makes the mistake of implementing an Expert System in the wrong context, applying it to problems to which it is not suited.

Expert Systems are however well suited to domains consisting of cognitive based, very specific expertise. The knowledge should be :

- Easy to capture
- Self-consistent
- Simple to explain
- Straightforward to represent
- Avoid dependency on common sense

Thus expert systems are well suited to very specialised deep knowledge areas, as opposed to areas which require a wide general knowledge and common sense. Such a system would for example be well suited to handling queries about a specialised aircraft jet engine and all the technicalities it represents.

3.3.1.b) Application areas of Genetic Algorithms

Genetic Algorithms are usually applied in handling large very complex areas of knowledge, and are often expected to "find their own way around". It is important to remember that genetic algorithms are best applied in finding a "general" solution to a problem very quickly, using high-speed processing as its main resource.

Genetic Algorithms do not however, always find the best solution to the specific problem at hand, and should often be used as the first step in determining a solution.

Two different scenarios where Genetic Algorithms could be applied successfully are:

- Single objective problems - Genetic Algorithms design many solutions until no further improvement can be achieved or some predetermined number of generations has evolved or when the allotted processing time is complete.
- Multi-objective problems – Genetic Algorithms give out many satisfactory solutions in terms of the objectives, and then allow the decision maker to select the best alternative.

Within these areas, Genetic Algorithms are very effective. The main advantage is that in their training phase, GA's require no user intervention or supervision.

Genetic Algorithms can thus be said to be best suited to problems which require solving within the shortest possible time, or problems which give management an informed choice between a few "best" options.

3.3.1.c) Application areas of Clustering

Clustering is a technique which can be applied to most data types, depending on the complexity and spread. It is a technique which can be applied most successfully to the following situations :

- Classification of data – Specifically situations where categorization is difficult, for example data with a large diversity, which a human analyst would find difficult to handle because of having to learn such a wide range of classifications in multiple dimensions.
- Problems where scaling capabilities are required.
- Situations where ample training time is available

Clustering techniques are thus best suited to mining problems where the organization is unsure of the exact scope of the possible solution, such as finding the best way to provide a new service (which has not been provided before). Also, where training (setup) time is available beforehand, the technique can provide improved results, as well as being highly adaptable throughout the implementation phase.

3.3.1.d) Application areas of Neural Networks :

The application of Neural Networks is an area which is under constant development, as the technology behind it is still developing rapidly. Neural Networks have often been touted as "Intelligent Systems", largely because they form a subset of Artificial Intelligence. In extracting information from databases or Knowledge Base Systems, as explained by Raggad (1996), Neural Networks are gaining popularity.

A major factor affecting success in planning and managing knowledge in Knowledge Base Systems is knowledge presentation. Users, experts and knowledge engineers should have the same objectives in each of the roles they play in the knowledge acquisition process. Nonetheless, this is difficult to achieve because of the heterogeneity of the knowledge in the specific domain of the Knowledge Base System (Raggad, 1996).

Neural Networks are best suited to applications where input is continuous, and a lot of adaptation is needed on the mining tool's side. For example, a database of internet search-engine trends needs the continuous learning ability of a Neural Network in order to retrieve relevant information. In today's fast-developing information world, most databases are moving towards continuous updating, and the intelligence of Neural Networks will make them the main tools through which mining will take place.

3.3.2 Implementability

As with any organizational strategy or process, Data Mining has to be practically implementable for it to provide value to the organization. For the purposes of this study, it is assumed that the organization has the following in place :

- Information Systems (IS) Infrastructure
- Databases of sufficient size to warrant Data Mining
- Data Mining goals

As most of the tools discussed are technological in nature, and make use of the Information Systems infrastructure, it would be prudent to take into account the impact each tool might have on the IS infrastructure.

Furthermore, it is assumed that the IS Infrastructure consists of the following components :

- Information Servers (in the form of file & database servers, web servers etc)
- Networking capabilities (LAN, WAN, Intranet etc)
- Workstations
- IS Support personnel

Databases are applications which run on their own, usually with their own dedicated physical servers.

This infrastructure is crucial to the successful implementation of any Data Mining strategy, and the impact on it from Data Mining tools has to be taken into account. The tools affect the components of this infrastructure as follows :

3.3.2.a) Implementability of Expert Systems

Expert Systems function at an application level. This means that an expert system is an entire program on its own which uses its own resources on a workstation or server to query the organization's database systems. Expert Systems are found in two forms :

- Query
- Search

Query Expert Systems function as an application which accepts questions. These questions are then processed through the inference engine and database, and an answer is output. This processing often occurs at both workstation and server level, and needs the user to specifically "ask" the system questions.

Search Expert Systems function as applications, but questions are compiled at one point and run through the system constantly or periodically. This form of the system does not require constant user input to be able to function, and can run on the server alone, without needing a workstation.

In both these cases, and for fuzzy as well as standard expert systems, user intervention (or "supervision") is high. Fuzzy expert systems are highly supervised, since like expert systems the knowledge engineer elicits fuzzy expert knowledge from either human experts through interviews or from textbook procedures. Knowledge elicitation is the most difficult and time-consuming aspect of developing fuzzy expert systems. (Gargano & Raggad, 1999)

The impact on infrastructure of Expert systems is thus relatively high – Constant user intervention is required, and the system often uses processing power on both a server and workstation level.

3.3.2.b) Implementability of Genetic Algorithms

Genetic Algorithms as systems usually run as programs in the background, conducting searches through specified databases. These programs are often referred to as "agents, crawlers, or spiders", and function on a server level. Genetic Algorithms are given search criteria, and can search through vast amounts of data rapidly, providing search results.

Genetic Algorithms are used by internet search company Google to provide rapid and extremely effective World Wide Web searching capabilities. The search services provided by Google enable a user to specify criteria, and receive results which are relevant to the query, and not just through a large number of keyword matches.

This technology is very important in terms of ease of use, and can traverse infinitely large amounts of data very quickly.

As user intervention is limited to simple search criteria specification, and the program runs as an agent at server level, impact on the IS infrastructure is relatively low. Because of the large amounts of data searched on the internet however, bandwidth use for the search itself may be astronomical. This is however determined by the scope of data to be searched through, and not necessarily by the tool itself.

3.3.2.c) Implementability of Clustering programs

Clustering programs, like genetic algorithms, also usually consist of search agents which conduct searches for clusters according to specified criteria. The programs are highly unsupervised, and after being given the criteria, conduct the processing on their own.

As clustering programs usually run as agents, the use of processing power is not massive, but still more noticeable than Genetic Algorithms. Because clustering programs perform a more thorough (and sometimes even elaborate) processing of the possible results, the server and bandwidth infrastructure is taxed more.

The overall impact of clustering programs on IS infrastructure is thus moderate, with very little impact on the user side, and a moderate processing impact because of sometimes complex searches.

3.3.2.d) Implementability of Neural Networks

As Neural Networks are in effect intelligent applications, they are run at application level, and may even require dedicated server infrastructure. Neural Networks require a substantial amount of "training" (initial input simulation and problem solving supervision by operators) in order to become "intelligent".

Once a Neural Network is up and running however, user intervention is limited to queries, and the system functions on its own. As it improves in ability to process and solve problems with an increase in queries, the system uses more processing power to continue "learning".

The impact Neural Networks have on the IS infrastructure could be quite substantial, depending on the complexity of the system and the demands for problem solving placed on it by users. As Neural Networks are the state of the art in Data Mining, one would expect organizations to invest in the costs of constantly upgrading these systems.

In summary, the infrastructure effect of a Neural Network is potentially substantial, both on bandwidth and processing power, and initial user "training".

3.3.3 Interoperability of Tools

The choice an organization makes to implement a Data Mining strategy is an active one, and the practical tool implementation may change more regularly than might be expected. As Data Mining tools are generally highly technical in nature, constant development of these tools takes place, at the same fast pace as other technology developments.

In the current technology environment, it is extremely rare to find any software / hardware operating independently without a connection of some sort to other systems. Even personal computers in the home are connected on demand to the internet, without much technical knowledge on the user's side. Thus these data mining tools seldom operate in complete independence, and it is prudent to consider the effects they might have on each other.

3.3.3.a) Interoperability of Expert Systems

As Expert Systems operate on an application level, they are able to function highly independently to other systems. All an expert system needs to function are rules, a database, and its own inference engine. This means that the system can run as a standalone system on a workstation handling queries from operators / users. It was mentioned previously that it is in the current technological environment extremely rare for any system to operate completely independently, but depending on the type of data, for the Expert System it is technically possible and even viable.

Even though independence is possible, the large majority of Expert Systems will have their database updated constantly through a network connection, which means that they are actually operating in an interdependent system.

Expert Systems can thus operate either independently, or in interdependently with other systems.

3.3.3.b) Interoperability of Genetic Algorithms

Genetic Algorithms process data from a variety of sources, only acting as a "search engine" for these data. As Genetic Algorithms run as agents, and not applications, they do not have their own databases or interfaces.

Genetic Algorithms are thus highly dependent on other systems, and can be implemented to run in the background (in agent form, as described). Genetic Algorithms can thus run interdependently with other systems, performing the search functions allocated to them as needed.

3.3.3.c) Interoperability of Clustering programs

Like Genetic Algorithms, Clustering programs run as agents, performing searches in the background. They do not have their own databases, and perform searches through outside databases specified to them.

These systems are thus also highly dependent on a platform on which to run, even though they run in the background. Clustering programs are a form of search engine, much the same as a Genetic Algorithm. The knowledge professional will find that a choice will often have to be made between either a clustering program, or a genetic algorithm, depending on the type of data searched through and the results needed.

3.3.3.d) Interoperability of Neural Networks

Neural Networks are systems which can run completely on their own, and often require their own infrastructure because of substantial processing needs. These systems can however operate in conjunction with other systems, depending on the specific needs of the organization.

As the design of Neural Networks is quite an involved process, the knowledge professional would be well advised to consider the Network's interaction with other systems when designing the initial architecture.

Neural Networks are thus very flexible, and can operate either independently of or interdependent on other systems, all depending on the architecture and data mining needs of the organization.

3.3.4 Interoperability Case Study : K-12 School databases
Adapted from source : Tillett (2000)

American School districts are starting to invest millions of dollars on business-intelligence tools to monitor student activity and progress.

As school administrators are made increasingly accountable for student performance, they are starting to realise that business intelligence tools can provide them with valuable information on educational trends.

Large K-12 school districts such as Gwinnett County, Ga., Broward County, Fla., and Clark County, Nev., are investing in data technologies to the tune of millions of dollars. Gwinnett County, for one, is halfway through a three-year data warehousing project worth nearly $11 million, according to a county official.

Although many schools cannot afford such large investments in technology, those that can seem to find it worthwhile.

IBM this month began marketing a suite of services designed to help school districts sift through the data they accumulate in stovepiped legacy systems and paper files-and to turn that data into usable information. The key in this exercise is merging information from different databases into a single stream, requiring interoperability of data mining tools and techniques.

"Typically, districts are information-poor and data-rich," said Jane Lockett, an IBM Global Services executive who specializes in business intelligence and education. Gwinnett County, moving measuredly toward its vision of a robust data warehouse full of useful information for gauging performance, next month will complete the third phase of its data warehouse project. It will roll into the warehouse information on students' performance on standardized tests such as the Preliminary Scholastic Aptitude Test, the Scholastic Aptitude Test and the Gateway Test, as well as student performance in programs such as English as a Second Language (ESL).

With the facilities provided by data warehouses, teachers and officials can analyse possible patterns in many areas, ranging from general educational performance trends to individual teacher or student details.

Warehousing and mining the data takes some guesswork out of running a school, said Jim Woolen, chief information officer for Gwinnett County Public Schools.

"When you put it into a data warehouse, you can start tracking trends," Woolen said. It also saves time over having county programmers pull data for administrators, he said.

Once the data warehouse (together with data mining tools) has provided answers to specific questions, school principals and other stakeholders can make decisions based on this information – such as reassigning teachers or changing specifics of a curriculum. The information provides a concrete justification for these decisions and possible expenditure.

Data warehousing and the accompanying data mining strategies are not usually seen as projects a school would engage in, as schools are not for-profit businesses.
But Nancy Stewart, an analyst who follows data warehousing for Survey.com, said business-type thinking for schools makes sense. "What they are faced with in their own organizations is in a lot of ways not different from what's faced in a business organization," she said.

The education community's interest in data warehousing has fallen largely on universities with both the funding and development expertise, according to Stewart. So, for now, vendors such as IBM are concentrating on large school districts with deep pockets, since a data warehousing and mining project might cost as little as $100,000 or as much as several million dollars, IBM officials said.

Data mining tools can thus operate interdependently, but a concerted effort has to be made to effectively integrate different systems. Technologically, it is almost always possible to integrate information from different platforms, with practicality usually being the only limiting factor.

3.4 Analysis and Synthesis :

Data Mining tools are highly technical tools which can be applied to meet almost any data mining needs within the modern organization. The majority of these tools require a substantial investment in an information systems infrastructure. In order to justify this investment, the organization needs to carefully look at its specific knowledge needs, and also the approach that it will follow in implementing these tools.

Most organizations will find that implementing a set of data mining tools is not as easy as it may seem at first glance, as these systems are highly specialised and usually need to be tailored to the exact specifications of the organization. In this regard, implementation of a combination of tools would be the best approach, as specific tools very seldom provide a total solution to the knowledge extraction problem at hand.

It seems logical that any organization which considers applying specific data mining tools should embark on an exhaustive analysis of its knowledge needs. More importantly, in the organizational learning context, the organization should carefully determine its *learning* needs before implementing specific data mining tools, as these tools can affect the knowledge outcome substantially. This chapter thus concludes the part of the objective concerned with the current nature and status of data mining tools.

3.5 Summary

This chapter provided an overview of the specific data mining tools available, together with an analysis of their implementability and interoperability. The study now turns towards the construct of organizational learning.

Chapter 4 : Organizational Learning – The Facilitating Factors

4.1 Introduction

4.1.1 Purpose of Organizational Learning

The rapid increase in competitive behaviour, widespread communications diffusion, and rapid technology development have all given rise to a need for organizations to adapt even more quickly than was once thought possible or necessary.

Organizations evolve over the years in accordance with what they have learned is necessary to operate in business. Organizations develop "best way" principles for planning, organizing, and controlling their day-to-day operations, based on yesterday's experience and knowledge base. This works in stable business environments since the gap between yesterday's knowledge and today's business demands are small. If the environment in which one operates today is evolving at ever increasing rates, then the way in which the organization learned and adapted in the past will not meet today's needs (Appelbaum & Reichart, 1998). Reg Revans, originator of the equation L>=EC stresses that learning must be equal to or greater than environmental change, or the organization will die (Hitt, 1995).

4.1.2 Organizational Learning defined

Organizational Learning as a field has received a large amount of attention during the past decade. It is a concept which defines how an organization disseminates information, and utilizes this information in improving its ability to perform organizational activities with maximum effectiveness. Without going into a detailed analysis of the field of Organizational Learning, a few authors' definitions of the concept are :

Argyris and Schön (1978) state that: "Organizational learning occurs when members of the organization act as learning agents for the organization, responding to changes in the internal and external environments of the organization by detecting and correcting errors in organizational theory-in-use, and embedding the results of their inquiry in private images and shared maps of organization". Furthermore, Argyris and Schön (1996) state that "an organization's learning system is made up of the structures that channel organizational inquiry and the behavioural world of the organization, draped over these structures, that facilitates or inhibits organizational inquiry".

Some authors have divided Organizational Learning into "old" and "new" definitions. Both old and new organizational learning would probably be seen as processes. Old organizational learning is about individuals learning as agents for the organization (e.g. Argyris and Schön, 1978). New organizational learning also means learning, by a collective (Cook and Yanow, 1993) or by humans as social beings (Brown and Duguid, 1991).

Peter Senge has put forward an organizational architecture and the concept of an implicate order and learning results (Senge *et al.*, 1994). For Senge, the learning organization appears to represent a combination of three architectural design elements, which are:

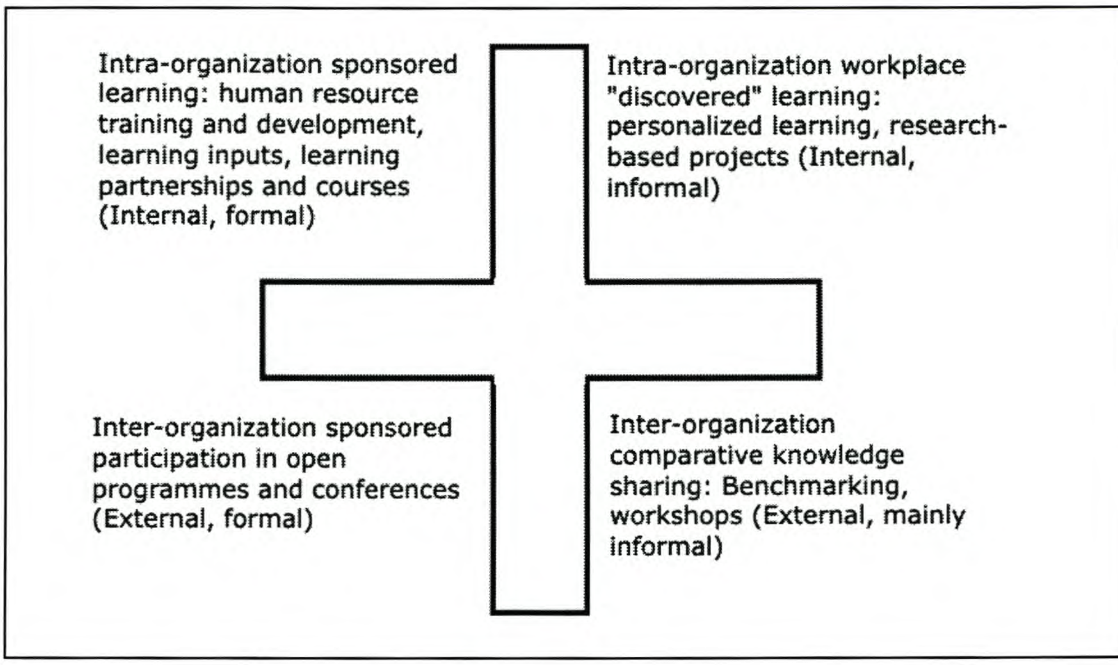
(1) guiding ideas;
(2) theory, methods; and
(3) innovations in infrastructure (Senge *et al.*, 1994).

It could be said that the goal of organizational learning is to create a learning organization. Although the majority of the literature written about the learning organization tries to define the general characteristics; it has tended to produce definitions that focus on:

- the importance of acquiring, improving and transferring knowledge;
- facilitating and making use of individual learning; and

- modifying behaviour and practices to reflect the learning. (Appelbaum & Reichart, 1998)

Figure 4.1 : Dimensions of organizational learning



Source : Adapted from Lorange (1996)

As can be seen in figure 4.1, there are different dimensions to organizational learning, including inter and intra-organizational dimensions. This implies that organizations can learn from each other as well, above and beyond learning within themselves.

The ultimate goal of organizational learning is to create a learning organization, one which has a sustainable competitive advantage because of its capabilities of continuous innovation. A learning organization is an organization skilled at creating, acquiring and transferring knowledge, and at modifying its behaviour to reflect new knowledge and insights. (West & Burnes, 2000)

As is the view of West & Burnes (2000), organizational learning is linked to individual learning, and there are stages that both the individual and the organization go through in the learning process. This view is illustrated in table 4.1. The ultimate goal, according to West & Burnes, is independency between the two.

Table 4.1 : Individual and Organizational Learning

	Individual learning ability	Organizational response
Stage 1	Foundation stage The individual is ready to learn; shows interest in acquiring the skills to learn; involvement in learning activities	Dependency stage The organization offers formal job training; remedial education; introduction to teamwork
Stage 2	Formation stage Self-development; independent learning; role interdependence; interest in teamwork	Transitional stage The organization offers job rotation and shadowing; wider industry training; opportunities for teamwork; experiential learning
Stage 3	Continuation stage The individual is self-motivated; has achieved independence as a learner; has developed a questioning approach; demonstrates autonomy at a group and individual level	Independency stage Organization offers linked career planning; shared responsibility for production and investment goals; broad commitment to work group autonomy

Source : West & Burnes (2000)

Organizational learning could thus be said to be an activity which focuses on acquiring, transferring and using knowledge to improve organizational effectiveness.

4.2 Organizational Learning's facilitating factors :

In order to make it possible for an organization to learn, an infrastructure has to be put into place which makes it possible for individuals and groups within the organization to engage in these "processes" of learning. This infrastructure is made possible by several facilitating factors defined by Appelbaum & Reichart (1998). These factors are :

4.2.1 Scanning imperative and performance gap

This factor measures whether there is an understanding or comprehension within the organization of the environment within which it functions. Also, how are analyses done of the differences / gaps between actual performances and targeted performance ? Also, if the organization determines that there is a gap (of whichever size) between the actual performance and what was targeted, the question is whether this realisation brings a move towards experimenting and development of new skills and insights. Mainframe computer manufacturers Cray, Unisys, IBM and the US auto companies in the 1970s failed to respond to developing changes that sound investigative work would have made visible (Nevis *et al.*, 1995). An awareness of gaps in performance helps the organization develop new strategies to try to acquire resources / skills that will help it close the performance gaps. The gaps in performance however have to be measured, and it is this measurement which is made easier by Data Mining.

4.2.2 Concern for measurement

How focused is the organization on measurement ? If the organization has put in place measurement tools, it has to actively use these tools and constantly develop them to adapt to changing circumstances. Many standard measurement tools are available, but they might not be suited to the organization's needs. Also, the areas in which measurement takes place are very important. Often, focus is placed on "traditional" business performance areas, while ignoring important areas such as quality and research & development. In this area, organizations introduce what is referred to as "balanced scorecards", where measurement is directly related to the results management would like to see. The correlation between what we measure and what we get is not perfect, but it is definitely positive (Hitt, 1995).

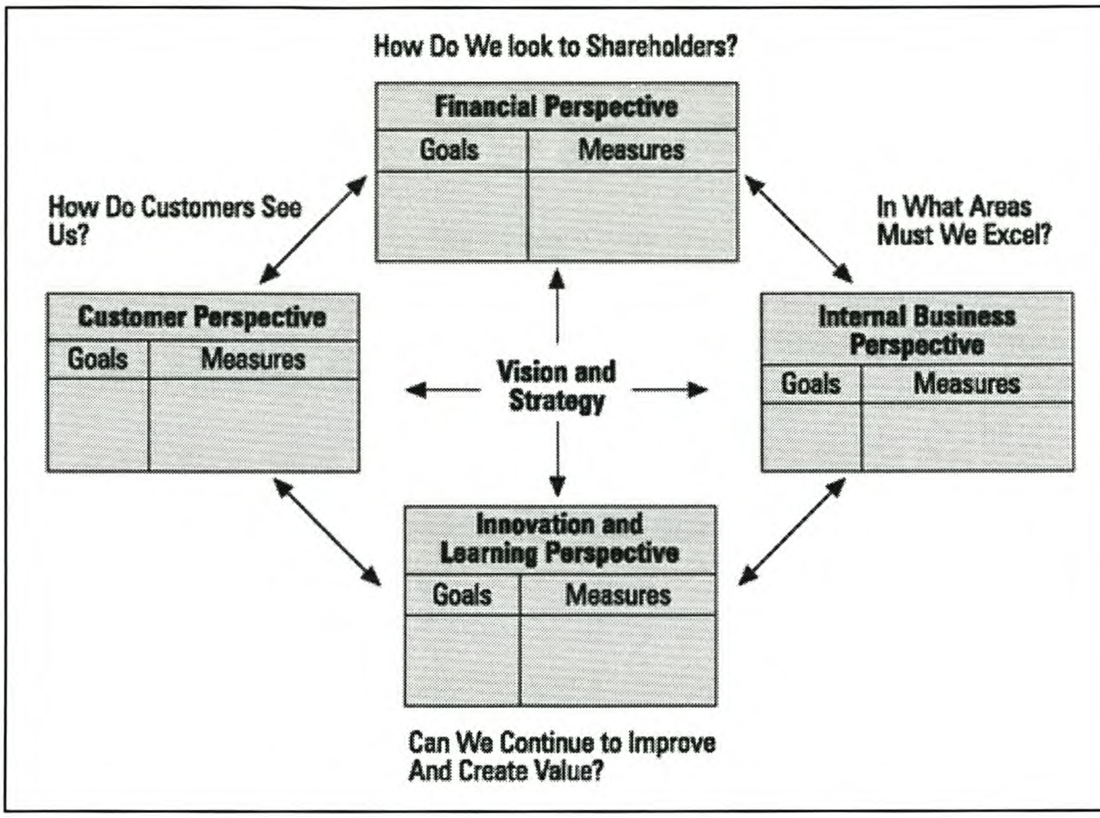
The balanced scorecard provides a systems perspective to the measurement process. The scorecard also provides priority to vision and strategic intent, and focuses less on operational goals / performance. Kaplan and Norton (1993) made these major points about the balanced scorecard :

- The organization's measurement system strongly affects the behavior of employees.
- No single measure can provide a clear performance target or focus attention on the critical areas of business.

- Managers need a balanced presentation of both financial and operational measures.
- The balanced scorecard complements the financial measures with operational measures.
- Operational measures should focus on customer satisfaction, internal processes, and the organization's innovation and improvement activities.
- The operational measures are the drivers of future financial performance.
- The balanced scorecard forces managers to focus on the handful of measures that are most critical.
- The scorecard brings together in a single management report many of the seemingly disparate elements of a company's competitive agenda.
- The scorecard guards against sub-optimization.
- The scorecard is a way to clarify, simplify, and then operationalize the vision at the top of the organization.

The balanced scorecard seems to be a popular way of looking at measurement, bringing into focus the true long-term goals related to vision and strategy. The organization has to carefully look at its goals when setting up a balanced scorecard. Balanced scorecards seem to have been effective so far because they are tailored to the exact needs of the specific organization, and not to a "type of organization".

Figure 4.2 : The balanced scorecard



Source : Adapted from Kaplan & Norton (1993)

4.2.3 Experimental mindset

The ability of a person to experiment relates directly to their ability to discover new skills and methods of doing things. The same holds true for the organization, which needs to expand its horizons and not become complacent. It is important to make the distinction between experimentation as a "desperate" effort to overcome difficulty (problem solving), and experimentation as a way to broaden horizons and discover new opportunities for future exploitation. There has been no significant evidence of experimental mindsets in industry with some notable exceptions at Motorola and Wal-Mart (Garvin, 1994).

Experimentation is usually divided into two forms :

- Ongoing programmes
- Demonstration projects

Ongoing programmes usually involve a series of small projects meant to improve organizational knowledge in small steps. This is a safe option, as experimentation often involves large costs, which could be negative if a large project is aimed incorrectly at the first stages, and continues along the original guidelines. Small ongoing programmes divide experimentation into blocks with separate measurable outcomes, constantly monitoring progress and direction. Ongoing programmes often manifest in the following forms (Appelbaum & Reichart, 1998) :

- statistical methods, i.e. design of experiments, that efficiently compare a large number of alternatives;
- graphical techniques, i.e. process analysis, that are essential for redesign in work flows; and
- Creativity techniques, i.e. story boarding and role playing, that keep novel ideas flowing.

Demonstration projects are usually of a more complex nature than ongoing programmes. They are designed to create organization-wide changes, and are often designed to bring about development of new organizational abilities. These projects are usually implemented at one site to test their implementability, and after validation of their success are extended organization-wide.

Demonstration projects are often a radical departure from "the way things are usually done", and require a change in thought patterns from some parts of the organization. Once again, Appelbaum and Reichart (1998), state the following about demonstration projects :

- They are usually the first projects to embody principles and approaches that the organization hopes to adopt later on a larger scale. For this reason, they are more transitional efforts than endpoints and involve considerable "learning by doing". Midcourse corrections are common.

- They implicitly establish policy guidelines and decision rules for later projects.
- They are normally developed by strong multi-functional teams reporting directly to senior management.
- They tend to have explicit strategies for transferring learning to the rest of the organization.

Both these forms of experimentation essentially have the same purpose – Creating understanding of the organization's purpose.

4.2.4 Climate of openness

An organizational culture of allowing members almost full access to all information has to prevail if sharing knowledge is going to be a successful initiative. Knowledge is only valuable if it is applied, sharing being an important catalyst. This climate of openness has to be applied from the top down, with top management setting an example and providing the vision to enable people at all levels within the organization to get used to sharing valuable intellectual assets in the form of knowledge with others.

Companies must review their successes and failures, assess them systematically, and record the lessons in a form that employees find accessible. This process is referred to as the "Santayana Review", citing the philosopher George Santayana, who coined the phrase, "Those who cannot remember the past are condemned to repeat it" (Garvin, 1994).

4.2.5 Continuous Education

Organizations which successfully learn continuously do not manage to do so just by implementing constant learning programs. Learning is not necessarily always a process which can be defined as a "business activity", and a large part of learning occurs spontaneously. It is this part which is often important, and the organization which instils a culture of learning will find that its members learn more effectively and routinely.

This "sense of learning" has to be present in all facets of the organization for it to be a true learning organization. This factor is another way of expressing what Senge (1990b) calls "personal mastery".

Highly skilled professionals are frequently very good at single-loop learning. They have spent much of their lives acquiring academic credentials, mastering one or a number of intellectual disciplines and applying those disciplines to solve real-world problems. This fact explains why so many professionals are often inadequate at double-loop learning. Because many professionals are usually successful at what they do, they rarely experience failure. (Appelbaum & Reichart, 1998)

Because these skilled workers rarely experience failure, they have often not learnt how to learn from failure. This is a critical factor in determining successful learning, and these skilled professionals need to expand their knowledge to include a wider range of organizational topics, working in teams, and moving towards reflecting on their own learning experiences in order to change and improve on their organizational behaviour.

4.2.6 Operational Variety

Businesses that focus on one specific set of operational processes often find that their competitors quickly find better ways to do things. If there is too much rigidity in operational process design, the organization might find that its members become so used to one specific process set that they cannot adapt quickly enough when needs change. On the other hand, organizations which are flexible and support variety in operational process design, policies and structures, might find that employees are eager to experiment and come up with improved ways of doing things.

The fundamental principle of organizational design at Japanese companies is redundancy, the conscious overlapping of company information, business activities, and managerial responsibilities. The term redundancy with its connotations of unnecessary duplication and waste sounds unappealing. Yet building a redundant organization is essential in creating a learning organization (Ikujiro, 1991). Redundancy provides the organization with inherently improved learning because employees have to constantly communicate and share knowledge to improve their own operational processes, while at the same time improving organizational processes as a whole.

Operational variety is thus an orientation which improves the "circulation" of learning processes throughout the organization.

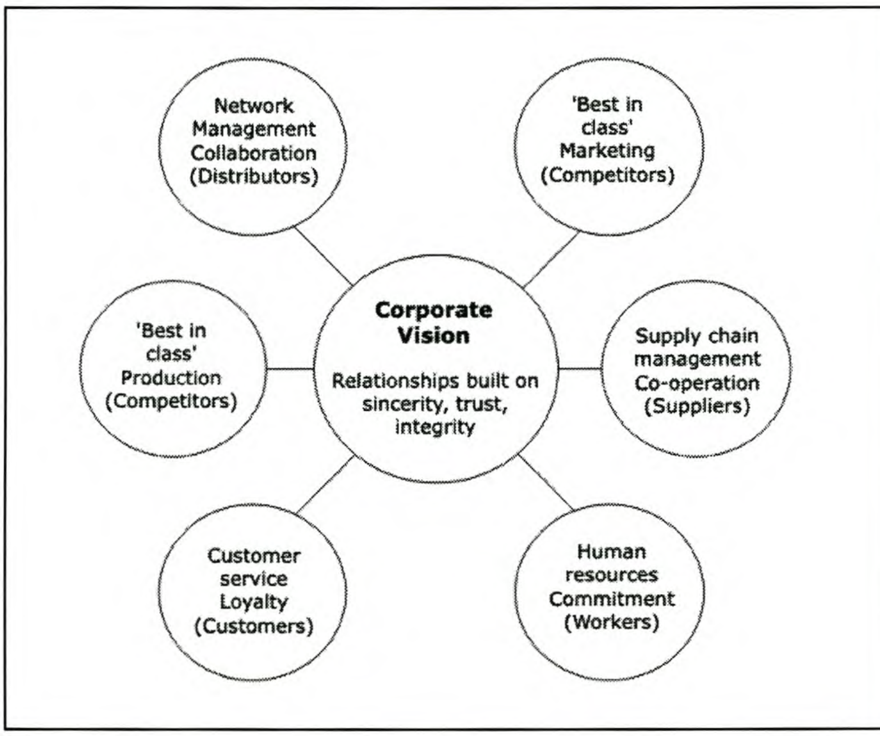
4.2.7 Involved Leadership

As with any organizational activity, implementation is only going to take place if leadership is provided from the highest levels. Leaders have to become visible role-models at all levels if employees are going to adopt the vision communicated by the leaders. Also, having multiple role-models will substantially increase the organization's success in promoting a learning culture. These role-models can be present in different levels of the organization, and need not be traditional or hierarchal leaders as such. Social leaders can also provide a large incentive and motivate employees to adopt the culture.

Peter Senge raised a provoking question "Who is considered the real leader of a ship ? Is it the captain, helmsman, or the engineer ?" Senge's answer is the ship's designer (Senge, 1990a).

The leader thus has to develop a vision, provide members with the tools to achieve this vision, and be a "designer" who constantly learns. In essence, the leader has to design successful learning experiences (Appelbaum & Reichart, 1998). The ways in which this corporate vision relates to core business activities are shown in figure 4.3.

Figure 4.3 : Relating corporate vision to core business activities



Source : Adapted from Chan (1994)

4.2.8 Systems Perspective

For any collective to function successfully, it needs a system on which to base its actions. Each member of the organization should in effect base his / her actions on the impact that it will have not only on their own performance, but the performance of the organization as a whole. Employees thus need to learn to foresee the effect of their actions not just within their domain, but with as much consideration for the larger interconnecting system.

A system is defined by Skyttner (1996, p. 35) as: "A set of interacting units or elements that form an integrated whole intended to perform some function". Miller (1995), described it as "A set of interacting units with relationships among them".

Ackoff (1981) set certain conditions for a system to be qualified. These are :

1. The behavior of each element has an effect on the behavior on the whole.
2. The behavior of the elements and their effects on the whole are interdependent.
3. However, subgroups of the elements are formed, each has an effect on the behavior of the whole and none has an independent effect on it.

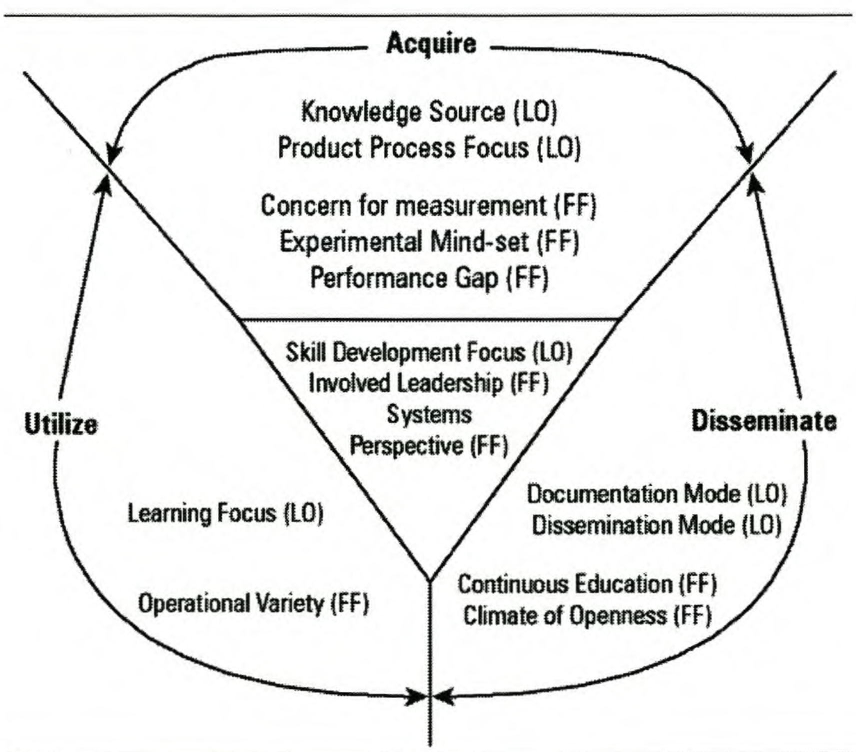
A system must thus be seen as a whole, and cannot be defined as such if it is divided into independent parts.

Once employees start having a systems perspective, they realise that learning constantly will improve their ability to contribute more effectively to the system, and also gain more from it. A systems perspective is thus a positive facilitator which needs to be present if an organization is going to take full advantage of the learning tools at its disposal.

4.3 Conclusions on facilitating factors

Taking into account the definitions of the facilitating factors listed, it is clear that any organization which wishes to learn has to, through a concerted effort, provide its members with the necessary infrastructure to do so in a variety of dimensions. This infrastructure forms the core of the learning approach and orientation, and needs to be carefully designed and implemented from its inception. Even if the organization does not necessarily have a specific goal in mind when approaching such a task, it needs to realise that flexibility and involvement at all levels is of paramount importance.

Figure 4.4 : Elements of the learning process



Source : Argyris (1977)

In figure 4.4 above, the facilitators are graphically represented as elements needed in the various phases of the learning process. They can be seen in this way as the infrastructure of the learning organization.

When looking at the infrastructure to be put into place, it is important to realise that a large part of the infrastructure has to be manifested in tools which can be applied and measured. As this study is concerned with the use of data mining tools in enabling organizational learning, the focus falls on infrastructure which relates to the application of specific data mining tools. These tools have been found to be largely technological and information systems related in nature, and thus the focus falls on the infrastructure put into place which is dependent on information systems for its effectiveness.

4.4 Infrastructure for improving learning

There are two general strategies for improving the learning process in an organization (Huber, 1991). One is to accept the existing style and improve its effectiveness. This strategy develops the fundamental part of the organization to its fullest extent, and builds on the notion of full acceptance of what has been accomplished to date.

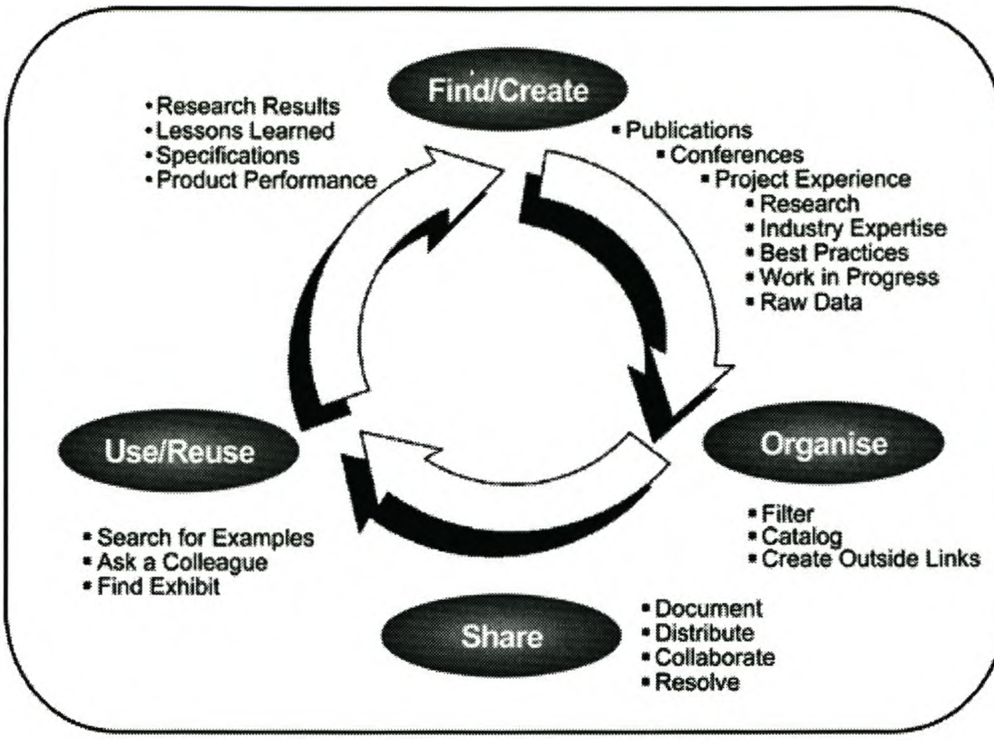
The second is to fundamentally change the learning orientations of the organization. What has been accomplished to date is not necessarily built on, and this strategy is sometimes seen as an attack on the organization's basic values.

Each strategy is approached through implementing a process with 3 main steps :

1. knowledge acquisition
2. knowledge sharing,
3. knowledge utilization.

This knowledge cycle is relevant when looking at all knowledge management activities, regardless of whether Data Mining is in fact involved in the process. This study does not have its focus on the broader range of Knowledge Management activities as such, but it is prudent to refer to the knowledge cycle. This cycle has been visualized by Burk (1999) in figure 4.5:

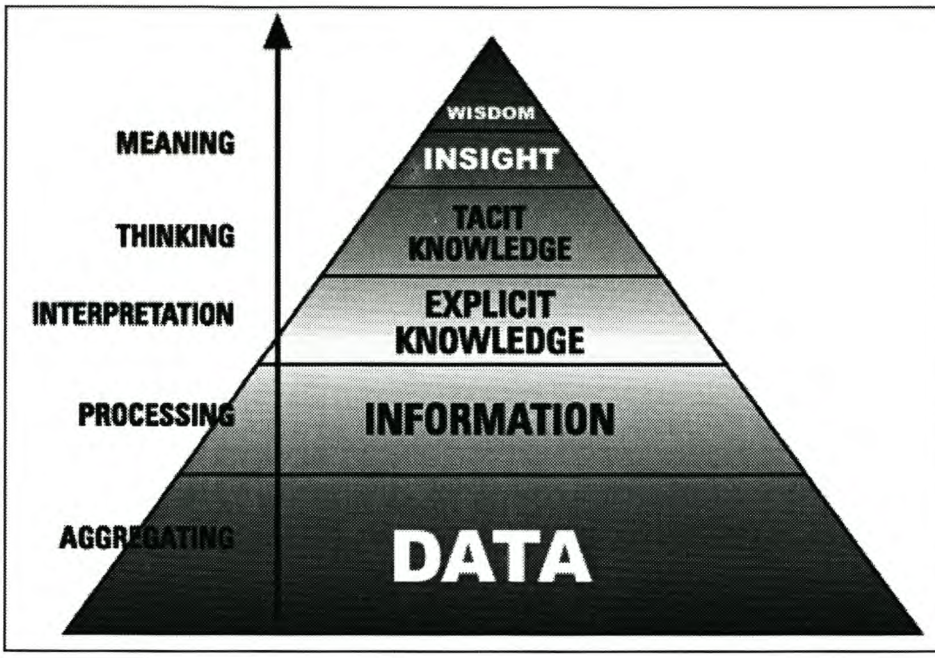
Figure 4.5 : The Knowledge Cycle



Source : Burk (1999)

It is important to mention the knowledge conversion cycle, as all knowledge-management related activities are usually built with this cycle in mind. The cycle has various forms and goes through various stages, but the basic components of it remain relatively constant. With this study more specifically in mind, moving from data to knowledge, the process has also been described by Rollo & Clarke (2001) in figure 4.6 :

Figure 4.6 : Data, Information and Knowledge



Source : Rollo & Clarke (2001)

As can be seen in *figure 3*, the ultimate goal of the process is to attain wisdom, which is only possible once insight into tacit knowledge has occurred. Data mining could be relevant at several stages in this cycle, including processing data to create information, and even gaining insight from tacit knowledge.

Improving the implementability of each of these steps involves focussing on improving a specific area within the organization. An organization could improve its learning processes, or increase the availability of facilitating factors to learning.

As stated by Appelbaum & Reichart (1998), maximizing the facilitating factors would add more to the organization's learning capabilities than enhancing or changing the learning orientation.

In aiming to gain as much as possible from knowledge and the mining process, one author has suggested that knowledge maps are a crucial prerequisite. Knowledge mapping is a consciously designed communication medium using graphical presentation of text, stories, models, numbers or abstract symbols between map makers and map users. Knowledge maps are excellent ways to capture and share explicit knowledge in organizational contexts. (Wexler, 2001)

It is important to recognize in all knowledge maps, even ones that are in common use (like the organization chart), that the relationship between map makers and map users is fluid and dynamic (Wexler, 2001). As has been stated in previous sections, a dynamic nature is somewhat ubiquitous in the technological world of data mining.

Wexler suggests that knowledge maps should be used by all involved in the knowledge management process, and not just by the IT function.

The knowledge map should yield economic, structural, cultural and/or knowledge management returns. The who, what and why of knowledge mapping is not a call for simplification but rather one for increasing the confidence of general managers to get involved in the process (Wexler, 2001).

4.4.1 The Information System infrastructure

Within the strategic direction explained, a narrower definition of strategies would be more appropriate in designing concrete ways through which to move towards improving organizational learning.

King (2001) has proposed six distinct strategies for creating a *learning organization*, which is often defined as being a completely different concept to *organizational learning* (Örtenblad, 2001). One of King's strategies is that of utilizing an information systems infrastructure.

An information systems infrastructure has become somewhat ubiquitous throughout most industries today. Where a cutting-edge IS was unique in the 1980's, and gave an organization a unique competitive advantage, the development of technology has made advanced systems much more accessible to a wider range of organizations.

Information systems refers to the paradigm that involves the collection of data and its transformation into successively more useful and more valuable explicit information. The data are primarily numerical, although textual data (e.g., reports of sales calls and commentaries concerning the economy as well as images and sound "clips") are also dealt with by modern information systems (Thuraisingham and Bhavani, 1992; Ryan, 1995).

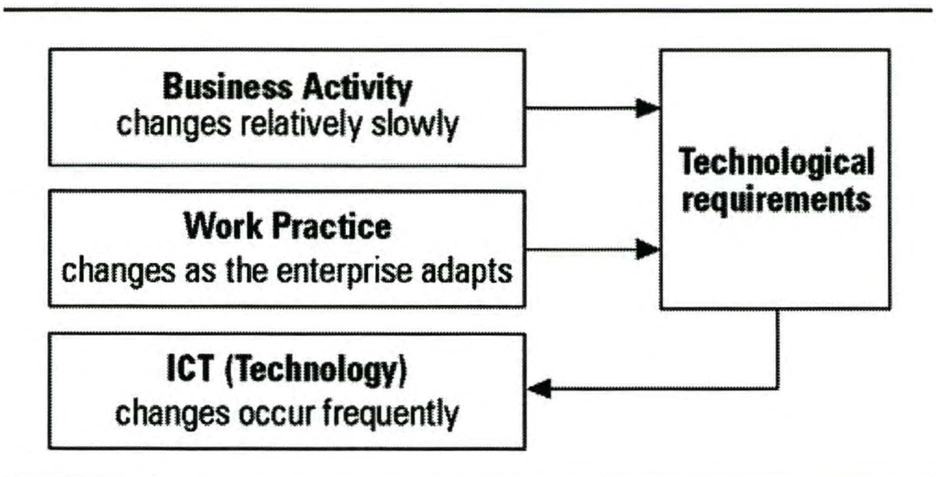
An organization that chooses to employ an information systems infrastructure strategy in pursuit of a learning organization does so by creating databases, inquiry capabilities, communications capacities, and other elements of an IS infrastructure. It typically attempts to have a "leading-edge" infrastructure that will enable and facilitate collective learning, information sharing, collaborative problem solving, and innovation. (King, 2001)

This information system infrastructure is more often than not simply put into place and maintained, without active training and utilisation monitoring. Many organizations do however actively endorse the use of the infrastructure, but do not “force” it, allowing individuals within the organization to make use of it within their own discretion. The hope with this kind of strategy is that, by simply demonstrating the system and communicating the benefits of it, organizational members will take the initiative and use it themselves. This ties in with a “freedom of choice” culture, and is often found in more liberal organizations.

A problem does however arise with a strategy of simply providing tools – If there is no proactive effort to guide what is fed into the system, the organization might find that its data (and possibly knowledge) resources are diluted to an extent.

It would therefore make sense to separately and proactively (in effect interdependently) manage both the organizational processes and the information systems infrastructure. By separating the two, the organization enables itself to guide each in the needed direction, making synchronisation easier. The separation is illustrated in figure 4.7.

Figure 4.7 : Separation of organisational and technological issues



Source : Franckson et al (1998)

Taking into account the focus on Data Mining as a tool, it seems that Data Mining is mostly applied in a technological form. Thus *the focus on an information systems infrastructure as a starting point and facilitator for an Organizational Learning strategy seems logical.*

4.5 Analysis and Synthesis

The study, through this chapter, turned to the construct of organizational learning. Before an analysis of the enabling effect of data mining on organizational learning is possible, it was necessary to determine the definition of organizational learning, and the extent to which it is influenced by various factors.

Specifically, the factors needed to facilitate organizational learning were discussed. Firstly, when looking at these factors, it is important to realise that they are simply *facilitators,* and not a guarantee that organizational learning will be successful. The factors do however have enough depth to suggest that implementation of an organizational learning strategy can benefit greatly from their presence.

When taking the objective of the study into consideration, the concept of organizational learning is very specific. The enablement of this learning process, more specifically the factors needed, have been discussed. The next step in the process is determining :

- If data mining tools can be used in the enablement of organizational learning.
- How this enablement takes place.

4.6 Summary :

This chapter provided an overview and definition of organizational learning. Factors needed to facilitate the organizational learning process were reviewed, and the infrastructure needed for the process was discussed.

The next chapter will bring to the fore the specific relevance of applying data mining tools in the information systems infrastructure in a narrow sense, and in Organizational Learning in a broader sense.

Chapter 5 : Enabling Organizational Learning through Data Mining

5.1 Introduction

With the advent of knowledge management, intellectual capital is gaining increasing recognition as the only true strategic asset (Hamel, 1998). The question is : How does an organization attain and leverage intellectual capital, and where does organizational learning come into the picture ?

Data mining as a concept and area of study has been probed in the previous chapters, with focus on the strategies and specific data mining tools available to the knowledge professional. The field of organizational learning has also been described, with specific emphasis on the factors which facilitate it, and within a narrower context the infrastructure needed to enable implementation. Organizational learning is a process aimed at leveraging knowledge, ultimately increasing the value of the organization's intellectual capital.

The study now turns towards bringing together the two areas, and examining the enabling effect data mining has on organizational learning. This is structured as follows :

Firstly, an overview is given of the relevance of applying data mining in organizational learning. This is needed to bring into context the structure of the study.

The next section looks at the application of specific data mining tools in improving the availability of organizational learning's facilitating factors.

Thirdly, an analysis is done of improving the implementability of organizational learning through improved availability of the facilitating factors.

Fourthly, an overview is given of the direct effects data mining tools have on organizational learning, and the last section deals with the enablement of an environment conducive to organizational learning

The question in this chapter is whether the use of data mining tools has a positive effect on the organizational learning process, and how this effect is manifested.

5.2 The Relevance of applying Data Mining in Organizational Learning

Data mining as a field, although relatively thoroughly documented, remains largely unexplored. Because data mining relies so heavily on technological capabilities, development occurs at a high speed. According to some, the development of technology in the larger picture is still in its infancy, thus one could expect the use of advanced data mining applications to increase in future.

Data mining applications form part of the Organizational Knowledge Management System. Taking Quinn *et al.*'s (1996) definition of intellectual capital, an Organizational Knowledge Management System can be seen as that which organizes a firm's know-what, know-how, and know-why into explicit knowledge resident in the firm's databases and operating technologies (Nonaka, 1991; Quinn *et al.*, 1996; Sviokla, 1996).

Knowledge in the correct form is priceless to the organization, and attaining it is often not a simple process. Bierly, Kessler & Christensen (2000) put forward a framework with four levels of learning, one for each in the data / information / knowledge triangle (described in Chapter 4). These factors are described according to their definition, process and outcome. In this case the ultimate goal is success through wisdom. This framework is illustrated in table 5.1.

Table 5.1 : Distinctions between data, information, knowledge and wisdom

Level	Definition	Learning process	Outcome
Data	Raw facts	Accumulating truths	Memorization (data bank)
Information	Meaningful, useful data	Giving form and functionality	Comprehension (information bank)
Knowledge	Clear understanding of information	Analysis and synthesis	Understanding (knowledge bank)
Wisdom	Using knowledge to establish and achieve goals	Discerning judgments and taking appropriate action	Better living/success (wisdom bank)

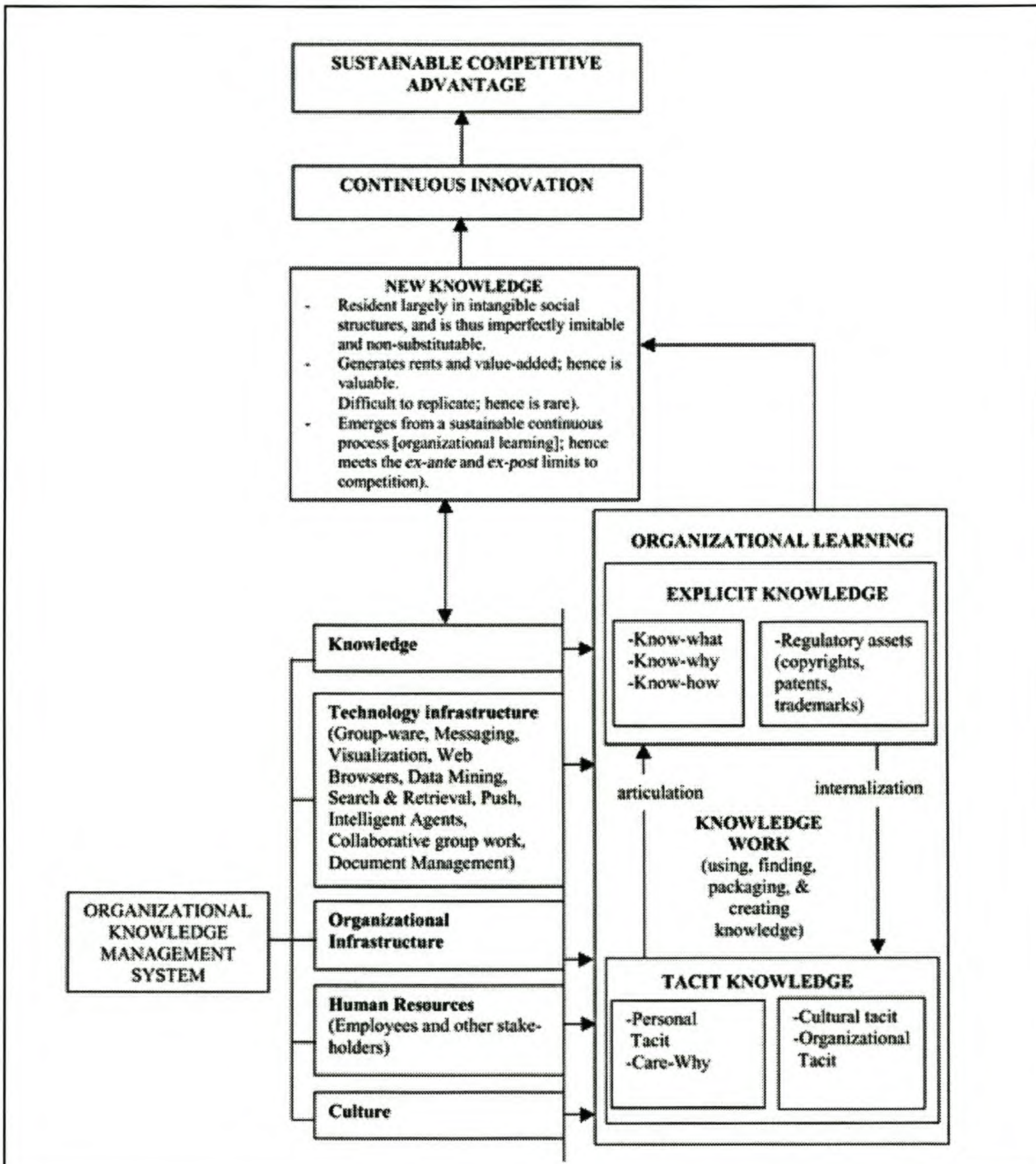
Source: Bierly, Kessler & Christensen,(2000)

Nonaka's spiral of knowledge creation leads to the definition of an organizational knowledge management system as that which supports organizational learning by enhancing the exchange and sharing of tacit and explicit knowledge (Meso & Smith, 2000)

When taking a systems perspective, one should realise that the organizational knowledge management system is more than simply the technology infrastructure. The system is made up of a combination of technology infrastructure, organizational infrastructure, corporate culture, knowledge, and people.

Throughout the knowledge management process, systems in their different forms have to be applied in order to gain as much from the knowledge as possible. Development of an organizational knowledge management system is only the first step in the knowledge management process.

Figure 5.1 : Organizational knowledge infrastructure and its relation to sustainable competitive advantage



Source : Meso & Smith (2000)

Figure 5.1 illustrates the continuous use of an organizational knowledge management system which needs to take place in order for the organization to gain a competitive advantage. The knowledge management system is only effective if it is utilized to its full potential through regular use. This of course is dependent on leadership (in whichever form) within the organization, and its commitment to making the system work.

In bringing the study back to organizational learning, a knowledge management system is a very important facilitator. In turn, data mining plays a crucially important part in any OKMS.

The direct effect is thus that Data Mining has a very high degree of relevance in the construct of organizational learning. It is this author's opinion that the relevance will increase even more as :

- Business processes move even more from a paper to digital base.
- Database ownership and availability increases because of this digitization.
- The amount of digital data which needs to be processed exceeds conventional methods of dissemination.

Data mining will thus increase in importance, with specific relevance to related organizational learning capabilities.

5.3 Data Mining tools – Improving the availability of Organizational Learning's facilitating factors

The use of Data Mining tools can be applied to several different situations. In this section we will look at how the use of the specific data mining tools analysed can improve or aid the availability of certain facilitating factors. These factors, listed as facilitators of organizational learning in chapter 4, are important in setting a positive stage, and in acting as catalysts to the organizational learning process. We will now bring the data mining tools into the picture to determine if they aid these factors. Some factors are specifically influenced by specific tools, while others are influenced generally by the range of tools.

5.3.1 Facilitating factor 1 - Scanning imperative and performance gap

The difference between what the organization has targeted performance wise, and what it has achieved, is measured in this factor. This also relates to the organization's understanding of the environment in which it functions. An understanding of the environment implies an awareness of :

- Environmental factors
- How these factors impact the organization
- How to influence the positive reaction to these factors

The organization thus needs an understanding of how it performs within the environment, and how it measures the gap between how it performs and how it has planned to perform. This in effect means that the organization has to be aware of how to influence those factors that cause the gap in performance.

In order to be aware of / measure the abovementioned "underperformance" factors, the organization needs intelligence from a wide variety of sources on a wide range of performance factors.

To compile an intelligence (or "status") report on organization-wide performance has been a daunting task in the past, handled by specific reporting teams, and done over long period (long in technology terms). All this is being changed by Data Mining. The specific data mining tools can be applied in the following ways :

5.3.1 a) Expert Systems :

As expert systems have been found to generally run as applications (refer to chapter 3), they can be applied as knowledge bases to be probed as needed. An expert system could for example be applied in keeping record of customer interaction types and effects thereof. The organization would be able to query this database on demand to determine which types of interactions had the highest positive effect on customer retention and repeat sales.

The organization can thus, on demand, obtain a generalised (or even specific) picture in a specific business area. Expert systems are not suited to constant automatic queries, as they have to be given specific "questions" to be able to provide feedback.

5.3.1 b) Genetic Algorithms :

Genetic algorithms are potentially very well suited to measuring performance gaps. As these algorithms function at agent level, they can be set up to constantly monitor specified data in the database, and look for unusual patterns or exceptions. Exception reporting has become a topic of importance, as it provides management with information which is relevant to foreseeing and *preventing* problems, instead of simply reacting to them.

A genetic algorithm agent could be set up to provide e-mail or mobile device alerts to the appropriate management level in case of an increase in a specified problem in customer interaction. The algorithm would function by detecting patterns which arise in, for example, the amount of customers buying from a different department on the introduction of a new warranty program by the vendor. Genetic Algorithms are best suited to provide alerts in cases where management would not necessarily have been aware of the possibility of any problems.

5.3.1 c) Clustering :

Clustering systems provide the organization with a way to match up specific data types and draw conclusions from these matches. These programs, like genetic algorithms, also run in the background as agents, and can conduct searches according to specific match criteria. If used in conjunction with genetic algorithms, a clustering system can provide the organization with reporting on areas where an awareness of data matches (or even similarities) would be beneficial to awareness of performance gaps.

An example of the above would be an automatic report on instances of customers making increased use of one service, while at the same time reducing their usage of another high-value service which would usually be complementary. This enables the organization to become aware of the fact that delivery of the second high-value service might not be up to the optimal standards and targets set for it.

5.3.1 d) Neural Networks :

As "learning systems", neural networks provide management with an unsupervised way to track multidimensional trends in different areas. When measuring gaps between performance and targets, awareness of factors impacting performance negatively might not be as simple as spotting matches between two previously identified compounding factors.

In many cases, the organization needs to keep track of several trends simultaneously, without necessarily always knowing the impact that these trends have on each other.

A neural network enables management to provide the system with some base rules and training, and from there the system "learns" by itself. The network could initially spot two trends emerging, and from its rules, know that a simultaneous rise in these trends impacts organizational performance negatively. Later on, it could spot a completely different set of trends emerging, and by matching up the outcomes (either negative or positive) of these trends, it could deduce a hidden relationship between these trends, and advise management on the appropriate action that needs to be taken – This advice being in the form of recommendations for increasing the delivery of one service element while for example decreasing another (where in some cases the two factors might have seemed completely unrelated without the system's recommendations).

5.3.2 Facilitating factor 2 – Concern for measurement

Without an awareness of how functions within the organization are performing, it is difficult for management to adjust behaviour (and thus performance) at any level. The measurement metrics used are important here. In this case, the balanced scorecard (Kaplan & Norton, 1993) has (as was discussed in chapter 4) made inroads into management's perceptions of how focused the organization is on measurement.

Employees adjust their behaviour within the organization if they are aware that measurement is done in a balanced way, and that positive contribution is visible.

In this factor, Data Mining tools can provide a visible "window" to the measurement system, and allow employees to have access to measured results and make their own conclusions. According to Appelbaum & Reichart (1998), a critical test of a scorecard's success is its transparency. The main contribution of data mining in this case is thus in providing transparency to the measurement process.

In dealing with this factor, it is not necessary to go into the specific transparency provided by each factor. It is however important to realise that the detail of measurement provided by Expert Systems, Genetic Algorithms, Clustering and Neural Networks gives the organization a way to provide a concrete measurement to its employees through mediums like :

- A corporate intranet
- Newsletters
- Measurement reports

It seems logical that the more employees know about how the organizations (and its different functions) are performing, the more motivated they will be to improve their own performance.

5.3.3 Facilitating Factor 3 – Experimental Mindset

The experimental mindset an organization has influences the way it manages databases and the knowledge inherent in them. Experimentation involves the systematic searching for and testing of new knowledge (Appelbaum & Reichart, 1998).

To make it possible for individuals within the organization to experiment, and develop an experimental mindset, the organization has to provide them with the tools to analyse data and draw conclusions from it. The best way to develop an experimental mindset is to empower employees to act on insights into knowledge. In turn, the best way to empower employees is to provide them with the tools needed to analyse data and draw conclusions from it. Data Mining tools are a vehicle for this application, and the interfaces can be adapted to meet the employees' needs.

The specific tools can be applied as follows :

5.3.3 a) Expert Systems :

The nature of expert systems enables the configuration of these systems to an on-demand format available as a utility to employees. The most effective way to implement this would be in intranet webpage form, with query fields and criteria.

Knowledge workers can use the expert system as a reference in areas outside of their usual knowledge domain, and through it gain insights into the activities of their fellow workers.

The intranet interface allows easy deployment of the system, and encourages experimentation because it is easy to use, and available to everyone within the organization.

5.3.3 b) Genetic Algorithms :

Because Genetic Algorithms provide for a degree of intelligent searching, it allows organizational members to find results *relevant* to their specific queries, and enables them to search outside of their usual domain without fear of being overwhelmed by unfamiliar knowledge.

5.3.3 c) Clustering :

Clustering allows employees to experiment through finding unexpected relationships between data sets. This "discovery" of new knowledge serves as a motivator to experimentation, and provides a way for functional teams to cross-collaborate according to new data relationships.

As long as clustering programs are delivered in a user-friendly format, the adoption rate should be high enough to warrant implementation.

5.3.3 d) Neural Networks :

Because of the nature of neural networks, one would expect them to only be utilised by top management. This is not the case, as these systems can function at all levels of the information systems infrastructure.

Neural networks improve experimentation because of their intelligence. A knowledge worker might input a problem into a neural system, and retrieve a completely unexpected (but insightful) result, leading to further exploration. As long as this "curiosity" is stimulated, experimentation will flourish.

5.3.4 Facilitating Factor 4 – Climate of openness

The way in which an organization deals with failure reflects its climate of openness. The goal in this regard is a mindset that enables companies to recognize the value of productive failure as contrasted with unproductive success. A productive failure is one that leads to insight, understanding, and thus an addition to the commonly held wisdom of the organization. (Appelbaum & Reichart, 1998)

In providing all members of the organization with complete (or near-complete) access to its measurement and knowledge resources, management is providing part of what is needed to facilitate a climate of openness. Information on organizational activities goes a long way towards encouraging open discussion – If employees are aware of a larger range of activities, they ask more questions, and thus openness is facilitated.

Data Mining tools in general thus assist a climate of openness, simply because they provide a channel through which awareness and knowledge can be shared, and through which discussion is encouraged.

5.3.5 Facilitating Factor 5 – Continuous Education

As far as Organizational Learning as a broader construct is concerned, continuous education is of vital importance. As was shown in figure 5.1 at the beginning of this chapter, new knowledge (through continuous education) leads to continuous innovation, which in turn leads to a sustainable competitive advantage.

Data Mining tools assist continuous education through providing an information retrieval tool which is available at all times. Through interfaces like corporate intranets, tools are available on-demand, with scalability to accommodate multiple users usually built into the system.

If data mining tools are available fulltime, the organization is able to constantly drive education and knowledge discovery processes. This constant drive as a whole benefits organizational learning positively, and in turn has an effect on continuous innovation.

5.3.6 Facilitating Factor 6 – Operational Variety

Operational variety implies that the organization does not have overly inflexible or fixed procedures for accomplishing operational tasks. Variations in procedures provide members with constant new learning experiences, which in turn serve to drive the organizational learning process as a whole.

By finding new ways to do things, knowledge and other workers are improving their existing skill sets, as well as developing new ones. The data mining tools can assist in the following ways :

5.3.6 a) Expert Systems :

Expert systems allow people involved with projects in operational areas other than their own to find answers to highly technical questions. Because these systems are available to "answer" queries, knowledge workers need only ask questions about the relevant topic on the specific system (which handles only the knowledge domain allocated to it) to get highly accurate technical answers.

Because of the easy availability of answers, cross-functional workers are encouraged to explore areas of knowledge outside their own, and develop a variety of skills.

5.3.6 b) Neural Networks :

Because of the neural network's ability to find unexpected relationships between unrelated data subsets, it allows the organization to create knowledge-sharing processes between internal divisions. This brings together people from different areas and in itself creates operational variety for individuals involved in the process.

5.3.7 Facilitating Factor 7 – Involved leadership

Leadership in the form of visionary managers are responsible for guiding the organization towards a strategic goal, and they can only do so if aided by the ability of the organizational knowledge management systems to constantly provide up-to-date information about all facets of operational and management activities.

The main data mining tool which is able to improve the availability of this factor is the Neural Network. This tool allows the leadership of the organization to constantly and consistently stay in touch with regards to :

- Performance measures
- Possible problems
- General communication & motivation

Neural networks allow management to become leaders through being informed in a timely fashion of any possible problems in operational areas. Being able to pre-empt problems frees more time to be involved with actual leadership activities such as mentoring and developing. Thus, the essence of a leader's role is to develop a shared vision, provide the resources needed to achieve the vision, delegate authority, celebrate successes, and be a learning architect. In essence, to design successful learning experiences. (Appelbaum & Reichart, 1998)

5.3.8 Facilitating Factor 8 – Systems Perspective

The systems perspective is arguably the most important of the eight facilitating factors discussed. It relates to all individuals within the organization seeing the organization as a complete system, which they are part of. This view has the effect of individuals realising that their actions have consequences beyond their operational area, and that they are interdependent on other individuals and operational areas within the organization structure.

This systems perspective allows for renewed efforts towards managing all activities with the goal of improving the system as a whole. The different Data Mining tools can contribute as follows :

5.3.8 a) Expert Systems :

Upon having access to a variety of expert systems, knowledge workers come into contact with information and knowledge from other operational areas, enhancing the learning experience (as has been discussed). These activities of probing expert systems gives the individuals involved a view from a different perspective.

Once members of the organization realise that not only *their* efforts are essential to organizational success (but also the efforts of other organizational members), they adapt and modify their behaviour to correspond with the system as a whole.

5.3.8 b) Genetic Algorithms :

Genetic Algorithms provide search results in context. It is this context which is important for the systems perspective. Once individuals start probing a variety of information sources through the algorithm method, they might realise that the activities which they perform are also relevant *in context*. This realisation aids a move towards a more systems-oriented perspective.

5.3.8 c) Neural Networks :

The Neural Network provides the state-of-the-art in systems thinking. It enables the ability to "learn" by itself across operational boundaries and bring information which is *important,* regardless of where it comes from, to the attention of those to whom it is relevant.

The Neural Network in this way provides all involved organizational members with knowledge which they might not previously even have been aware of, because it was contained in the knowledge base of another organizational area, but which is highly *relevant* to them.

5.4 The direct effects of Data Mining tools on Organizational Learning

Data Mining allows the organization to monitor predefined sets of

- Performance measures
- Movement patterns
- Research
- Transactions
- Customer interactions

at predefined intervals, or if needed, in real-time. This implies that there is constant availability of information regarding any operational activities included in the monitoring process.

Data Mining also allows the organization to extract new knowledge from existing static databases, which would otherwise be "gathering dust". The best applications for Data Mining are however dynamic databases, as it is current information and future trends which are most powerful and useful in the business environment.

If organizations are to successfully adopt the learning approach to competitiveness, they need to understand both the theory and practice of organizational learning. (West & Burnes, 2000) This learning process can only be accomplished if timely access to information and knowledge is available – Which data mining provides.

Many writers (Coopey, 1996; Garvin, 1993; Hirschorn and Gilmore, 1992; Nonaka, 1991; Senge, 1991) argue that learning organizations are skilled at a range of activities that enable them to develop and integrate their learning. These include their capacity for:

1. Systematic problem solving, which underlies much of the quality movement and is essentially focused on transformation in management and organizational activity;
2. Experimentation: actively seeking and testing new knowledge and learning from mistakes;
3. Drawing upon memory and past experience;
4. Learning from and with others;
5. Communicating effectively within and beyond the organization;
6. Systems thinking and developing shared ideas/models of the current organizational position.

These 6 "skills" reflect to a large extent the 8 factors discussed as facilitators of organizational learning.

In connection with this final point, as Hendry *et al*. (1995) note, the importance of groups and teams in the organizational learning process seems, surprisingly, to have been neglected. This is possibly one of the most important contributions Data Mining tools can (if properly implemented) have on organizational learning – Enabling effective collaboration because of universal availability of information in the relevant form.

One could thus say that the direct effect of Data Mining tools on Organizational Learning is that of providing the constant information presentation and dissemination ability which is crucial to the learning process.

5.5 Effects of a Data Mining orientation on Organizational Learning :

Although the direct effects of data mining tools on organizational learning have been described, it is prudent to consider the effect that a general orientation towards data mining can have on organizational learning.

Because implementing a data mining approach and strategy (as was described in chapter 2) usually necessitates a move towards processes and procedures that facilitate and enable the mining process, it creates a general orientation towards making data mining possible.

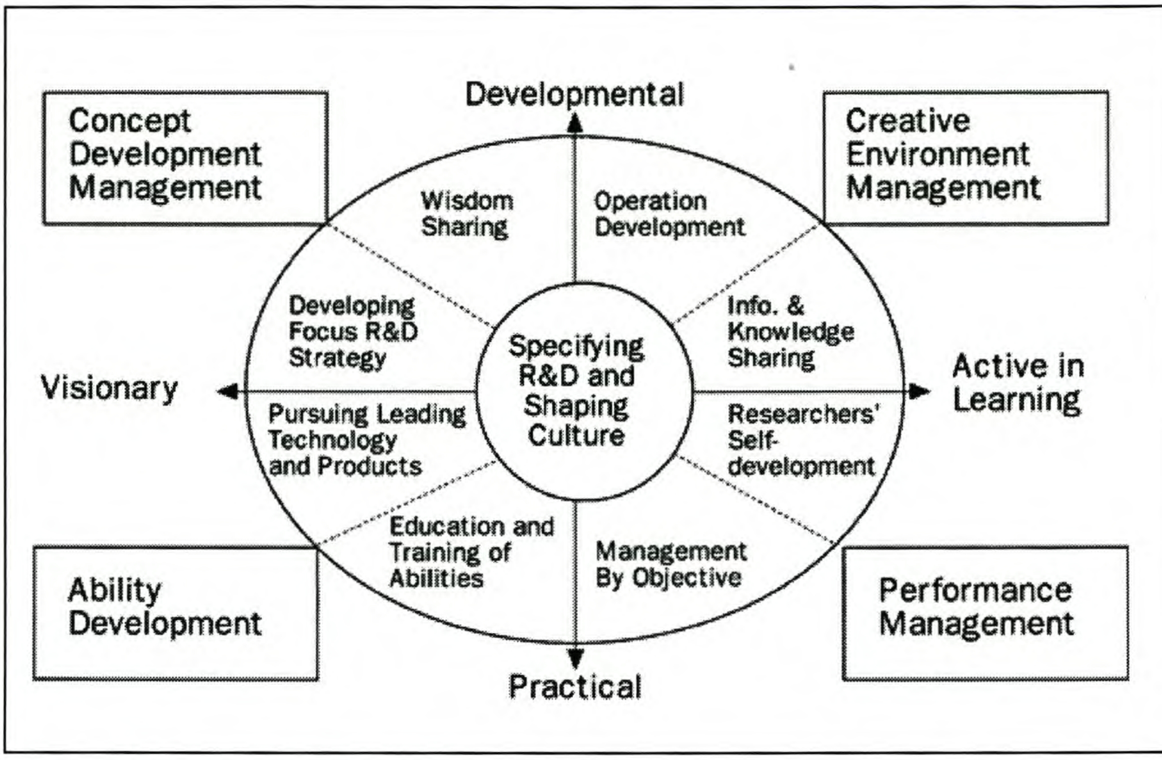
This orientation usually takes the form of an environment in which :

- Employees are geared towards extracting as much data as possible from sources (such as leads and prospects).
- Facilities to input and process data are readily available.
- Management uses mined information to aid decision-making.

This environment in turn encourages innovative management, which impacts on several other organizational functions.

Figure 5.5 shows the model for innovative management, centering on specifying Research and Development and shaping corporate culture. What is usually termed as the vision of a business is the consensus reached within the organization on the focus R&D strategy with the aim for leading technology and products. To be developmental is to share wisdom learned through developing operation processes. (Hong & Kuo, 1999)

Figure 5.5 : The model of innovative management



Source : Hong & Kuo (1999)

An environment where innovative management flourishes is ideal for progression of several enablers of organizational learning (as discussed), and through progressing these enablers, a general attitude of learning is fostered. Through this fostering, learning is constantly encouraged, and the process of organizational learning is accelerated.

5.5.1 Internal Benchmarking :

Another factor through which a data mining orientation can possible influence organizational learning is internal benchmarking. To remain internationally competitive firms must sustain a high rate of internal learning that both refines current practices and adopts new ones (Hyland & Beckett, 2002). Benchmarking is a way for an organization to identify the best practices in a particular area, and adapt to these practices to improve its own operations.

Benchmarking falls into two classifications :

- External Benchmarking
- Internal Benchmarking

Both these forms have in common the measurement against a predefined standard. Internal benchmarking however, measures the organization's performance against itself.

External benchmarking measures the best practices of other organizations (in effect the "world best" practices). The problem however is that best practices identified in one organisation are not necessarily easily transferred to another (Hyland & Beckett, 2002). Some of the reasons for this were investigated by Szulanski (1996), who suggested that some aspects of organisational culture and communications were the main inhibitors.

Internal benchmarking is thus an important factor to consider when looking at the effects of data mining on organizational learning. Internal benchmarking is enabled by data mining through increased data for benchmarking being available. This in turn makes it possible for increased organizational learning to take place through the information (and ultimately knowledge) gained from benchmarking.

5.6 Analysis and Synthesis

The study up to this chapter has provided definitions and explanations of the current nature and status of data mining tools. It has also focused on the construct of organizational learning, with a listing of some of the factors necessary to facilitate it.

This chapter provides insight into the actual enablement of organizational learning, more specifically through application of the various tools reviewed. However, before these applications could be considered, it was necessary to determine if data mining is in fact relevant to the organizational learning process. It was found that, because data mining forms a crucial part of the organizational knowledge management system (which is cited by many authors to be one of the prerequisites of organizational learning), it is in fact completely relevant to the learning process.

Once relevancy had been determined, enablement was discussed in three dimensions :

1. The use of data mining tools to improve the availability of organizational learning's facilitating factors.
2. Direct effects of data mining tools on the organizational learning process.
3. Effects of a general data mining orientation on organizational learning.

The learning process is influenced extensively throughout the three dimensions, both in direct and indirect ways. The most surprising influence is that which the availability of mining tools has on the general orientation of employees (and the organization as a whole). The changes in behaviour brought about by being "encouraged" to gather as much data as possible (through the tools being easily available), is not to be discounted.

From the analysis, it seems as if Data Mining (both as an orientation and specific application) is beneficial to the general enablement of an organizational learning strategy. The environment created by being data mining-oriented can only have positive effects on organizational learning as a whole.

5.7 Summary :

In this chapter, the focus was placed on the enablement of organizational learning through data mining. Firstly, an analysis was done of the effects of the specific data mining tools on improved availability of organizational learning's facilitating factors.

Secondly, the direct effects of data mining tools on organizational learning were discussed. Lastly, the focus was placed on the effect that a data mining orientation has on organizational learning.

Chapter 6 : Summary, Conclusions and Recommendations

Throughout the study, Data Mining was analysed in terms of both its generic and conceptual characteristics, and specific applications through the tools discussed in chapter 3.

Data Mining as a field is very dependent on technology, and the technological side of the equation was considered as an important infrastructure requirement.

The importance and effects of having a data mining orientated environment within the organization were also discussed.

6.1 Summary :

6.1.1 Defining Data Mining :

Data Mining was defined as a group of activities performed within an organization with the goals of extracting specific data (and more importantly, information) from a database, or multiple databases.

Linking with the concept of data mining is the data warehouse, a repository of information in various possible forms. Data Warehouses consist of relational databases and information management systems which are the focal point of all end user queries. With the development of technology, it has become possible to store data in Data Warehouses in several different forms, including :

- Figures (numbers)
- Text
- Images
- Audio
- Video

Because mass data storage has become accessible to larger markets (as discussed in chapter 2), data mining has become an increasingly important tool in the organizational move towards dissemination of the increasing masses of data.

In the discussion on the concept of data mining, it was also determined that the application of data mining is not always found as a specific purchasable tool. Instead, data mining is a process which involves several functions within the organization, as well as an orientation towards learning.

This "learning orientation" - one which was found in chapter 5 to often manifest in an innovation oriented corporate culture – and a data mining strategy were found to be mutually beneficial to each other.

6.1.2 The need for Data Mining :

The environmental factors leading to increased interest in the field of data mining were discussed in chapter 2, with the conclusion being that the increased pace of technology development serves as a prime motivator. Because of this development pace, the informational needs of individuals and organizations are changing dramatically, and the nature of the organization is changing as a result of this.

6.1.3 Approaches to Data Mining :

In developing a data mining strategy, it was found that the approach used can be critical in determining the success rate. Three categories of approaches were found as options :

- Based on the database
- Based on the knowledge

- Based on the techniques

These approaches become important in the sense that they determine to a large extent which area of the business the strategy is mainly implemented in.

Within these categories of approaches, it was found that there are several specific approaches, including statistics, artificial intelligence, decision trees and visualization. Within these approaches, even more specific tools are found.

6.1.4 Comparative analysis of Data Mining tools :

As generalised Data Mining approaches had already been defined in chapter 2, the discussion moved towards specific tools in chapter 3. These tools are the practical manifestations of a data mining strategy, and differ in their applicability to specific mining needs and situations.

Several types of Data mining tools were listed and described, among them expert systems, genetic algorithms and neural networks. The tools were then defined and compared according to their application areas, implementability, and interoperability.

Application areas within the organizational knowledge management framework were found to be widespread, including:

- Providing high-speed access to large amounts of unclassified data.
- In-depth probing of specialist knowledge areas.
- Constant monitoring of information to extract trends.

Concerning implementation, it was found that most of the tools require a relatively powerful and comprehensive information technology infrastructure for them to benefit the organization. As most of the tools are technological (more specifically software) in nature, their power and functionality increases as technology develops.

The ability of Neural Networks to adapt to problems and "learn" are at this stage limited only by micro processing speed, and future prospects for advancement seem positive.

In terms of interoperability, the use of combinations of tools was examined. The conclusions were that data mining tools are seldom implemented in isolation, and that they are often dependent on each other. Most mining strategies would involve the use of a combination of tools integrated as a system. Integration is the key in this case, as was illustrated in the K-12 schools case study at the end of chapter 3.

6.1.5 Enablement of Organizational Learning :

Because the objective of this study concerns the enablement of organizational learning through data mining, focus was placed on defining the concept of organizational learning in Chapter 4. Organizational Learning as such is a subset of the knowledge management field.

Organizational Learning was defined as activities which are performed by organizational members (or by the organization) in reacting to environmental changes / challenges. These activities usually have the effect of changing the organization from within.

The ultimate goal of organizational learning (as was stated by various authors) is a learning organization, one which constantly adapts to changes and becomes more effective through this adaptation.

To enable organizational learning, it was stated by one author that several facilitating factors need to be present. These factors were also reviewed in chapter 4, and range from a concern for measurement, experimental mindset and continuous education to involved leadership and a systems perspective.

6.1.6 Enablement infrastructure :

On consideration of the factors needed to facilitate organizational learning, it became clear that an infrastructure to enable learning has to be created within the organization. This infrastructure was reviewed, with focus on the need for:

- A general environment and organizational orientation
- Information technology systems infrastructure

Both these infrastructure requirements were discussed, taking into consideration the goals of knowledge acquisition, sharing and utilization.

When analysing the infrastructure, it becomes clear that cooperation between, and integration of organizational functions is crucial to the success of the data mining (and ultimately organizational learning) process.

6.1.7 Using Data Mining Tools to enable Organizational Learning :

An analysis of the application of data mining tools to enable organizational learning is stated as the main objective, and the study turned in this direction in chapter 5. As data mining, its specific tools, and the concept of organizational learning had already been explored in previous chapters, the focus of this chapter was to determine whether data mining does in fact impact positively on organizational learning.

The relevance of applying data mining techniques in an organizational learning context was examined first, with focus placed on the organizational knowledge management system. It was stated that the organizational knowledge management system is an important facilitator of organizational learning.

The systems perspective (regarded as important throughout the study, and cited specifically as one of organizational learning's facilitating factors in chapter 4) was shown as being crucial to designing a data mining strategy.

The conclusion was that data mining has very high relevance in the field of organizational learning.

After establishing the relevance of data mining, the study turned towards an analysis of the effects of specific data mining tools on the availability of organizational learning's facilitating factors.

The influence of tools on each facilitating factor was reviewed, with emphasis on practical application areas. The tools were proven to impact positively (with varying degrees) on the facilitating factors, with a large range of possible applications mentioned.

Possible direct effects that tools have on organizational learning were also listed. The primary direct effect was concluded to be that constant information presentation and dissemination abilities are available to organizational members.

The effects that an orientation towards data mining as such can have on enabling organizational learning were also taken into consideration. It was found that, because moving towards a data mining strategy requires the use of certain processes and procedures, an environment is created within which :

- Employees are geared towards extracting as much data as possible from sources
- Facilities to input and process data are readily available.
- Management uses mined information to aid decision-making.

This environment in turn encourages innovative management, which is a crucial part of organizational learning.

6.2 Synthesis and Conclusions

This study has had as its purpose from inception the enablement of organizational learning, with an investigation into the steps necessary to make this enablement possible as the logical analysis process.

In chapters 2 and 3, the current nature and status of data mining (and its specific tools) was reviewed and analysed. Thereafter, in chapter 4, the nature of organizational learning was studied, with specific reference to the factors needed to facilitate it.

In chapter 5, the relevance of applying data mining to the organizational learning process was determined. This relevance in turn led to an analysis of exactly *how* data mining's specific tools influence (and ultimately enable) the organizational learning process.

The objective of "**an analysis of the current nature, status and relevance of data mining tools to enable organizational learning**" was thus accomplished, with the conclusions drawn from the logical process as follows :

In the introduction of chapter 1 a point is made about efficiency (doing things efficiently) vs. effectivity (doing *the right things* efficiently). It is effectivity which forms the cornerstone of thinking behind this study.

After considering effectivity and realising that, to make it a reality, one needs to know *what to do*, the following questions arose :

- How does one know what to do ?
- What are the right things to do ?

This author believes that data mining, applied correctly, can provide answers to these two questions. Data mining, in its simplest form, provides whoever uses it with a way to analyse and extract information from a database according to specified criteria. Although defining this criteria is often overlooked as a means towards an end, the criteria is defining of the end. Criteria are not just simple parameters, but in effect the essence of what the user wants from the information. If the criteria did not exist, there would be no system. More importantly, if the *right* criteria do not exist, there also might as well be no system.

The question could thus be asked : What is data mining about ? The answer lies in providing the correct people with information relevant to them. Not just information in the form of data, but information in its defining form, because it is available in context. Referring to the knowledge creation spiral, recall that data in context is information, and information in even more context in the hands of the correct people is knowledge. This is what data mining is about – Creating knowledge by routing information to the correct people.

Once the correct people have the correct information in their hands, they can learn from it. Learning is defined as "gaining skill or knowledge by study". Study can occur in many forms, and most of these can be accelerated by Data Mining. Once studying becomes seamless and context-related, learning progression reaches a different level. Organizational learning is learning, but applied on different levels between numbers of individuals.

In this study, the factors needed to facilitate (and ultimately enable) organizational learning were reviewed. In turn, specific focus was placed on the infrastructure needed to enable these factors. It was found that, to enable the technological component of the infrastructure to be applied successfully, the organization needs to place emphasis on creating a culture of using this technology. Once this culture exists (and is nurtured), the benefits of using the infrastructure tools which are available begin to appear.

Furthermore, this study has proven that the specific technological tools with which to mine data do exist. They exist in different forms, applied through different approaches and techniques, by different people within an organization, but they do exist. Whether or not the organization can utilise data mining tools depends on the attitude towards it. Data Mining is only as effective as the way it is applied, which in the organization's case depends on the infrastructure, goals, data and criteria provided to the system.

The answer to the question "Does organizational learning benefit from data mining, and more specifically data mining tools – Do these tools enable organizational learning ?" is provided partly by this study :

1. Specific data mining tools can be applied to provide the organization with knowledge (assuming the correct criteria are specified).
2. Data mining as a whole can provide the organization with a learning attitude, and a learning attitude is beneficial to successful mining attempts.

The conclusion is thus that data mining tools can be used to enable organizational learning – as long as the mining process is implemented with care, a logical approach and techniques are applied, and the process is constantly monitored, providing the right people with the right information to enable learning to progress to a higher level.

6.3 Recommendations for Business Applications :

Recommendations for improved enablement of the organizational learning process through utilising the appropriate data mining tools (and other technologies) could include, but are not limited to :

- Establishing core technology plans or "roadmaps" as a crucial component of the strategic planning process. Without a defined and more importantly, integrated IT strategy, it is difficult to determine the informational and knowledge needs of the organization. As has been determined, the information systems infrastructure is a very important component of the data mining process, and should not be overlooked or classified as a "technical issue".

- Creating and nurturing learning environments within the organizational structure. By creating an atmosphere conducive to learning, the learning process is accelerated, and adoption of learning processes occurs much faster.

- Simplifying the data mining interface. Because the learning process is dependent on organization-wide adoption, simplifying the interface could facilitate increased usage among all staff levels. The redundancy created by knowledge distributed throughout different operational areas has been proven to be a powerful learning success factor.

- Providing users at all levels access to all information. By promoting an environment of transparency and accountability, the learning process could be accelerated. Data mining tools are an appropriate interface through which to provide access to the mass of information that might be contained in the organization's databases.

6.4 Recommendations for further research

This study investigates the different approaches to data mining and specific tools available to enable the organizational learning process. What it does not investigate is how successful organizations apply data mining to become more profitable. The question is thus : Does applying data mining (and therefore enabling organizational learning) provide an organization with a sustainable competitive advantage ?

More specifically, there is a need for further research in the areas of competitive behaviour in the data mining arena. It would be beneficial to the fields of data mining and organizational learning to determine exactly how organizations use data mining tools to gain measurable advantages over their competitors.

Recommendations are thus that further empirical research is conducted in the area of practical competitive and strategic application of data mining.

Finally, taking organizational learning into consideration, the question is whether organizations become learning organizations through their application of data mining (and subsequent success), or whether they become learning organizations in order to be successful in the first place.

Bibliography :

Ackoff, R.L, 1981, Creating the Corporate Future, John Wiley & Sons, Inc, New York, NY.

Anderson T, Fare R, Grosskopf S, Inman L, Song X, 2002, "Further examination of Moore's Law with data envelopment analysis", Technological Forecasting & Social Change, 69, 5, 465-478.

Appelbaum, S.H, Reichart, W, 1998, "How to measure an organization's learning ability: the facilitating factors - part II", Journal of Workplace Learning, 10, 1, 15-28.

Argyris, C, 1977, "Double loop learning in organizations", Harvard Business Review, Sep/Oct, 115-25.

Argyris, C, Schön, D.A, 1978, Organizational Learning: A Theory of Action Perspective, Addison-Wesley, London.

Argyris, C, Schön, D.A, 1996, Organizational Learning II; Theory, Method and Practice, Addison-Wesley, Reading, MA.

Ashford, J, 1997, "An Information-space model for the development and application of computer-based tools in information creation and dissemination", Journal of Documentation, 53, 4, 351-73.

Berson, A, Smith, S.J, 1997, Data Warehousing, Data Mining, & OLAP, McGraw-Hill, New York, NY.

Bierly, P.E, Kessler, E.H, Christensen, E.W, 2000, "Organizational learning, knowledge and wisdom", Journal of Organizational Change Management, 13, 6, 595-618.

Bollinger, A.S, Smith, R.D, 2001, “Managing organizational knowledge as a strategic asset”, Journal of Knowledge Management, 5, 1, 8-18.

Brown, J.S, Duguid, P, 1991, "Organizational learning and communities-of-practice: toward a unified view of working, learning, and innovation", Organization Science, 2, 1, 40-57.

Caudill, M,1991, "Neural network training tips and techniques", AI Expert, 6, 1, 56-61.

Chen, M.S, Han, J, Yu, P, 1996, "Data mining: an overview from a database perspective", IEEE Transactions on Knowledge and Data Engineering, 8, 6, 866-83.

Coles, S, Rowley, J, 1995, “Revisiting decision trees”, Management Decision, 33, 8, 46-50.

Cook, S.D.N, Yanow, D, 1993, "Culture and organizational learning", Journal of Management Inquiry, 2, 4, 373-90.

Coopey, J, 1996, "Crucial gaps in the learning organization: power, politics and ideology", Starkey, K., How Organizations Learn, Thomson Business Press, London.

Davenport, T. H, DeLong, D. W. and Beers, M. C, 1998, ‘Successful knowledge management projects’, Sloan Management Review, 39(2), 43–57.

Quinn, J.B, Anderson, P, Finkelstein, S, 1996, "Managing professional intellect: making the most of the best", Harvard Business Review, March/April, 71-80.

Cupello, J.M, Mishelvick, D.J, 1988, "Managing prototype knowledge/expert systems projects", Communications of the ACM, 31, 534-541.

Drucker, P.F, 1993, Post-Capitalist Society, Butterworth-Heinemann, Oxford.

Everitt, B., 1980, Cluster Analysis, 2nd ed., Heinemann Educational Books, London.

Fayyad, U, Piatetsky-Shapiro, G, Smyth, P, 1996, "From data mining to knowledge discovery: an overview", Fayyad, U, Piatestsky-Shapiro, G, Smyth, P, Uthurusamy, R, Advances in Knowledge Discovery and Data Mining, MIT Press, Cambridge, MA.

Firebaugh, M.W, 1998, Artificial Intelligence - A Knowledge Based Approach, Boyd and Fraser.

Franckson, M, Hall, J, Helmerich, A, Cañadas, R, Dehn, M, 1998, "ADDE: Application Development for the Distributed Enterprise", Internet Research: Electronic Networking Applications and Policy, 8, 5, 452-465.

Galagan, P.A, 1997, "Smart companies", Training & Development, 51, 12, 20-24.

Gargano, M.L, Raggad, B.G, 1999, "Data mining - a powerful information creating tool", OCLC Systems & Services, 15, 2, pp. 81-90

Garvin, D, 1993, "Building a learning organization", Harvard Business Review, Jul/Aug, 78-91.

Garvin, D.A, 1994, "Building a learning organization", Business Credit, 96, 1, 19-28.

Gates, W.H, 1999, Business @ the speed of thought: succeeding in the digital economy, Penguin Books, London, p 414-421.

Grant, R.M, 1991, "The resource-based theory of competitive advantage: implications for strategy formulation", California Management Review, 30, 3, 114-35.

González, E.L, Fernández , M.A.R, 2000, "Genetic optimisation of a fuzzy distribution model", International Journal of Physical Distribution & Logistics Management, 30, 7/8, 681-696

Gopal, C, Gagnon, J, 1995, "Knowledge, information, learning and the IS manager", Computerworld, 29, 25, SS1-7.

Hamel, G, 1998, Strategic Flexibility: Managing in a Turbulent Economy, Wiley, Chichester, UK.

Hendry, C, Arthur, B, Jones, A, 1995, Strategy through People, Routledge, London.

Hirschorn, L, Gilmore, T, 1992, "The new boundaries of the boundaryless company", Harvard Business Review, May/June, 104-15.

Hitt, W.D, 1995, "The learning organization: some reflections on organizational renewal", Leadership & Organizational Development Journal, 16, 8, 17-25.

Holland, J.H, 1975, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI.

Hong, J, Kuo, C, 1999, "Knowledge management in the learning organization", Leadership & Organization Development Journal, 20, 4, 207-215.

Howitt, P, 1996, "On some problems in measuring knowledge-based growth", Implications of Knowledge-based Growth for Micro-Economic Policies, The University of Calgary Press, Calgary, 15.

Huber, G.P, 1991, "Organizational Learning: The Contributing Processes and the Literature," Organization Science, 2(1), 88-115.

Hyland, P, Beckett, R, 2002, "Learning to compete: the value of internal benchmarking", Benchmarking: An International Journal, 9,3, 293-304.

Kaplan, R.S, Norton, D.P, 1996, Translating Strategy into Actions: The Balanced Scorecard, Harvard Business School Press, Boston, MA.

Kerwin, K, 2000, "At Ford, e-commerce is job 1", Business Week, Feb, 74-8.

Kettenring, J, Pregibon, D, 1996, Committee on Applied and Theoretical Statistics: Workshop on Massive Data Sets, Washington, D.C.

King, W.R, 2001, "Strategies for creating a learning organization", Information Systems Management, 18, 1, 12-21.

Laffey, T, et al, 1988, "Real-time knowledge-based system", AI Magazine, 9, 1, 27-45.

Lee, S.J, Siau, K, 2001, "A review of Data Mining Techniques", Industrial Management & Data Systems, 101, 1, 41-46.

Lewin, A.Y, Minton, J.W, "Determining organizational effectiveness: another look, and an agenda for research", Management Science, 32, 1986, 514-38.

Lorange, P, 1996, "Developing learning partnerships", The Learning Organization, 3, 2, 11-19.

Lu, H, Setiono, R, Liu, H, 1996, "Effective Data Mining Using Neural Networks", IEEE Transactions on Knowledge and Data Engineering, 8, 6, 957-61.

McAdam, R, Reid, R, 2000, "A comparison of public and private sector perceptions and use of knowledge management", Journal of European Industrial Training, 24, 6, 317-329.

Meso, P, Smith, R, 2000, "A Resourced-based view of organizational knowledge management systems", Journal of Knowledge Management, 4, 3, 224-234.

Michalisn, M, Smith, R, Kline, D, 1997, "In search of strategic assets", International Journal of Organizational Analysis, 1-39.

Miller, J.G, 1995, Living Systems, University Press of Colorado, Niwot.

Morris, R, 1996, "Developing a Mission for a Diversified Company", Long Range Planning, 29, 1, 103-115.

Nevis, E.C, DiBella, A.J., Gould, J.M., 1995, "Understanding organizations as learning systems", Sloan Management Review, 36, 2, 73-85.

Nonaka, I, 1991, "The knowledge creating company", Harvard Business Review, Nov/Dec, 96-104.

Örtenblad, A, 2001, "On differences between organizational learning and learning organization", The Learning Organisation, 8, 3, 125-133.

Peteraf, M.A, 1993, "The cornerstones of competitive advantage: a resource-based view", Strategic Management Journal, 14, 179-91.

Piatetsky-Shapiro, G, Frawley, W.J, 1991, Knowledge Discovery in Databases, AAAI/MIT Press.

Quinn, J.B, Anderson, P, Finkelstein, S, 1996, "Managing professional intellect: making the most of the best", Harvard Business Review, March/April, 71-80.

Raggad, B.G, 1996, "Neural network technology for knowledge resource management", Management Decision, 34, 2, 20-24.

Rollo, C, Clarke, T, 2001, International Best Practice: Case Studies in Knowledge Management, Standards Australia, Sydney.

Rosenblatt, F, 1962, Principles of Neurodynamics, Spartan, New York, NY.

Ryan, H.W, 1995, "Losing Clarity: Data Information, and Knowledge Administration," Information Systems Management, 12, 4, 56-58.

Senge, P.M, 1990, "The leader's new work: building learning organizations", Sloan Management Review, Fall, 7-23.

Senge, P.M, 1994, "Moving forward, thinking strategically about building learning organizations", Senge, P.M, Kleiner, A, Roberts, C, Ross, R.B, Smith, B.J, The Fifth Discipline Field Book, Strategies and Tools for Building a Learning Organization, Doubleday, New York, NY, 15-47.

Senge, P.M, 1990a, "The leader's new work: building learning organizations", Sloan Management Review, 23, 1-17.

Senge, P.M, 1990b, The Fifth Discipline: The Art and Practice of the Learning Organization, Doubleday, New York, NY, 423.

Skyttner, L, 1996, General Systems Theory : An Introduction, Macmillan, London.

Soliman, F, 1998, "Optimum level of process mapping and least cost business process re-engineering", International Journal of Operations and Production Management, 18, 5, 810-16.

Sviokla, J.J, 1996, "Knowledge workers and radically new technologies", Sloan Management Review, Summer, 25-40.

Szulanski, G, 1996, "Exploring internal stickiness: impediments to the transfer of best practice within the firm", Strategic Management Journal, 17, winter special issue, 27-43.

Tapscott, D, 1996, The Digital Economy, McGraw-Hill, New York, NY, p95-121.

Technology Forecast: 1997, Price Waterhouse World Technology Center, Menlo Park, CA.

The Economist, 1996a, "The world economy survey", The Economist, 10.

Thuraisingham, B, Venkataraman, V, 1992, "A New View of Information Modeling: A Bridge Between Data and Information," Information Systems Management, 9, 2, 29-36.

Tillet, S.L, 2000, "Reading, writing and Data Mining", Internetweek, Issue 823, p18.

Thibodeaux, M.S, Favilla, E, 1996 "Organizational effectiveness and commitment through strategic management", Industrial Management & Data Systems, 96, 5, 21-25

Tufte, E.R, 1983, The Visual Display of Quantitative Information, Graphics Press, Cheshire, CN.

Weiss, S.H, Indurkhya, N, 1998, Predictive Data Mining: A Practical Guide, Morgan Kaufmann Publishers, San Francisco, CA.

Wernerfelt, B, 1984, "A resource-based view of the firm", Strategic Management Journal, 5, 171-80.

West, P, Burnes, B, 2000, "Applying organizational learning: lessons from the automotive industry", International Journal of Operations & Production Management, 20, 10, 1236-1252.

Wexler, M.N, 2001, "The who, what and why of knowledge mapping", Journal of Knowledge Management, 5, 3, 249-264.

Zadeh, L, 1962, "From circuit theory to system theory", Proceedings of Institution of Radio Engineers, 50, 856-65.

Internet Sources :

Burk, M, 1999, Knowledge Management: Everyone Benefits by Sharing Information, http://www.fhwa.dot.gov/km/prart.htm.

Malhotra, Y., 1998, "Knowledge management for the new world of business", @brint.com (Online). http://www.brint.com/km/whatis.htm