

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Mapping the relationship between knowledge management and information architecture

A dissertation submitted in fulfilment of the requirements for the award of the degree of Master of Philosophy

Under the supervision of Dr. J.G. Smith

Faculty of the Humanities

University of Cape Town

April 2007

Louise Taljaard (tljlou001)

COMPULSORY DECLARATION

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Signature: Signed by candidate Date: 10 April 2007

Abstract

This dissertation defines knowledge in terms of traditional epistemological ideals and as a strategic resource. Knowledge management is defined in terms of the ability of organizations to manage knowledge as a strategic resource in order to gain an advantage from it. In the knowledge management framework, knowledge is presented as a continuum consisting of tacit, implicit and explicit knowledge. Tacit and implicit knowledge is managed through the acknowledgement of the social nature of knowledge. One method to achieve this is communities of practice. On the other end of the spectrum, explicit knowledge is very close in nature and character to information. Due to the expansion of available information resources the design and structure of information (explicit knowledge) for effective retrieval has become very important.

Information architecture is a field that specializes in the design and structure of information for effective retrieval. Traditional information architecture tools such as metadata and subject classification address some of the issues, but experience difficulty in heterogeneous environments such as the Internet. Topic maps are considered as a possible solution to the concerns of metadata classification and subject based classification. Due to the extent and nature of the information recorded in a topic map, it becomes an information resource in itself.

Topic maps also act as an enabling technology for knowledge management as it maps the complex relationships between concepts and include a range of information resources. The conclusion of this dissertation is the representation of a conceptual model based on the themes developed in this dissertation. The main advantage of the conceptual model is the clear and direct link between knowledge management and information architecture.

Table of contents

ABSTRACT	1
CHAPTER 1 – INTRODUCTION	1
1.1 Introduction.....	1
1.2 Information age.....	2
1.3 Knowledge and information.....	4
1.4 Epistemological qualities of knowledge	6
1.5 Rational for dissertation.....	8
1.6 Research problem and research questions	9
1.7 Goals and Objectives	10
1.8 Methodology and research design	10
1.8.1 Qualitative research	11
1.8.2 Exploratory research	13
1.8.3 Inductive approach.....	13
1.8.4 Grounded theory	13
1.8.5 Unobtrusive research and content analysis	15
1.8.6 Objectivity, reliability and validity	16
1.8.7 Conceptual models.....	17
1.9 Dissertation outline	18
CHAPTER 2 – KNOWLEDGE AND KNOWLEDGE MANAGEMENT	20
2.1 Introduction.....	20
2.2 Factors of production and resources.....	22
2.3 Knowledge and sustained competitive advantage.....	25
2.3.1 Communities of Practice.....	29
2.3.2 Concluding remarks.....	32
2.4 Definitions and understanding of knowledge management.....	32
2.5 Tacit and explicit knowledge.....	36
2.6 Conclusion	38

CHAPTER 3 – INTRODUCTION TO INFORMATION ARCHITECTURE	41
3.1 Information architecture.....	41
3.2 What is information architecture?	42
3.3 History and development of information architecture as a discipline	42
3.4 The Zachman Framework	43
3.5 Information architecture shift from technology to information.....	46
3.6 Information retrieval.....	47
3.7 Information seeking behaviour.....	49
3.8 User experience and usability	53
3.9 Conclusion	54
CHAPTER 4 – INFORMATION RETRIEVAL AND INFORMATION ARCHITECTURE	57
4.1 Introduction.....	57
4.2 Recall vs. Precision.....	57
4.3 Classification and classification schedules.....	58
4.4 Metadata	60
4.4.1 Dublin Core.....	61
4.4.2 Concluding remarks	62
4.5 Subject indexing.....	63
4.5.1. Controlled vocabularies	64
4.5.2. Natural language indexing	67
4.6 Conclusion	70
CHAPTER 5 – TOPIC MAPS	72
5.1 Introduction.....	72
5.2 Topic maps in general.....	73
5.3 Topic names and types.....	74
5.4 Associations	75

5.5	Occurrences	77
5.6	Scope.....	78
5.7	Subject identity.....	79
5.8	How topic maps promote information retrieval	80
5.9	Topic maps as an information architecture tool	82
5.10	Topic maps as enabling technology for knowledge management.....	82
5.11	Why topic maps are not more widely applied	84
5.12	Conclusion	84
CHAPTER 6 – DISCUSSION AND CONCLUSION		87
6.1	Introduction.....	87
6.2	Summary of main conclusions derived from the literature	87
6.3	Conceptual Model	95
6.4	Suggestions for further research	101
6.5	Conclusion	103
REFERENCE LIST.....		1

Chapter 1 – Introduction

1.1 Introduction

Information is playing an ever-increasing role in society and in the modern economy. This has been fuelled by better technology that enables the easy gathering and dissemination of information. It is, however, paradoxical that due to the explosion in information available many suffer from information overload and resort to using only centralised sources of information. These factors have thus led to an increase in information related professions and industries that focus on reducing the information overload and providing the right information to the right person at the right time. A typical example is the evolution during the so-called information age of Knowledge Management (KM) and Information Architecture (IA). Although these fields have generally been accepted and the terms commonly used in many spheres of professional and academic life, they are still in the process of being defined and the relationships between them are still being explored.

This dissertation therefore aims to contextualise the current debate regarding the conceptual issues that relate to knowledge management and information architecture. A framework was thus developed to map the relationship between concepts and to specifically focus on the role that information architecture can play in supporting and promoting knowledge management in the information age and modern economy. A consequential outcome of the investigation was the realisation that topic maps can effectively be used both as a tool to address the short comings of traditional information architecture techniques and as an enabling technology for knowledge management. This technique is then further employed to depict the relationship between information architecture and knowledge management.

1.2 Information age

The term, information age, has been used in various contexts, but it is not a well-defined term and means different things to different role players. For example, many definitions focus on digitisation as the driver for the convergence of different information sources, such as television, computers and multimedia. For politicians one of the consequences of the information age is that communities and countries are becoming borderless and interconnected. From a business and economic point of view the information age means that information is regarded as a critical resource (see discussion in chapter 2.2) and businesses have to evolve into learning organizations, while workers become knowledge workers (Jastroch, 2001).

A number of visionaries, like Daniel Bell (1974) and Peter Drucker (1969), have predicted since the mid 20th century that knowledge and information would begin to play a larger role in the economic activities of mankind. More specifically, they commented on the increased knowledge expectations of workers, both on factory shop floors and in offices, by employers. Workers would be expected to contribute to the organization through better work practices, not only through more hours or physical output. This leads to a very different working environment that includes different workspaces and changes in the organization of co-workers.

In 1974 Bell wrote that the post-industrial society would be a knowledge society. He supplied two reasons for this argument. Firstly, a greater number of innovations would evolve due to research and development, which is a knowledge activity. Secondly, an increasing percentage of Gross Domestic Product (GDP) and employment was occurring in the knowledge field. Both Bell (1974:374) and Drucker (1969:40) predicted that a new class would emerge within the economic society; namely that of a professional class that will derive their wealth from knowledge rather than property. This means that knowledge would be the key resource and not physical assets as had been in the past agricultural or industrial age.

The start of any economic age is hard to pin down. As changes take place gradually over a period of time a new phase is born. Bell (1974:346) saw the period between 1945 and 1950 as the estimated “birth years” of the post-industrial society. During this period the first digital computers were developed and they spread quickly and the pace of development increased. Therefore, there is a direct link between the post-industrial or information age and the advances in technology, especially in information technology and computers.

When the post-industrial society is compared to the industrial society, the major difference is that the industrial society was a “goods producing society” whereas “the post industrial society is an information society” (Bell, 1974:467). This does not mean that society will cease to produce goods. Rather, it means that value is also generated where information is used and applied and not solely through the production of goods. This is a complex relationship. If information is used to produce goods more cost effectively then it might seem that the value is added at production. However, it is the appropriate application of information that is adding the value. The concept of information and knowledge being valuable and strategic resources will be further addressed and developed throughout this dissertation.

The Internet¹ and the World Wide Web² (WWW) are manifestations of the information age. One of the advantages of the Internet is that it has given people the opportunity to interact more with the media presented to them. In addition the Internet has made larger amounts of information available. The importance and possibilities embodied in the Web “implies a new information-driven society that implies different kinds of work and forms of knowledge” (Burnett & Marshall, 2003:10).

iiiiiiiiiiiiii—

¹ Definition of Internet: An interconnected system of networks that connects computers around the world via the TCP/IP protocol. (American Heritage Dictionary)

² Definition of World Wide Web: The complete set of documents residing on all Internet servers that use the HTTP protocol, accessible to users via a simple point-and-click system. (American Heritage Dictionary)

The development of the Internet and the World Wide Web was part of a greater development in globalisation that made it important to move information around the globe. Although globalisation and the Internet clearly have a close relationship, the nature of the relationship is not always explicitly understood. Globalisation is an economic development that embraced the advantages provided by the Internet, rather than the changes in technology itself causing globalisation (Burnett & Marshall, 2003:18). However, technology is causing countries to become borderless and the number of participants in a market and the products available to consumers has increased substantially. There are also a greater number of distribution channels available to organizations. The result is that organizations need more information about their competitors, target markets, etc. Organizations also need this information at a greater speed, because they need to react more quickly. Therefore, the role of information and who possesses the information becomes more important. For this reason, information and the application of it in the form of knowledge can demand a premium because of the value it adds (Prusak, 2001:1002).

The greater availability of information affects both the inputs used in organisations and the output that they produce. As Drucker (2001) pointed out, having access to this amount of information will not only change the way production is done, but also what type of “production” is done. In terms of monetary value, knowledge has become the key ingredient in most goods and services. It is the source of the value-added portion in many production facilities, such as computers and other machine hardware (Drucker: 2001).

1.3 Knowledge and information

In order to place the discussion surrounding knowledge management and information architecture in context, the concepts knowledge and information have to be clarified. When considering the differences between knowledge and information, the knowledge hierarchy is a natural start (Ponelis and Fairer-Wessels, 1998:3). Within the knowledge hierarchy, data is the term used to refer to values, numbers, characters or symbols that can be captured, stored or processed (Concise Dictionary, 2000). Therefore, data can also

refer to “structured records of transactions” and form the raw material from which information is created (Davenport & Prusak: 1998: 2). Information is the term used to describe data that has been formatted, summarised and generally processed - i.e. given substance (Beckman, 1999:5).

Davenport and Prusak (1998:3) see information as a message that is transmitted between a sender and a receiver. The communication between the two parties is considered to be information sharing if the sender has added value to the original data in the view of the receiver. If the receiver perceives that there was no value added, then the communication is said to consist of data and not information.

Ponelis and Fairer-Wessels (1998:2) regard knowledge as information applied within context and experience. As one moves up the knowledge hierarchy the concepts are increasingly subjective and abstract, while the quantity decreases. For example, much data can be captured and collected; from this trends (information) can be identified, which in turn may result in one course of action (knowledge). Another aspect of the knowledge hierarchy is that the higher levels also require increased human involvement and therefore imply greater value added (Ponelis & Fairer-Wessels, 1998:2). The greater value attributable to knowledge compared with information and data is its relative nearness to action (Davenport & Prusak, 1998:6). This distinction is especially important in an organisational environment as opposed to the discussion within epistemological contexts. From the discussion above it can be seen that knowledge is scarcer than information. Despite the perceptions of information overload, knowledge is a rare commodity.

As data is the source for information, so information is the source for knowledge. But the process of converting information into knowledge requires a greater human input than converting data into information (Davenport & Prusak, 1998:6). Some authorities argue that knowledge can only exist within humans due to the inherently personal nature of knowledge. As part of this discussion, we will now consider the nature of knowledge in more detail.

1.4 Epistemological qualities of knowledge

Epistemology is much more than the science and study of knowledge and is concerned with the properties expected of our beliefs. As a study of knowledge, epistemology aims to define knowledge and to determine what can be considered legitimate knowledge (Aarons, 2006:166). This provides a theoretical background from which to consider knowledge and its function within an organization. One of the ideals of epistemology is that our beliefs should be coherent, because coherent beliefs are perceived to be more likely to be true and justified. A belief is considered coherent if it adheres to the logical principles of its field and can be substantiated when seen in connection with other beliefs. Beliefs can be incoherent for two reasons. Firstly, they could fail to apply the principles of logic. Secondly, beliefs are incoherent when they exist despite strong evidence to the contrary. The problems with incoherent beliefs are that they are difficult to defend and therefore, they may be false. Although, coherency is the ideal, it cannot occur in 100% of the cases. People have different beliefs and different views on the sources of information that support their beliefs. It can be that the perceived credibility of the sources that inform our beliefs, cause the difference in the perceived coherence of beliefs between different people (Morton, 2003: 1-3).

There are a number of concepts within epistemology that have to be clarified for further discussion. Firstly, beliefs can be rational or irrational, depending on how that belief was arrived at. For example: if someone has a belief despite evidence that points to the contrary it is both irrational and incoherent. Irrational beliefs also often include words such as "all", "never" and "always" (Morton, 2003: 4-5). The quality of information that informs decisions and knowledge will not affect beliefs if they are irrational. The irrational belief may be changed if the source of the original information is discredited or if additional information leads to another conclusion. For the purpose of this dissertation, the focus will be on how to use information in order for an individual to have as many as possible true and justified beliefs.

A belief is true if it describes the world accurately or if the belief exists because the reality of the belief exists. Alternatively, a belief is false if it is not true, but still has meaning. A meaningless belief is neither true nor false. It is not a given that truth and rationality will always go together. For example something might have strong historical evidence to suggest that it is true, but in fact it is false. An example would be if a conclusion was reached based on all the available evidence. If a key piece of information (that would change the conclusion) is missing, but you are unaware of it, the belief resulting from the conclusion would be rational. However the belief is untrue as it does not reflect the reality of the world. The belief is rational as it is based on sound reasoning and the available evidence, but the belief is not true (Morton, 2003: 5, 181, 183). It is very important that people will have the right information in order to influence the truthfulness of their beliefs. If knowledge is based on faulty information the person may believe that his/her conclusion is appropriate while it is not.

The issue of evidence is very important for the discussion on beliefs. Evidence is the reason for holding a certain belief and is often derived from an experience (Morton, 2003: 181). Evidence is often presented in order to change someone's beliefs. However, their belief should only change if they consider the proof rationally and believe the evidence as presented to them. The process of considering evidence is known as reasoning, although reasoning can also take place without evidence. When reasoning is done without evidence it is a mental exercise (Morton, 2003: 5-6).

Evidence also affects the justification for our beliefs. If the evidence is misleading but we have no reason to suspect that it isn't true, then our belief based on that information/evidence would be justified even if the belief is false. A justified belief is one supported by evidence and correct reasoning. The greater the number of justified beliefs one holds the greater the possibility that one has many true beliefs. Many theories of knowledge focus on when beliefs are justified and how to acquire knowledge rationally (Morton, 2003: 6, 182).

The basic qualities of beliefs held by individuals should be truth, rationality and good evidential backup. Although it is not always possible to obtain all of the characteristics, aiming for these basic qualities should allow an individual or an organization to ensure that most of the knowledge contained within it is useful and timely.

1.5 Rational for dissertation

The above discussion about knowledge and beliefs highlights the importance of evidence in the construction of coherent and justified beliefs. The information people collect throughout their lives informs their beliefs and knowledge about their surroundings and their world. In an organizational setting it is important that employees and stakeholders have access to valid and up-to-date information to inform their decisions. The information age (and the technologies that accompany it) has made access to information easier, but there is debate about the quality of information available on the Internet, which includes the drawbacks of information overload. When information resources are poorly supported, then knowledge resources are poorly supported. Information architecture and knowledge management are two fields that aim to address the support and management of information and knowledge resources.

Both information architecture and knowledge management came to the fore as autonomous and distinct subject fields in the 1990's. Their emergence coincided with the development and rapid expansion of information technology. Both have appeared in the corporate literature as an important management concept for the new information based economy. However there is a lack of discussion in the existing literature on the specific relationship between knowledge management and information architecture and how they influence each other. The purpose of this research project is thus to help define information architecture and knowledge management and present a model of the relationships between them. It is envisioned that this will contribute to a better understanding of both the academic and practical aspects of these fields. In addition, the 2006 Information Architecture Summit listed *Firm wide information architecture: How information architecture supports knowledge management* as a potential topic for

discussion in its call for submissions (www.iasummit.org/2006/index). This indicates an interest in the topic from those currently involved in information architecture and shows the relevance of this research project.

1.6 Research problem and research questions

The primary research problem investigated in this dissertation relates to the lack of understanding about the relationship between knowledge management and information architecture. From this the following main research question that underpins this study evolved:

- What is the relationship between knowledge management and information architecture?

This lead to the subsidiary questions of

- How can each field and their relationship be clearly conceptualised?

In the process of investigating the above primary and subsidiary research questions, a number of related questions arose. The inductive and exploratory research processes (cf. the discussion of methodological issues in chapter 1.8.4) that were adopted for this project identified key shortcomings in certain core components in the field of information architecture. This lead to the research question:

- What technologies or tools exist that could improve traditional information architecture techniques?

Topic maps, as discussed in chapter 5, emerged as a potential answer to this question and while investigating and obtaining a better understanding of topic maps the following additional questions evolved:

- What potential do topic maps hold for the field of knowledge management?

And

- Can topic maps be used to clearly explicate the relationship between knowledge management and information architecture?

Therefore, the reiterative process of data collection and analysis to answer the original research question lead to more questions which aim to address the original problem, but from a different angle and with a more specific focus.

1.7 Goals and Objectives

The aim of this research project is therefore to derive a conceptual model that investigates the linkages between knowledge management and information architecture. Although these fields are linked conceptually, explicit linkages have not been made in the literature. In the process of clarifying these linkages, knowledge management and information architectures will be mapped and the issues involved discussed. Potential solutions, such as topic maps, will also be discussed.

The outcome of the investigation will be in the form of a conceptual model that explicates and contextualises these relationships. The use of topic maps as a tool in this process will further also be demonstrated.

1.8 Methodology and research design

Different branches of learning were consulted in the research process for this dissertation. The conceptual fields of knowledge management, information architecture and topic maps derive from a number of, often disparate, disciplines which are represented in a wide variety of academic literature that ranges from library sciences, computer science, management studies and information systems to various other lesser fields. Although, some fields have their roots in the more exact sciences, this research dissertation will draw most of the research design from social science research methods. This decision was based on the premise adopted by the researcher that knowledge management and information architecture, although grounded in various disciplines, have a distinct humanistic focus and since

“Social science research is a collaborative human activity in which social reality is studied objectively with the aim of gaining a valid understanding of it” (Mouton & Marais, 1990: 7)

this would be an appropriate procedure to adopt.

As outlined in 1.7 the aim of this dissertation is to review existing practices, as represented in the literature, of knowledge management and information architecture as human activities and then to evaluate and synthesise these findings, in order to answer the research questions as outlined in section 1.6.

1.8.1. Qualitative research

Research designs are often divided into two groups: qualitative and quantitative. One of the main differences between qualitative and quantitative research approaches is that the researcher in quantitative research aims to be an objective outsider to the subject matter, while with qualitative research, the research is conducted from an insider’s point of view. The goals of qualitative and quantitative research also differ. Qualitative research aims to describe and understand, while the focus in quantitative research is explanation and prediction. Because these research paradigms have different starting points and different goals they also often have different methods of data collection and data analysis.

Tesch (1990: 3) points out that “qualitative data” might be a confusing term in the English-speaking world, leading people to think that relates to superiority of the data, i.e. meaning high quality data. This is not the case. Qualitative data has come to mean information that is non-numerical and often textual in nature.

Qualitative research has greater flexibility and it employs a wider variety of strategies to collect and analyse data than quantitative research. This is because new insights are expected to emerge during the research phase, which can influence certain changes to the original research design, naturally leading to a more flexible approach (Bradley, 1993: 434). Other authors further point out that qualitative researchers are not eager to standardize and set in stone the methodological aspects of qualitative research and

analysis (Tesch, 1990: 4; Ambert, et al. 1995: 881). As the focus of social science research is the social workings of people or groups of people (and this can be very unpredictable) researchers need to be flexible and creative in their approach to studying social phenomena and hence the applicability of qualitative research methods.

According to Ambert, et al (1995: 880) qualitative research has a number of goals. The first is that it seeks an in-depth understanding instead of a broad general understanding. Secondly, qualitative research is more concerned with the how and why of human behaviour as opposed to the 'what'. Qualitative research goals can also be situated on various levels, ranging from micro to macro. Importantly, qualitative research is concerned with discovery rather than verification of data. This gives the research the advantage that it is not limited by existing knowledge, but can add to current debate and expand on existing understandings of certain phenomena. Finally, Ambert, et al (1995: 881) further suggest that qualitative research is often aimed at theory building, through the emergent nature of the process. Concepts are developed based on the data and then "shaped and reshaped according to the results of ongoing observations".

Based on the above discussion of the characteristics of qualitative research and the nature of this research project, a qualitative approach was adopted. The focus is on describing and understanding knowledge management and information architecture, rather than predicting or forecasting behaviour in these fields. The data used and the subsequent analysis is non-numeric and textual since various written documents are analysed. As the research progressed the direction of enquiry and focus shifted as new information came to light, especially about topic maps and the application opportunities for information architecture and knowledge management. In addition, one of the main aims of this research project is to build on existing literature and to add new insights and theories for future research. This aim thus requires an explorative approach to the research design.

1.8.2. Exploratory research

An explorative research design is specifically flexible to encourage extensive consideration of all relevant factors. This study, as many other exploratory studies, has the goal of formulating a conceptual model that can be tested in future studies.

Exploratory research may be conducted by means of three methodological approaches. The first is a review of related and relevant literature. The second method is to question persons who are familiar with the area of study and thirdly, examples can be studied that highlight some of the issues related to the area of study. These methods can be applied in isolation or in combination (Selltiz, et al, 1965: 50-52). This dissertation has focused primarily on the review of existing literature.

1.8.3. Inductive approach

The process of analysing different sources and creating a conceptual model based on the material collected follows the inductive approach to research. Induction is the process of generalizing or creating a theory based on the observations that were made (Blaikie, 2004: 486). The opposite of the inductive approach in logic is the deductive process. When deductive logic is applied in research, specific hypotheses are developed based on the general theories and tested to see if they prove or disprove the general theory (Rubin & Babbie, 1993:689).

The advantages of inductive theory construction are that new theories can be developed, without being constrained by existing theories (Rubin & Babbie, 1993: 52). This provides greater flexibility for hypotheses development and is more likely to lead to new discoveries. The disadvantages are the potential lack of objectivity and the potential for selection bias. (See discussion on objectivity, reliability and validity in chapter 1.8.6.)

1.8.4. Grounded theory

Grounded theory is one method that is often used to analyse qualitative data inductively. The aim of grounded theory methodology is to construct theories through systematic and

structured qualitative data analysis. The data analysis is achieved by coding text based on the properties and characteristics of the text and then comparing text with the same coding in order to identify emerging themes (Charmaz, 2004: 440-444; Tesch, 1990: 64).

Grounded theory was originally developed by B.G. Glaser and A.L. Strauss. Their original approach to grounded theory placed the emphasis on allowing themes and theories to emerge from the data, without forcing it. Glaser (1992:21) even suggests that one should not formulate a research problem or question before analysing the data or literature in the field of interest, but to let the problem emerge from the data. This is achieved by first using open coding, where data is coded with no preconceived concepts in mind. After the initial open coding, conceptualisation takes place by comparing different concepts and incidents and then categories are developed from the emerging patterns. This can lead to further, more focused, data collection. This again leads to more analysis, until a theory has been developed (Glaser, 1992).

Grounded theory is then based on symbolic interactionism and this means that analysis takes place while data collection is taking place. Despite the interactionism, the first phase of the research is dominated by collection while the second phase is dominated by analysis. Categories are developed from key concepts chosen by the researcher and/ or derived from the data. These categories should not limit the expansion of the topics for the researcher. With further analysis more categories emerge and through better understanding some categories amalgamate or are excluded. Also, once categories are understood the relationships between them should be explored and explained. Categories are also clustered together as conceptual entities and this helps with the analysis and development of theory. Visual mapping can also help to analyse the concepts as they emerge. Tesch (1990: 85) further points out that grounded theory not only seeks to identify a pattern, but also to explain that pattern. This is where theorizing comes in and the development of a conceptual model. Even though the research process is not rigid it continues to be systematic and comprehensive (Tesch, 1990: 86 - 99).

The principles of grounded theory informed the methodology for this research project. This research project specifically started with an in depth investigation into the meaning and definitions of the main conceptual components of this study, knowledge management and information architecture. Qualitative sources, i.e. journal articles, encyclopaedia entries, books and electronic sources were consulted and the literature grouped into knowledge management or information architecture. From the literature a number of subsidiary concepts were identified that are related to these fields of study and which could potentially influence the relationship between them. Examples of such subsidiary terms included information, knowledge, taxonomy, intellectual capital and “communities of practice”. At one stage more than 50 related terms had been identified. A pattern emerged which pointed to the lack of a definitive description of the relationships between information architecture and knowledge management. This was followed by an in-depth comparative evaluation of the two domains. Out of this analysis, a theory evolved that a positive relationship exists between the two. Further investigation focused the discussion on information retrieval techniques as a key concern for information architecture. This also narrowed the number of relevant terms. The investigation of information retrieval technologies lead to the identification of topic maps as a very important technology that addresses the shortcomings of traditional information retrieval techniques. While further investigating topic maps, it was found that it could also be fruitfully employed as an enabling technology in knowledge management. Finally, it was seen that topic maps could very effectively be used as a visual mapping technique (see chapter 6.3) to clearly explicate the relationships between the various concepts for better clarification and definition.

1.8.5. Unobtrusive research and content analysis

The method described above is a form of unobtrusive data collection and more specifically content analysis. Unobtrusive research is undertaken when data is collected in a way that the researcher does not interact directly or indirectly with the subject matter. Content analysis is one way of conducting unobtrusive research and can be applied to any form of written material. Although content analysis can be performed in a quantitative manner, this dissertation undertook to evaluate the sources in a qualitative manner.

The advantages of content analysis in this context include the saving of time and money, relative to doing a survey. Conducting a content analysis study allowed the researcher to access more material from a geographically diverse background. This form of research design has also allowed the researcher to interact with the material in an unobtrusive manner (Babbie & Mouton, 2001: 374-393). The disadvantage of content analysis and specifically within this research project is that the analysis is limited to written material only.

Although, every attempt was made to investigate as comprehensively as possible a representative sample of material, it was not feasible to attempt to investigate all material in the field. Especially since new and potentially relevant material is released continuously. Every effort was however made to include the most relevant and up to date material.

1.8.6. Objectivity, reliability and validity

Reliability refers to the likelihood that the same result will be achieved if the study is repeated, in other words how consistent the results will be if the tests are rerun. In quantitative analysis, reliability is measured using statistical methods. In qualitative analysis the measure is slightly different, because the nature of qualitative research is such that it is often difficult to repeat exactly the same study, i.e. achieve 'external' replication. (Smith, 2004: 957; Dey, 1993:221). The researcher can still aim to achieve 'internal' replication, which refers to "the principle of ensuring that such procedures as are followed can stand up to scrutiny" (Day, 1993:221). In this dissertation, materials used are clearly referenced and the logical reasoning explained. This was done with the aim of achieving internal replication, which improves the reliability of the findings.

Validity, on the other hand, is a more complex and important measure. Validity of research refers to the ability of the research to accurately reflect the area under consideration. Face validity means that the research project conforms to the general consensus about a certain topic. This means that although researchers might differ about the detail or the extent of the validity, they will agree that it is valid to some point. In this

research project, every effort was made to clearly explain concepts, especially in the light of their ‘face validity’. A form of triangulation was also employed in order to enhance validity. Many different sources were investigated in this research process, and although new connections and explicit relationships were identified, consensus from the material regarding definitions of concepts was sought (Babbie & Mouton, 2001: 122-124; Smith, 2004: 957).

Objectivity is related to validity and refers to the ability of the research conclusions to reflect reality accurately, i.e. for the conclusions to be true. Various factors can negatively influence objectivity, for example researcher bias and motives. Again, every effort was made in this dissertation to accurately reflect the results of the enquiry without bias. However, due to the inductive nature of the research project it is acknowledged that full objectivity cannot be guaranteed (Hammersley, 2004: 250-251; Babbie & Mouton, 2001: 275).

1.8.7. Conceptual models

Science is supported by logical thought and observations of the world around us. Our beliefs have to be rational and correspond to the observations done in the real world. In general, scientific activity involves three areas: theory, research methods and statistics. Theory represents the logical and conceptual aspects; research methods relate to the rigorous observation methods that are employed to collect data and statistics evaluate that which was observed within a logical framework and expectations (Babbie, 1995: 26).

Engelbart (1962: 33) defines a conceptual model as a “theoretical model with which to organize our thinking and action”. Because conceptual models represent theory, they also underpin all research, either explicitly or implicitly. A conceptual framework becomes even more important when that specific area of study is in an era of change (Jarvelin & Wilson, 2003). Although knowledge management and information architecture have been part of academic discourse for some time, the two fields are still in the process of defining themselves and of establishing identities. Furthermore, the relationship between

the two fields of study has not been clearly conceptualised. All these facets clearly point to a need to develop a conceptual model that explicates such a relationship.

When developing a conceptual model certain aspects should be specified. Firstly, the essential components that relate to the system to be investigated should be specified. This shows an understanding of the issues involved. Related to that is the specification of recognized relationships between the different concepts. Next, the ways that changes in objects or the relationships between the objects affect the system, should be specified. Finally, potential goals or methods of research should be specified (Engelbart, 1962).

The above discussion on the research design applied in this dissertation attempted to describe how this researcher went about investigating the essential components of the relationship between knowledge management and information architecture. Through the course of this investigation the relationships between these concepts and related topics were identified and specified. The results are represented in a conceptual model (see chapter 6.3). The second aspect suggested by Engelbart is to specify the way that changes in these concepts or their relationship would affect the system. This is beyond the scope of this research, but could form the basis for future research.

1.9 Dissertation outline

Chapter 2 discusses knowledge as a resource in some detail and what this means for management strategies. Knowledge management is also discussed. Chapter 3 introduces the concept of information architecture and considers what aspect of information seeking behaviour needs to be taken into account when designing an information architecture for information retrieval. The tools available to information architects to improve information retrieval are discussed in Chapter 4. This chapter also highlights the challenges of information architecture in relation to effective information retrieval, especially for heterogeneous collections and ambiguous classification. Chapter 5 introduces topic maps as an information architecture tool that addresses the challenges highlighted in chapter 4. This chapter also describes how topic maps support the

objectives of knowledge management. Chapter 6 summarises the main conclusions derived from the discussion in the dissertation, graphically represents the conceptual model developed through the preceding discussion and suggests areas of further research.

University of Cape Town

Chapter 2 – Knowledge and knowledge management

2.1 Introduction

Ultimately, the goal of organizations is to gain a competitive advantage using the factors of production and resources available to them. Many strategic management studies have focused on what enables a firm to gain a strategic advantage and very importantly what enables a firm to sustain that competitive advantage. However, many of the studies used the same framework for their analysis; namely the analysis of strengths and weaknesses of an organization. This type of analysis made implicit assumptions that influence where strategic advancement would be sought. Firstly, it is assumed that all firms within an industry or group have identical strategically relevant resources. In economic theory regarding the traditional factors of production this would mean that all firms in the industry have equal portions of the same land, labour and capital available to them for utilization. Traditional strategic analysis also assumed that any advantage, i.e. additional labour, would be short lived because resources are mobile and will move to the place where they receive the maximum return. However, these assumptions do not necessarily reflect reality (Barney, 1991: 99-100; Wade & Hulland, 2004: 108). Management studies as a discipline is engaged with the question of how to use land, labour and capital to achieve a strategic advantage, even if it is not sustained indefinitely.

The resource based view of the firm is one strategic management theory that addresses the validity of the two assumptions above and attempts to specify the conditions where a firm's resources will act as a source of sustained competitiveness. The resources in the resource based view of the firm are defined to include all assets (tangible and intangible) that is controlled by the firm and that can be used in the execution of the strategies planned by the organization. More specifically, the attributes of a firm are only considered resources when they enable the firm to develop opportunities or eliminate threats. Barney (1991: 101) explicitly includes knowledge and information in the definition of firm resources. The resource-based view of the firm also defines a

competitive advantage as a time when the organization implements a strategy that adds value, which is not being implemented by the competitors of the organization. This means that a competitive advantage is that which helps an organization achieve more than its peers and allows it to survive turbulent economic times. Competitive advantages have to be sustained or new ones created in order for the organization to maintain its advantage and remain economically viable. A competitive advantage is defined as sustainable when current competitors and potential competitors are unable to duplicate the strategy. This means copying, done by competitor firms, will not erode the strategic advantage. This does not mean that the strategic competitive advantage will last forever. The strategic advantage could still be undone through new technology or sudden changes in availability of key resources (Barney, 1991: 100-103, 106).

If all firms in an industry are assumed to have identical strategic resources and those resources are perfectly mobile it is impossible for a firm in that industry to have a sustained competitive advantage. As all firms have the same resources and opportunities all firms will be able to conceive of and implement the same strategies. Any strategic competitive advantage that is gained will not be sustained as competitors could copy it. Therefore if a firm wants to have a sustained strategic competitive advantage they will have to focus on developing resource heterogeneity from other firms in their industry or group (Barney, 1991: 103-104).

However, heterogeneity and mobility are not the only characteristics resources need to have to provide sustained competitive advantages. According to the resource based view of the firm there are four other characteristics that a resource should possess to provide a sustainable competitive advantage. Firstly a resource should have value and be rare. Next the strategically important resources should be “imperfectly imitable” and without strategic equivalents (Barney, 1991: 106-112).

Throughout this dissertation knowledge will be evaluated as a resource, using the criteria set out above to discuss if knowledge could be a source of sustainable strategic advantage and what the implications are for management strategies and specifically the management

of knowledge. The first step is to evaluate knowledge in comparison to traditional factors of production and definitions of resources.

2.2 Factors of production and resources

In the neoclassical economic tradition there are three factors of production: land, labour and capital. Each of these factors represents the resources required to produce goods (and services) and bring economic value to the society or profits to the firm. Land includes all natural resources, such as oil, timber and other raw materials used in production. Capital refers to manmade items used in production and includes the finances used to fund production or buy other resources. Finally, labour refers to the work done by blue and white-collar workers. Each of the factors of production receives their own reward for participation in the production process. Labour receives a wage, capital receives interest and land receives rent. To this can be added the profit that the entrepreneur requires for putting all the factors of production together in an economically viable way. However, the role of the entrepreneur is often grouped together with the activities of labour or the role the entrepreneur plays is neglected in neoclassical economic theory (Felderer & Homburg, 1992: 35; Kermally, 1999: 138).

In economic theory a number of simplifying assumptions are made. Firstly, in the short term land and technical knowledge is taken as constant. This means that the business owner or entrepreneur can only use labour and capital investments to produce the maximum output. Output is given as a function of labour hours and capital stock (Felderer & Homburg, 1992: 35). If however, Bell (1974) and Drucker's (2001) predictions are true, then knowledge (technical and organizational) is not constant or unchanging in the short term. Knowledge can be applied, in the short and long term, to influence the productivity of labour and capital stock. This means that output is not only a function of labour and capital, but also of the knowledge applied to improve their productivity. Because all traditional factors of production are limited to some extent and can be substituted for each other, it can also be argued that the real source of their

productivity is knowledge (Drucker, 1969: 151-152). This means that knowledge could be viewed as the new production factor and as an essential resource in the economy.

The second important aspect of economic theory as it applies to factors of production is that of the law of diminishing marginal returns. Diminishing marginal returns means that every unit of additional production will produce a slightly lesser yield than the unit before it. Although this might not be strictly true from the first unit of production – i.e. the second unit of production will yield a slightly higher return than the first, because the first unit is insufficient to fully exploit the resources used for production – the law of diminishing marginal returns applies to traditional factors of production at the level of perfect competition. In the state of perfect competition firms will aim to optimise the level of production. Optimisation will occur in the range of diminishing marginal returns. While marginal units of production yield an increasing return, production will be increased to take advantage of the marginal increase, until a point is reached where marginal units produce a diminishing return. It is also assumed in this model that the most profitable or rewarding work will be done first and then the next most profitable work as a strategy of optimising profits for the firm (Felderer & Homburg, 1992: 38). Interestingly the law of diminishing marginal returns does not apply to knowledge as a factor of production. Knowledge is the only resource that is not reduced as it is used. Actually, knowledge increases the more it is used. Therefore knowledge provides increasing marginal returns over the whole production spectrum. This has very important applications for economic theory. Importantly, it further reinforces the need to address knowledge as a factor of production and to understand how it will impact on the competitiveness and strategic decisions of organizations.

Knowledge as a production factor is supported by information as a resource. Human resources support labour as a production factor, in terms of people that work to produce goods and services. Natural resources influence the quantity and quality of the production factor land that is available for production. Similarly, technology, machinery and money are the resources that combine to form the production factor capital. Furthermore, the substantial increase in technology does not only result in substituting labour with

technology. The definition of labour in the traditional definitions of factors of production has also changed (Kermally, 1999: 190). Although the terms factor of production and resources are used interchangeably in many instances, there is a distinction to be made. People or human resources are critical to knowledge as a factor of production. However, it is information as a resource that influences knowledge and the effectiveness of the application of knowledge. People are the vehicle through which the information as a resource is given meaning and economic value in the form of knowledge as a factor of production.

The three traditional factors of production— land, labour and capital – each have two aspects that influence the value that they can potentially add to an organisation. The first is the availability and the quality of the resource, for example, the skills and abilities of labour or the quality of land for agricultural production. The second element is how the resources are applied and managed to maximise output. This means that the quality of the resource has to be considered in the context of the abilities of an organisation to develop the resource. In the example of capital the first aspect would be the availability of sufficient funds and amount available for investment. If the organisation has funds, but there are limitations on what the funds may be spent on, then the availability and “quality” of the funds could be incompatible with certain investment decisions. The second aspect is the application of appropriate funding to projects or departments that would be able to make the most efficient use of it. Good financial management in this example will increase the resources (money) available to the organisation. The same applies to knowledge as the new factor of production in the information age. The information resource that is available and generated within an organisation has to be of high quality. In addition, information has to be used, shared and expanded in the most efficient way to maximise the outputs of the organisation. To be truly effective, information as a resource and knowledge as a factor of production have to be managed.

The dichotomy of the factors of production is represented in a Table 1 as understood by the author. An instance of each factor of production is used to illustrate how the concepts of quality and management would be applied.

Table 1

Production factor and supporting resource	Availability and quality	Application and management
1) Land and natural resources: i.e. Agricultural produce	Seeds and crop quality.	Fertilisation and harvesting.
2) Labour: i.e. Programmers	Information technology skills and enough people to programme.	Human resource management and training of personnel.
3) Capital: i.e. Machinery	Technological advancement of machinery and the right number of machines.	Production planning and scheduling of regular maintenance.
4) Knowledge: i.e. Information	Type of information and significance. Also information technology that supports information.	Use and sharing of information. Information and knowledge management

As clarification for the table: In order for agricultural produce to add value, the quality of the seeds and crop quality have to be high enough. The crop yield should also be sufficient to cover cost and provide a return on investment. The management of agricultural produce includes fertilization and harvesting at the appropriate times. There are many types of labour, but if labour is needed to produce a software package, for example, then programmers will be needed. These programmers will need to have the necessary technical skills to execute their duties. The organization will also need the right number of programmers to complete the project within a given time period. In order for a factory to be economically viable the machinery used for production has to have the right level of technological advancement and the optimum number of machines is needed for production. In addition, production has to be planned around the available machinery and maintenance needs to be performed on a regular basis. The above examples highlight a number of different instances where the quality and management needed for each factor of production is important. When the same criteria are applied to knowledge as a factor of production, then the type of information and the accuracy of that information are important. Information also needs to be managed by creating channels through which the information can be shared and used effectively.

2.3 Knowledge and sustained competitive advantage

Although the analogy to other factors of production and resources helps one to understand the importance of knowledge and information within an organizational

context it could lead to the misunderstanding that knowledge is an object. Knowledge is an asset, but due to its intangible qualities it should not be considered an object in the same sense as physical or financial assets. The value knowledge adds to the organization or production process does not happen where the knowledge resides, but rather when the knowledge is transferred, shared or applied. Therefore, knowledge as an object does not have the intrinsic value as a building or bank account, knowledge only has value when it is put into action or used. The value is in the process and therefore when it is said that knowledge is an asset, the knowledge referred too is not an object, but a process or activity (Al-Hawamdeh, 2002).

This has important implications for the management of knowledge. Most importantly it takes the focus of the discussion of “what is knowledge” away from the individual (where the knowledge object resides) to the collective or applied environment (where the process of knowledge exists).

When knowledge is considered as a process it is considered in a social context. Knowledge exists within people’s minds and therefore a knowledge management initiative cannot exclude people from the plans. As a survey conducted in 1997 showed, the main difficulty in knowledge management implementation is “changing people’s behaviour” and the main barrier to knowledge transfer is “culture” (Ruggles, 1998, 87-88). Churchman (1971, quoted by Malhorta, 1999: 37) points out: “Knowledge resides in the user and not in the collection. It is how the user reacts to a collection of information that matters”. Therefore, even though the top technological solutions may be in place, it only acts as complementary to the people and their behaviour. People and their behaviour do not operate in a vacuum, but in a social context. Therefore the social context of knowledge or the “human view” of knowledge needs to be considered.

Because the “human view” initially did not receive sufficient attention the potential impact of knowledge management programmes was not always realised. The disproportionate focus on technology was partially because of the relatively established relationship most organisations had with information technology and information

management by the time that knowledge management and its importance became more prominent. However, this has led to the criticism that knowledge management is just another name for information technology management. As more critics pointed out that knowledge management is intricately linked with the social milieu in which knowledge is used, the literature around the human aspects of knowledge management expanded greatly. However, this is still an area where more research can be done.

Part of the reason for the change in attitude towards knowledge management is the change in the view of knowledge itself. The world and what we know about it used to be considered knowable and predictable. Therefore knowledge was considered to consist of a collection of facts that can be collected, stored and shared with others (Thomas, et al. 2001: 864). This compact view of knowledge made it an easy resource to manage through the effective use of information technology. However, this view only paid attention to the explicit element of knowledge and as a result addressed only a fraction of the challenges associated with knowledge management.

As discussed in the introduction, knowledge can be defined as a personal item and as an "individual's true, justified belief" (Aarons, 2006: 166). Therefore, knowledge is considered a personal attribute and the ideal is that an individual should have as many true beliefs as possible. However, there are other aspects of knowledge in the social context beyond the correctness of it, such as knowledge transfer, knowledge use and knowledge acquisition that should be considered.

When managing intellectual capital within an organisation, the knowledge contained on a personal level does play a role, but managers and executives cannot focus exclusively on whether employee's beliefs are true and rational. This will cripple production and will not contribute to the efficient working of the organisation. In order to improve the working conditions of a firm, knowledge generation, knowledge transfer and knowledge use become more important. Within a productive environment, whether it is for goods or services, the absolute truth of beliefs are not the most important criterion. It is more important to ensure that the processes are continually improved and that the current

information is used to the advantage of the firm. As previously discussed, ideally knowledge is based on good evidence. Information gathered and collected acts as the evidence for future actions to be taken. Therefore information is also the evidential basis for knowledge (together with experience and insights) and some definitions of knowledge focus very strongly on knowledge being information acted on or information applied. In this way information management is important to ensure that the knowledge managed in an organisation is of good quality. Chapter 3, 4 and 5 will discuss the different strategies that can be applied through information architecture in order to support knowledge management.

The collaborative view of knowledge is not common in the traditional study of the philosophy of knowledge, but in the philosophy of science there has been some investigation into knowledge creation and knowledge use. Such work has especially focused on how scientists create new knowledge through a shared understanding of language, measurement and experiments. It also considers the social aspects of the shared processes. By incorporating the social aspects, more than just the explicit aspects of knowledge are included. The other strength of these methods is that it doesn't focus on the one size fits all solution. Rather specific solutions for specific problems in specific contexts are found and generalizations are avoided (Aarons, 2006: 168 – 169).

Another area of study that contributes to the discussion about knowledge in an organisational setting is social epistemology. As opposed to traditional epistemology, social epistemology considers knowledge to be a shared product of groups. More specifically, social epistemology considers how people interact and generate a "social path" to knowledge, rather than the individual's journey to knowledge. Therefore knowledge is considered collective and not individual property. Social epistemology also considers that knowledge could be distributed amongst individuals and pay attention to the specific factors that influence the distribution within a group. This is not to say that the real properties and process of knowledge are irrelevant, rather that there are additional dimensions to the knowledge process, especially in a group environment. The disadvantage is that this focuses on knowledge production and not knowledge use

(Aarons, 2006: 169; Goldman, 1999:4). Nevertheless, social epistemology is a field that can add value to future developments in knowledge management.

According to the resource based view of the firm, another criterion that a strategic resource should have to create a sustainable competitive advantage, is that it should be imperfectly imitable. Resources, or combinations of resources, might be imperfectly imitable because of the way the relationships between the resources developed over time. As the relationships and interconnections were developed in a specific time and place, other organizations will be unable to replicate the resource that is needed to implement the competitive strategy. Examples could include social networks or organizational cultures that develop uniquely for each organization. Ironically, a resource could also be imperfectly imitable if the source of the competitive advantage is not known, because the unknown cannot be replicated. Another reason for resources to be imperfectly imitable is that they might fall beyond the realm of systematic management and influence. Therefore the strategy cannot be copied, because it is unique (Barney, 1991: 107-108).

Knowledge is imperfectly imitable. It develops over a period of time in a unique way for each individual or group that possesses it, depending on informational and other inputs such as experience. Many argue that knowledge is beyond the realm of strategic management and often outsiders cannot know the exact nature of knowledge, so competitors are unable to perfectly replicate it. One of the examples of how knowledge can be managed to enhance its uniqueness is the concept of communities of practice. Communities of practice is also an example of how the social aspects of knowledge is specifically addressed in a knowledge management framework.

2.3.1 Communities of Practice

As discussed at length, knowledge is not an asset that thrives in isolation. Rather it is the one asset that grows through use and sharing. But it is also not correct to assume that knowledge work includes as many people as possible. Rather, studies have found that knowledge is best communicated through informally defined networks. In order to understand these networks, one also has to have an understanding of the customs and

interactions between members of these networks in context. Some of these networks are known as communities of practice (CoP). (Thomas et al, 2001: 866)

Communities of practice are one of the most popular strategies for managing tacit knowledge that resides within people. A community of practice is more than an unstructured team and it differs from a structured team structure in a number of ways. Firstly, members are assigned to standard teams by the organisation or organisational structure. The formation of a community of practice is however voluntary and based on shared interests or experiences. These differences in formation also affect the authority relationships differently. In a team environment, leadership is assigned, whereas in a community of practice, leadership positions emerge based on experience, expertise and the interaction of the members of the community. Teams are created to achieve certain objectives and deadlines determined by the organisation. Communities of practice are only responsible to their members and the goals they set themselves. Finally, the output processes and formats differ between teams and communities of practices. As with other aspects, team reporting structures are prescribed by the organisation, while communities of practice can develop their own reporting structures and processes (Stork & Hill, 2000, quoted by Lesser & Storck, 2001: 832).

A number of factors interact to contribute to the success of a community of practice. Firstly, shared background and experiences greatly enhance the performance of the communities. It not only contributes to setting up a community but also leads to even more shared experiences and more sharing of knowledge and ideas (Thomas et al, 2001: 867).

The second factor that influences the success of communities of practice is motivation. For many the phrase "knowledge is power" means that what they know is a source of influence and they view it as 'protection' of their position within the organisation. If incentives are incorrectly aligned it can further re-enforce the idea that people have to hoard what they know. For example, if employees are rewarded for their competitive talents they will be more protective of what they know. On the other hand, if employees

are rewarded for how many people use their knowledge outputs and how often they are contacted for information, they will be much more willing to share their knowledge (Olsen and Olsen, 2000: 156).

Thirdly, social ties and the physical space that the network occupies, influence the interactions within networks. Many of these interactions happen by chance and can be very fortuitous, but the relationships and work spaces can be cultivated to encourage more of these seemingly unsystematic meetings (Huxor, 1998 quoted by Thomas et al, 2001: 867). Other research also points to the importance of workspace design in both virtual and physical environments that encourages "weak ties" that further enhance knowledge sharing (Churchill & Bly, 1999: 103).

From the three aspects that improve the workings of communities of practice: shared background, motivation and social ties, one can see that for knowledge to be effectively managed the focus should not be at the individual level, but rather on the social context within which a number of individuals share and develop knowledge. Although philosophers have debated very long what the qualities of our different beliefs should be, when the definition is applied in a functional environment the debate should be about what characteristics allow people to improve the quality and quantity of their beliefs in a social environment. The same granularity can also be applied to how knowledge is viewed. It is no longer the factual, context-free collection of items that describe the world around us: rather knowledge only adds value in context (Thomas et al, 2001: 867).

The benefits of communities of practice are manifold. However tapping into these benefits requires different organisational structures and different value analysis of time spent on knowledge acquisition and sharing. The traditional organisational structure is hierarchical and tends to prescribe team structures and goals. To allow knowledge to be shared people have to be encouraged to form communities, and allowed to develop their own work processes (Wilson, 2002). This will come about when communities of practice are seen not only as being beneficial to the participants, but also to the organisation. The benefit will accrue to the organisation if the agent from the organisation participating in

the community of practice has the ability to identify and acquire knowledge from the community activities that fits strategically with the organisation's goals and objectives (Malone, 2002: 10).

2.3.2 Concluding remarks

A resource is defined as valuable if it contributes to the ability of the organization to initialise and implement strategies that will increase productivity and effectiveness (Barney, 1991: 105). The above discussion highlighted the fact that knowledge is as valuable as traditional factors of production and it can even enable the other factors to become more productive. Because knowledge is also imperfectly imitable and rare it meets the criteria set out by the resource based view of the firm theory to provide a sustainable competitive advantage if managed correctly. The potential value of knowledge should be considered in the social context of knowledge and this affects the management of knowledge. Due to the social nature, technology should be applied in a knowledge management strategy as an enabling technology to assist people to communicate and share knowledge.

2.4 Definitions and understanding of knowledge management

The above discussion already mentioned one knowledge management technique, namely communities of practice. In addition, having established that knowledge is an important resource and therefore needs to be managed, the discussion turns to the importance and meaning of the term, knowledge management. Knowledge management, although it is a contested term for various reasons (as the discussion on criticisms of knowledge management will highlight), is a term that has entered Management, Information Technology, Library Sciences and other literature and is here to stay. That said, there is no definitive definition of knowledge management that would please all parties. The diversity and multiplicity of views is clearly reflected in the following selection of definitions related to knowledge management:

“Processes of capturing, distributing and effectively using knowledge”
(Davenport, 1994)

“Knowledge management is defined as the collection of processes that support the creation, dissemination and utilization of knowledge between appropriate individuals, groups within an organization and independent organizations” (Back, et al, 1999: 11-2)

“Knowledge management is a management practice that uses an organization’s intellectual capital to enable the enterprise to achieve its organizational mission.” (St. Clair, 2003: 1486)

“The need to manage the creation and flow of intangible assets like ideas, innovations, best practices, corporate policies and other aspects of intellectual capital has caused many organizations to invest in information systems (IS) generically referred to as knowledge management (KM) systems” (Brown, et al, 2005: 49)

“[as] an integrated, systematic approach to identifying, managing and sharing all of an enterprise’s information assets, including databases, documents, policies and procedures, as well as previously unarticulated expertise and experience held by individual workers.” (Army Knowledge Online, as quoted by Butler, 2006)

These are wide ranging definitions that include a variety of components of knowledge management. The first short quote by Davenport indicates the attitude prevalent at the time that knowledge should be (and could be) stored. It also indicates, as the other quotes do, that knowledge management is about using and applying knowledge. As the later quotes indicate, knowledge management evolved and came to include more processes and systems to manage intangible aspects of knowledge.

But the lack of definition has led to complications. There are two broad categories of criticism of knowledge management. The first is that knowledge management is a term that is used to describe various information management or information technology strategies. Knowledge management is seen as a temporary buzzword and a way for information technology firms and consultants to repackage information management and information technology solutions. The second criticism of the term knowledge

management is that knowledge cannot be managed and that the term is therefore misleading. Various points of view on both criticisms will be discussed.

Wilson (2002) did a study of academic journals, business consultants and business schools to determine if there was a consensus about the definition knowledge management and what it really referred to. From the titles of academic papers he concluded that the majority of papers that had knowledge management in the title appeared in special issues of the particular journal. In addition, there was no agreement on what the term meant within those papers and often the terminology seemed to be incorporated only to allow publishing in the special issue. Finally, the academic papers that address the issue of what is knowledge management conclude that knowledge cannot be managed and that it rather refers to management of information or management of work practices.

Secondly, Wilson considered how knowledge management was interpreted by management consultants, as they are often the proponents of knowledge management. Here he also found that knowledge management meant different things to different companies, many of whom described information management under the heading of knowledge management. Often the confusion was caused by casually interchanging the terms knowledge and information. Notably, even those that are credited with starting the movement (Sveiby, Drucker and others) acknowledge that the term, knowledge management, is a misnomer. They consider knowledge to be within a person's mind and therefore incapable of being managed by outsiders. Knowledge management was also not one of the preferred management tools used by businesses sampled by Bain and Company – it was ranked number 19 out of 25 management practices. However, even though companies in this survey may have been found not to be using knowledge management, it is reflected as one of the growing trends and considered by the Economist Intelligence Unit 2005 to be one of the top five issues to be addressed by multi-national companies within the next 15 years.

Wilson's third area of interest is the development in business schools. He concludes that the most prominent business schools do not engage with knowledge management. Furthermore, those that do mention knowledge management programmes fall into the same difficulty as seen in the review of academic papers and journals, i.e. confusion between knowledge management and information management.

The three areas investigated by Wilson can be subdivided into two main categories: academic and private sector. Both are unclear about what exactly knowledge management is and appear to regularly interchange the terms knowledge and information, indiscriminately. Furthermore, there is evidence of information technology initiatives being labelled knowledge management. Thus, there is evidence that the term knowledge management has been misused and misapplied. Part of the reason for this is the enthusiasm that follows a new management fad and the resulting misunderstandings of a new practice (St. Clair, 2003: 1488). However, even Wilson admits that the concept of knowledge sharing and empowering people to contribute creatively to the organization would be beneficial to the organization. This was confirmed in the Economist Intelligence Unit Forecast 2020 survey where knowledge management was listed as one of the top 5 management issues to face the corporate world in the next 15 years as judged by the top companies. Knowledge management was also listed as the area of activity that has the most potential to ensure productivity gains in the same period.

Therefore, if the view is taken that knowledge management is "a set of systematic and disciplined actions an organization takes to obtain the greatest value from the knowledge available to it" (Marwick, 2001: 814), the concept of managing knowledge does not seem so abstract. This is an extension of the view of knowledge as an asset and that knowledge needs to be administered to add value to the company. As with other assets, the management of knowledge requires a combination of organizational, social and even technological programs to be effectively executed to ensure that returns are maximized. Even though knowledge resides in people's heads, the organizational environment has to be organized in such a way as to encourage employees and agents to achieve their highest

productivity, which will result in successful achievements of the organizational strategic objectives (St. Clair, 2003: 1486).

Therefore, knowledge management when taken in the broadest sense could be a feasible management practice. However, the meaning of knowledge management has as yet not been unified into one universally accepted definition. There is consensus however that knowledge management involves the organisation and sharing of knowledge between individuals with the aim of increasing the organisation's competitive advantage (St. Clair, 2003: 1486; Beckman, 1999: 6).

2.5 Tacit and explicit knowledge

One of the most recognized distinctions of knowledge in knowledge management literature is the tacit-explicit distinction. The term "tacit knowing" was first used by Polanyi (1967) to highlight the fact that there is knowledge that an individual has, but that they cannot clearly articulate - "we can know more than we can tell". After a detailed investigation into the different ways of knowing, Polanyi (1967: 24) summarizes tacit knowledge as the state of having a valid understanding of the problem and being able to sense the solution while moving towards a possible understanding of the implications of the solution.

These qualities of tacit knowledge are intuitive to most people, but not easy to express clearly and coherently, which also leads to it being difficult to codify and capture electronically.

Nonaka (1991) builds on Polanyi's work in his theory of the knowledge creating company. Nonaka (1991:98) defined explicit knowledge as "formal and systematic", qualities which make it easier to capture in digital format and disseminate to different groups. This is in contrast to tacit knowledge which is very personal, not easy to capture with technology and difficult to communicate to others.

Many authors argue that in order to manage knowledge in an organizational environment the aim should be to have it in an explicit, easy to communicate format. In order to describe how tacit knowledge becomes explicit, Nonaka (1991) developed the “spiral of knowledge”-paradigm for teams or organizations. This model can then also be used to understand how new knowledge is created in an organizational environment. The first step in the knowledge creating cycle is the “socialization” of knowledge through one-on-one tuition. The student assimilates knowledge from the teacher through imitation, observation and practice. The tacit knowledge of the teacher is now also the tacit knowledge of the student.

The second step is for the master (or the student who has grasped the tacit knowledge) to express that tacit knowledge so that it can become explicit. This articulation allows the knowledge to be shared with colleagues and co-workers, rather than each member having to go through the same “socialization” phase. The third step is for all the expressed (i.e. explicit) knowledge in an organization or team to be combined. Different pieces of knowledge are put together in such a way as to form a new whole. Ideally this combination should also take into account that which is still tacit in the group in order for the new knowledge to be as complete as possible. Finally the new knowledge that was created within the team must be internalised by the members and become part of their shared base of understanding (Nonaka, 1991: 99).

The spiral of knowledge paradigm has made knowledge management practitioners think differently about the asset they are managing and started the debate about what is knowledge and how do people share it in the context of organizations and making a profit. However, there has been some criticism of this understanding of tacit knowledge. Nonaka gives the impression that all tacit knowledge can be made explicit through the right translations and conversions. This is contrary to the tacit knowledge Polanyi was referring to. “True” tacit knowledge cannot be expressed and therefore cannot be made explicit. Also, when knowledge is explicit it has the same characteristics as information and therefore knowledge management should not be overly focused on managing explicit knowledge, because it will be managing information.

One of the consequences of Nonaka's spiral of knowledge is the focus on making all knowledge explicit so that it can be codified and managed with technology. This has led to an over emphasis on technology within the knowledge management domain. As discussed in this chapter knowledge is a social process and therefore humans have to play a central role. Technology, however does have a role to play by facilitating the processes and mediums that humans use to communicate knowledge. Topic maps is one example of technology that can assist humans to expand their knowledge and will be discussed in chapter 5. Communities of practice is also the type of environment that would be conducive to the exchange of tacit knowledge. As members interact and develop a shared background they will also share a large amounts of tacit knowledge about their field of interest or expertise.

An additional dimension could be added to build on the work related to the tacit-explicit discussion and classification, namely implicit knowledge. Even when people express what they know there is always some context or assumptions that are not made explicit. This knowledge is not tacit in terms of people not being able to express it, but it is not made explicit either. Implicit knowledge falls between explicit and tacit on the knowledge continuum. Sometimes tacit and implicit knowledge is used interchangeably. They share some of the same characteristics, but are different in certain aspects that is especially important for knowledge management. Certain knowledge management initiatives can focus on making implicit knowledge explicit, as this is an achievable goal and would improve the communication framework within an organization. Another term for this type of implicit knowledge would be "know-how" (Al-Hawamdeh, 2002).

2.6 Conclusion

The resource based view of the firm considers all assets that are used in the execution of strategies at organizations to determine which assets develop opportunities or eliminates threats. In other words, which assets have the potential to provide a strategic competitive advantage. Such a strategic competitive advantage can be sustained through resource

heterogeneity and by ensuring that strategic resources are rare, valuable, imperfectly imitable and without strategic equivalents.

Traditional factors of production include land, labour and capital. Traditionally output is given as a function of labour and capital in the short term, but knowledge plays a more important role than traditionally considered. Knowledge is not constant in the short term and can positively influence output. Knowledge adds value to an organization by improving the potential of the traditional factors of production. Therefore, knowledge can be considered to be the new factor of production and an essential resource in the information age. The difference between knowledge and the traditional factors is that knowledge does not decrease in value through use. To the contrary, knowledge only adds value when it is applied, not when it is stored or gathered. Knowledge resides in people, but information is the resource that supports knowledge as a factor of production. As with other factors of production, the availability and quality of the resource, plus the way that it is managed, influences how strategically relevant and valuable the resource will be.

The value of knowledge in an organisational environment is realised when it is treated as a process that involves interaction between people and their surroundings. Therefore there is a need to move beyond traditional epistemology to social epistemology in order to understand knowledge in a social context. The social nature of knowledge also ensures that each individual and group develops its own unique knowledge. Communities of practice is a initiative that developed to build on the social nature of knowledge and to provide a space where knowledge can be shared in order to realize the full potential of knowledge as a factor of production.

Because the concept of application and use is so important to knowledge, a field called knowledge management has emerged. No definitive definition of knowledge management exists, but a common understanding of knowledge management is emerging. The initial confusion around the term knowledge management did cause damage to the image of knowledge management and has lead to the criticism that knowledge management is in reality information technology or the field of information

systems repackaged. However, because of the potential value that knowledge can add to organisations, there is now agreement that some effort should be made to manage knowledge as an asset. As with other assets, the management of knowledge requires organizational, social and technological programmes.

Within knowledge management literature, knowledge is often distinguished as tacit or explicit. Tacit knowledge is difficult to articulate and therefore difficult to capture on information systems. Explicit knowledge, on the other hand, is easy to capture and store in an information system. Between these two is implicit knowledge that could be articulated, but is not always. Implicit knowledge shares some characteristics with both tacit and explicit knowledge, but is more often associated with tacit knowledge. Tacit, implicit and explicit knowledge exist on a continuum with no clear boundaries between them. However, explicit knowledge has many properties of information because it can be defined and analysed. As a result, although the current debate in knowledge management is on tacit and implicit knowledge, explicit knowledge should still be addressed in order to have a balanced view of what knowledge management entails.

The argument here is different from the initial knowledge management practices that focused on making tacit knowledge explicit and over-emphasizing the role that technology can play in that process. The argument in this dissertation is that explicit knowledge (or information) should be recognized for the role that it plays in supporting knowledge and not as the end goal. The next two chapters will discuss the importance of information retrieval in an information society where information overload is a concern. It will also discuss how information retrieval can be enhanced through the effective application of information architecture.

Chapter 3 – Introduction to information architecture

3.1 Information architecture

Information architecture, like knowledge management, is a contested label for a field that is thriving despite the controversy surrounding the name. Due to the history and development of information architecture (that deliberately included a wide range of disciplines) the field of information architecture is wide-ranging and includes a broad spectrum of topics.

The technological advances of the information age have led to an exponential increase in the number of information sources available to people and organisations. Although heralded as a good thing, people soon began to realize that more information does not necessarily translate into more knowledge, greater productivity or even better decisions.

“There is a popular sense that contemporary culture has an overload of information that is counterproductive at both the individual and society level. The Internet’s burgeoning supply of information has often become the current shorthand for this information overload... In other words, the Web produces a massive surplus of content that may not actually be useable as information” (Burnett & Marshall, 2003: 23).

Even though technology is part of the problem of information overload, technology is also becoming part of the solution. With most information existing in digital format it is only natural that most of the solutions to information overload should be in the digital and specifically “online” domain. However, just as knowledge management initially focused excessively on technology as the key ingredient, so the field of information architecture should also take note of the role of more traditional architectures when developing a comprehensive, organizational framework. Also included in the information architecture framework should be recognition of the contribution that Library and Information Science can make due to their long history in making information as accessible as possible.

3.2 What is information architecture?

In order to answer what a field of study/interest refers to, one should possibly consider what a person operating in that field would do. Richard Saul Wurman is usually credited as being the first to define the term “information architect”. In 1996 he listed the characteristics of an information architect as being:

“(1) the individual who organizes the patterns inherent in data, making the complex clear; (2) a person who creates the structure or map of information which allows others to find their personal paths to knowledge; (3) the emerging 21st century professional occupation addressing the needs of the age focused upon clarity, human understanding and the science of the organization of knowledge” (Wurman, 1996, as quoted by Rosenfeld & Morville, 1998: 10).

This definition paints a very broad picture of what an information architect can be. The definition can be summarized as highlighting two main functions, namely the information architect makes the vast amounts of information more accessible to people through grouping together of like items and by highlighting the ways to find the right information. As more information is available to organizations, generated internally or externally, these two functions of the information architect become vital to allow the organization to manage information effectively. Managing information effectively is important because information is a building block of knowledge.

3.3 History and development of information architecture as a discipline

Information systems architecture is a set of architectures that represent different perspectives of the information and technological needs of the organization at various levels (Zachman, 1987: 291; Segars & Grover, 1996: 381). In the 1980’s information systems architecture developed because of an increase in the scope of design and levels of complexity of information systems as technology improved. When computers were first developed, the logical construction of information systems was not a concern

because the technology available did not allow for much complexity. There was only so much that could be achieved in terms of automating an organization with computers that only had limited processing capabilities. Another part of the motivation for information systems architecture is the rate with which scope and capacity increased, adding an additional dimension to the complexity of designing information systems (Zachman, 1987: 276). The focus in the initial stages was to ensure that the technology was sufficiently employed to answer the strategic needs of the organization. The focus evolved as technology available to organizations changed. At first it was only the development of stand alone applications and later the integration of these applications. The recognized authority in this era was the Zachman framework (Evernden & Evernden, 2003: 95).

3.4 The Zachman Framework

Although RS Wurman coined the phrase “information architect”, JA Zachman developed one of the pioneering works on information systems architecture while at IBM and published it in 1987, with an extended version in 1992. Whereas Wurman’s work is linked closely to Library and Information Science, the Zachman-framework is much more information technology oriented and specifically geared towards information systems in the digital domain. As a starting point the Zachman-framework is useful because it uses the analogy of architecture to explain processes involved in information systems design. One of the motivations for using architecture as an analogy is that it is an independent reference and that lessons can be learned from disciplines that have experience in complex engineering projects. Because of the increase in technological advances, developing an information system starts to resemble a complex engineering project more closely. The aim of information architecture during this phase was to define and control the interfaces and integrate all the components of the system. At this stage in the evolution of technological systems, the driving force behind the development of information architecture was the need to improve the systematisation of the development of the different systems (Zachman, 1987; Evernden & Evernden, 2003: 95).

The key aspects of the Zachman framework is that it consist of more than one view, depending on your perspective or on the description of the system being applied. Although the complete model has a greater number of representations, the three fundamental representations are from the point of view of the owner, the designer and the builder. Translated into information systems architecture, the owner is the business manager or department head who requests the development of the information systems design for their business or department. The designer is the information architect who has been contracted to translate the owner's perspective into a well-planned system. The builder is the technological expert who takes the conceptual ideas and translates them into a reality based on the constraints of nature and available technology. These three players represent the main participants in the development of an information systems architecture. These views are not substitutes, but rather complement each other. Most importantly, the different perspectives do not only represent different levels of detail, they differ in nature and have different roles to fulfil (Zachman, 1987).

Another way of differentiating between the various views is to consider the different constraints each perspective faces. The scope of an information system's architecture is dependent on the financial considerations. While in the enterprise model, the aspects that guide usage and policies within the organisation will determine what is included in the representation. The systems model will be constrained in terms of the structure and operations available to the designer. As the most easily understood example, the technological model, which represents the builder's perspective will be constrained by the available technology. And the "out-of-context" model, which represents the programmer's view will be constrained by what can be implemented on a practical level. These constraints add to each other from the lower levels to the higher levels. For example, an information system architecture that cannot be implemented is not feasible even if there are no financial constraints. Where there is a conflict or incompatibility between levels, alternatives have to be negotiated between those responsible for the different levels (Sowa & Zachman, 1992: 593).

Next, the Zachman framework points out that different views are also possible depending on the description of the information system or product that is viewed. In common language, the information system can be described in terms of what, how and where. The functional (or “how”) description describes the process that the system applies to the data. The focus is on the transformation of data into useful information. The network (or “where”) description reveals the flow of data and information through the organisation. This will not only focus on the technological connections, such as intranets, etc. but will also focus on the communication networks between people and their colleagues. The material (or “what”) description refers to the data that is contained and represented in the information system. This is where information design and structure are considered (Zachman, 1987: 282-283, 292; Sowa & Zachman, 1992: 595).

In the extended framework the people, time and motivation (the who, when and why) abstractions were added to the description of the information systems architecture. For the sake of conciseness the extended framework is not discussed in this dissertation.

The different descriptions of the product as defined by the Zachman framework, the what, where and how, are different abstractions from the same product specifications and therefore are related to each other, but are considered independently for simplification. However, the relationships between these abstractions and the impact that changes that one abstraction will have on the others during the design process are important to avoid inconsistencies in design and ultimately implementation. Another reason to be cognisant of the different abstractions is to avoid optimising the one aspect, i.e. function, over the others, i.e. data or process (Sowa & Zachman, 1992: 595).

Combining the two groupings of views related to the development of an information systems architecture, means combining the participant views with the product descriptions. The result is that each participant will have a different view depending on the description of the product they are looking at. For example, when describing the data aspect of the information system, different views can be taken depending on the perspective of the participant.

The Zachman framework represents a very detailed discussion on the different representations that will result from a detailed development of an information systems architecture. The benefit of including all the participants from top management to the programmers is that when the focus moves from technology to information it can still be included in the framework.

The Zachman Framework was the first of its kind and was developed in a specific time frame. Although the framework is general enough to encompass many other aspects, subsequent developments have led to new focuses and improvements.

3.5 Information architecture shift from technology to information

As discussed in previous chapters, information has become a strategic resource in the information age. However, as the information systems architecture described above has illustrated the focus in this field has traditionally been weighted towards technology rather than the structure and design of information itself. There is a distinction between the resource, information, which provides value through use and good content, and the technology that supports the information (Evernden & Evernden, 2003, 95). Because of the close relationship between information and the technology that supports it, it is sometimes difficult to distinguish between them. For example, innovative information structures, such as hyperlinks and topic maps, are not possible without technology. But technology is the means, while information design and structure is the end. A natural world example would be sky-scrapers that would not be built without modern tools, such as cranes. In information architecture the focus has shifted to the design of sky-scrapers, while the quality and capabilities of cranes continue to improve.

The shift of focus within information architecture also brings the focus closer to knowledge management. As discussed in chapter 3, information and explicit knowledge are very similar. As the focus of information architecture moved away from technology to information design, the focus shifted towards explicit knowledge by association. The

process moved closer to considering how information architecture can contribute to the management of the knowledge continuum, including tacit and implicit knowledge.

As the focal point of information architecture shifted away from being purely technological, towards information design and structure it also moved back to the more traditional Library and Information Science concerns of classification, content management and information retrieval. Due to technology advancements, additional considerations such as navigation, online searching and web development have also been added. Building on the concepts of the previous information architectures, the focus now shifts to individual elements of information design. As will be described in the next chapter, some of the most important aspects of information design are information retrieval and metadata.

Information retrieval can be described as the process through which information is extracted from a system. The alternative is to “push” information to users. Marketers and managers concerned about the existing knowledge of their employees would be concerned with encouraging greater participation from their respective audiences. This is a potentially interesting area of research, but this dissertation will focus on systems and information designs that facilitate the finding of information.

3.6 Information retrieval

With information as a strategic resource it is important that the resource can be found in a timely way, i.e. information that is gathered, should be easily retrieved. This may seem like an elementary statement. However, in practice it is more complicated. Information retrieval is very closely linked to the organization of knowledge. It is easier to find information in a well organized system. Because of the expansion of information resources available, the organization of those resources has become more important and the retrieval of useful information out of the vast amounts available, more difficult. Organization of information resources is vital to information retrieval, because without knowledge about the system used to organize the resources, retrieval in a well-organized

collection will also not be possible. Libraries have well-established methods of organizing and labelling information and methods for information retrieval (Rowley, 1992:3-5). These can be tailored to the general information architecture field. As the earlier quote by Wurman (1996) indicated one of the main tasks of an information architect is to organize data and this is what Librarians have been practicing from the outset.

When a system is developed with the user and the user's experience in mind, the system can be evaluated based on user outcomes (see chapter 3.6 on user experience and usability). The effectiveness of an information retrieval system can also be evaluated based on the quality, time and cost associated with that system. Even though each of the components should be optimised, there will ultimately be a trade off between the criteria. The quality of a retrieval system depends not only on the content of the database (or other information storage system) that supports it, but also on the ability of the system to retrieve relevant documents based on a subject request from the database. The ability to retrieve relevant documents is known as recall and the ability to identify non-relevant items and not to present them to the user is known as precision. These will be discussed in detail in chapter 4.2. The final aspect that ensures the quality of the effectiveness of a retrieval system is the flexibility to provide the information in a form or format that is easy to use and most convenient for the user. Both the cost and time associated with using an information system is influenced by its ease of use and the effort that is needed to obtain the relevant information. To maintain the competitive advantage that information as a resource provides, the effectiveness of the information system should be optimised in such a way that the investment of time and money is still exceeded by the benefit gained from the information retrieved (Lancaster, 1986: 131-132).

When designing an information retrieval system the following considerations need to be taken into account. Firstly, the level of searching expertise the typical audience is likely to have and what type of information needs users will have. Other considerations are the type and quantity of information being searched (Rosenfeld & Morville, 1998:105). End users information seeking behaviour will differ depending on their specific need and

situation. Designers of information retrieval systems therefore need to take the information seeking behaviour of end users into account when designing an information retrieval system. Furthermore, the time and effort that indexers invest in setting up the linkages and cross references can be weighed off against the time and effort end users have to put in to get the desired results. For example, control can be exercised at the input stage (indexing or labelling) or at the output stage (searching or browsing). In order to exercise control at the input stage, while still having an effective retrieval system, the indexer has to put in more time and effort. Likewise, if the control is at the output stage, then end users have to simulate the controls of the indexer through effective querying of the system (Rowley, 1988: 60-63; Lancaster, 1986: 8; Warner, 2004:181; Garshol, 2004).

The next chapter will consider information seeking behaviour in detail. In the next chapter the tools available to indexers and designers of information architectures to improve information retrieval will be discussed.

3.7 Information seeking behaviour

Information retrieval can be improved through correct classification, but consideration should also be given to the methods end-users will employ to find information. Information seeking behaviour is a complex topic and much has been written about it. Searching and browsing are two methods of information seeking. Although the terms are often used as synonyms for each other they are different in many technical aspects. This chapter will explore these differences. Another facet of the information seeking process is the display of information in such a way that humans can effectively process it and respond in the appropriate manner (Jacso, 2003: 17).

Rowley (1992: 484-495) defines searching as “the process of looking for something, such as information or a document”, and to browse as the activity of investigating the content of a collection without a clearly defined strategy. The implied difference is that when searching the ultimate goal is known but browsing is more exploratory. This is further reinforced by the definition of a ‘search term’, by the same author, as the words or

phrases a user uses to find the documents that have been indexed under that term. In order to do an effective search the user has to have some background information about the indexing terms used. If the user has that information he/she is aware of the relationships between the index terms as presented in a thesaurus, he/she can use this to find additional relevant information.

In the situations where user's needs are ill defined or where they are unable to express it clearly in a query, the user can employ a browsing strategy. For example, when people are engaging in the process of acquiring new knowledge (such as when doing an exploratory search) and are unfamiliar with the subject matter, they are unable to use the terminology or to phrase the query in a succinct and exact manner. In such a situation, browsing might be more effective as it allows the user to find links to more information sources without being limited to what they already know. In a traditional library environment a user might use the shelf number of a familiar item to find the general area in the shelves where they can look for more information. Once in the shelves they will page through different books and scan book titles for items they consider relevant to their information need. However in an online environment this is difficult to simulate, but a number of ways to encourage "serendipitous" knowledge discovery have been developed. Browsing in the online and digital environment is accomplished through the navigation systems employed in a website or software (Warner, 2004:178; Rosenfeld & Morville, 1998:47).

Websites are informational spaces. Finding information within a digital domain is quite different from finding one's location in a physical space. Online environments do not have the same signposts to guide users. However, becoming lost in an online environment is as frustrating as getting lost while driving. Becoming lost or not being able to find their position within a website also prevents users from finding the information they require. As a supporting facet to the overall goal of assisting users to find the right information navigation is a powerful tool. Well-designed hierarchical structures decrease the chances of users, especially new users, becoming lost. A navigational scheme that complements the organizational structure will provide context

and give the user more flexibility of movement within the website (Rosenfeld & Morville, 1998:47).

Appropriate navigation systems can also support knowledge management initiatives as the navigational system supports associative learning by highlighting material related or relevant to current content. Examples include side bars that have “Related topic” links displayed. The challenge for website design and information architecture is to balance the flexibility and informational value of navigational links with the risk that the user will be overwhelmed and confused by the multitude of options (Rosenfeld & Morville, 1998: 47-48).

Search functions are included in web sites for a number of reasons. Firstly, browsing is more time consuming than searching (although it can also be more rewarding) and users do not always want to spend the time browsing through web content and links to find what they are looking for. Users also sometimes use the search function in a website, even if they do not know exactly how to phrase their queries. Providing a search function assists them by making the option available, even if it is not the most appropriate tool. A search function is also very important where there is the risk that users will experience information overload, such as the case may be with very large and complex sites. The other very important reason to include a search function is if the website contains dynamic content that would result in frequent updates to the navigation tools that support browsing. In this case, search functions are cheaper and easier to maintain. A search function should not be included in a small website that does not really need one in order to compensate for bad browsing or inadequate navigational systems (Rosenfeld & Morville, 1998:96-101).

Users turn to search options for different reasons. Understanding the way in which users search and the purpose of their search will enable the search system to be designed to effectively meet the user’s requirement. Different user information requirements have different types of solutions. For example, a “known item search” is well defined and consists of a single, correct answer. Two alternative information requirements arise when

a user has heard of a concept or idea. They either want to establish that it exists (existence searching) or find out more about an idea or concept (exploratory searching or browsing). For existence searching to be successful, the source or agents need to understand the user's question, which may be very vague if they are very unfamiliar with the topic. In exploratory searching the user is more prepared to take the time to learn about a new concept, and a few pieces of information would be sufficient for their needs. Comprehensive or research searching, on the other hand, requires a high recall rate. It is important to note that users will often go through a number of iterations while refining their search terms. People also tend to use more than one source for a particular search and combinations of search and browsing techniques (Rosenfeld & Morville, 1998:101-104). Furthermore, individuals can also use each other as information sources in their quest for knowledge.

An electronic search system should support these different modes of searching. In addition, because users might shift between searching and browsing modes, the two systems should be closely integrated. It is also important to display the search results in a user friendly way. The type of information and amount of information displayed about the search results will depend on user requests and on the degree of structure that the content has. The more results have been retrieved the less information about each item should be displayed. The order of the retrieved documents is another factor in the display of results. Although results can be displayed alphabetically or in reverse chronological order, many search options display search results based on relevance. However, relevance is sometimes a problematic issue. Different relevance algorithms calculate relevance statistics differently. That is why it is important to tell users how relevance is determined, while ideally they should be given the choice in the order of the results (Rosenfeld & Morville, 1998:105-122).

Searching and browsing are two methods employed by end users to find information in any collection. Although similar, the two methods each have their own strengths and weaknesses and are used in different ways. In the online environment, navigational tools assist users to find the information, through either searching or browsing. Navigational

systems have three roles to fulfil. Firstly, a navigational system provides context to prevent the user from getting lost. Secondly, dynamic navigational systems improve flexibility of movement through the website. These two roles combine to fulfil the third role of navigational systems, which is to assist the user in finding the correct information. While attempting to fulfil these roles of the navigation system, designers need to consider the following: the specific goals of the website, who the potential audience will be and what content will be included in the website. If uncertainty exists, usability tests could be included in the website design in order to facilitate learning more about user patterns and preferences (Rosenfeld & Morville: 1998: 70).

3.8 *User experience and usability*

The other aspect of effective information retrieval is user experience and usability of information retrieval technologies and methods. This is where information architecture has a role to play. The aim of good information architecture is that users will be able to find the information they need to expand their knowledge and make economically viable decisions. It is necessary to differentiate between information requirements and retrieval techniques for business and personal purposes, for example online shopping for gifts and business-to-business online purchasing.

Jared Spool et al (1999) did research to establish what components of web sites helped and which hindered the finding of information by first time users. Although this research does not consider the sales and marketing aspects or familiarity that users can develop through repeat visits to a web site, it does give insight into web site design aspects that affect information retrieval. Information retrieval can assist the discovery and sharing of knowledge and is the focus of this research project. Furthermore, although the research conducted by Spool et al (1999) was exploratory, the principles can be applied in intranets and other information architecture designs.

The research is based on observing first time users trying to find specific answers to specific questions from specific sites. The questions ranged from fact finding to

comparing judgments. The first finding of the research is that graphic design does not help, but it also does not hurt a web site.

For information retrieval text links are vital. Links are important to navigate through a web site and navigation and content are inseparable. When the navigation and content are planned simultaneously, the links are more likely to provide clues to where they will lead and this improves the information retrieval capacity of users. Information retrieval is a very different activity to surfing the web. The result is that websites need to be designed differently if they are aimed at surfing or at information retrieval (Spool, et al, 1999: 3-14).

The other interesting comments from the above mentioned study is the theory of 'skimming'. This means that users do not read all the text but only look for aspects that seem plausible and then read the text that surrounds it. This is very different from (the assumptions around) traditional print media. This also has implications that relate to the notion that people are not gathering all the data and information available, but that they are focusing their attention on the bare necessities to speed up the process. The study also found that web sites with less white space did better than those with more white space. Again, they theorized that it is because people are hunting for information and that "white space spreads out the information and slows down the hunt" (Spool et al, 1999: 76).

3.9 Conclusion

The information age has coincided with exceptional increases in available information resources, but more information does not automatically translate into more knowledge, improved productivity or better decisions. Technology has made the expansion of available information resources possible and is also assisting in creating solutions to information overload. However, new technology is not the only solution, traditional Library and Information Science techniques also have a role to play.

An information architect is a person who is concerned with making vast amounts of information more accessible. This is achieved through classification and highlighting ways to find information. Managing information in this way is important because information influences knowledge creation.

Information systems architecture developed to deal with the increased complexity and scope of design that improvements in information technology provided. Initially the focus was on automating existing practices, but recently the focus has been shifting towards the structure and design of information. The Zachman Framework is one of the pioneering works in information systems architecture. He used the analogy of construction to illustrate the importance of project management in information architecture and that different role players will require different representations of the same project. The Zachman Framework is technology oriented.

It is often difficult to distinguish information design from the technology that supports it, but there is a shift to focus on information design rather than the supporting technology. By moving the focus from technology to information the focus is also moving towards explicit knowledge. Therefore the fields of knowledge management and information architecture are getting closer while still pursuing their own goals.

Information retrieval is a key task of an information architect because it involves organizing information so that others can find their own path to knowledge. Information retrieval assists in making information (the strategic resource) more accessible. An information retrieval system should be evaluated based on quality of results and the time and cost associated with implementing and using the system. There is a trade off between the time and cost spent at the input stage (indexing and labelling) or at the output stage (searching and browsing).

User requirements and expertise will also influence the development of an information retrieval system. The two most common information seeking behaviours is searching (for known item requests) and browsing (for knowledge expansion). Navigational tools such

as search functions, help users find information in an online environment. The navigational system should provide context to users and assist them with moving through the website. This ultimately helps them to find the right information on a website. In this way navigation systems can also fulfil the goals of information architecture.

This chapter considered how information architecture evolved and what the aims are of information architecture. Information retrieval is one of the primary objectives of information architecture. For this reason the chapter also considered how users will interact with an information retrieval system. The next chapter will consider the tools available to the creators of an information retrieval system to fulfil the goals of an information architect as defined by Wurman.

University of Cape Town

Chapter 4 – Information retrieval and information architecture

4.1 Introduction

Chapter 3 introduced the field of information architecture and highlighted some of the features of user behaviour that needs to be considered when designing an information architecture. This chapter will focus on the various features that are available to information architects, especially in the field of information retrieval.

4.2 Recall vs. Precision

Recall is defined as the proportion of relevant documents retrieved to the total number of relevant documents in a collection for a given search query (Warner, 2004: 185; Rowley, 1988: 55; Lancaster, 1986: 132). When a user searches an information system, they are looking for information that will be relevant to their specific needs. Although all the items recalled might relate to a specific topic, not all will relate to the user's requirements. There are a number of reasons why items retrieved would not be useful to the reader: it may be in a language they do not understand, have the wrong level of detail for their requirements or it may be written at the wrong academic level (Foskett, 1996: 15). Precision measures the extent to which the documents recalled accurately match the needs of the user. Due to the different objectives of precision and recall there is a natural tension between them. The more items that are recalled the less likely it is that all of them will be relevant to the users needs. Nevertheless, both concepts remain critical, as it is important that relevant information is not overlooked due to poor recall, but that the user's time is not spent on irrelevant material due to poor precision.

Applying the recall ratio in a large database is problematic. Determining the total number of relevant items for any given search would take very long, even in a static database. In a web environment where the information and its storage is dynamic, the problems become

even more complex. Therefore, recall ratio is a difficult measure to apply in a web situation and its use is debatable. Despite this, the concept of recall as the ability of the information system to recall relevant items remains important. This is especially so where information overload is a common concern and having the correct information at the right time is vital. The main technique to improve the recall is the correct classification of material and the matching of search terms to the classification labels of the material. Although all items that are recalled should in some way be relevant to the search terms, not all material will have the same relevancy to the specific intended use of the material. Recalling documents with high relevancy will increase the precision of the search, but decrease the recall impact (Warner, 2004: 185-186).

On the other hand, in the web environment precision has become a very important measure. As ever-increasing amounts of information have become available, the challenge of presenting the user with information that is relevant to their particular need is becoming greater. One method of increasing precision of search results is through classification schemes that accurately reflect the contents of the collection that is being searched (Warner, 2004: 185-186).

4.3 Classification and classification schedules

Classification plays two crucial roles in information retrieval. Firstly, classification is the arrangement and grouping based on likeness, i.e. all the items in a group share at least one characteristic that items in other groups do not share. The second role of classification is to show the relationships between concepts or classes of concepts. Often the second function of classification, which is very important for knowledge discovery, is overlooked and excessive emphasis is placed on grouping of items. The correct classification of items will lead to improved recall during a search for relevant information. Classification also improves the precision, for example by refining the search based on the hierarchical structure of the classification, which improves the likelihood of finding the right information. There are a number of different standardized classification schemes that have been developed that highlight some of the norms in

classification and provide guidance on setting up new classification schemes (Buchanan, 1979: 9; Rowley, 1992: 485).

According to Rowley a classification scheme in the context of a library is an “arrangement of library materials in a systematic sequence, according to their subject and, to a lesser extent, their form” (Rowley, 1992: 485). She further states that a classification schedule is the “main published subject listing of a classification scheme” (Rowley, 1992: 485). In terms of organizing information on a web site Rosenfeld and Morville (1998: 26) define organizational schemes as representing the shared characteristics of content items and this is influenced by the logical grouping of those items. The definition of organizational schemes and classification schemes are very similar and the terms will be used interchangeably. The different types of relationships between the content items and groups are defined in an organization structure. Rosenfeld & Morville (1998: 26) highlight the importance of organizing information in web site development and that organizational structure also influences navigation, labelling and indexing. (See chapters on navigation (3.7) and subject indexing (4.5) for more details.)

Organization schemes or classification schemes can have different forms. For example, content can be arranged alphabetically, chronologically or geographically. These are examples of exact organization schemes. However, not all schemes can be arranged in that way. For instance, although listing library materials by author or title will result in an exact organization scheme, listing material by subject will result in an ambiguous organizational scheme. Despite guidelines to direct classification of material within a subject categorization, it remains an approach that is open to personal bias from the indexer or prejudicial influence based on personal experience (Rosenfeld & Morville, 1998: 27-29).

Due to the nature of ambiguous classification schemes there are certain challenges. For example, information is expressed in different forms of language. Natural language can have many ambiguities, which mean that words can be interpreted in more than one way. A number of vocabularies and other methods have been developed to deal with the

multiple interpretations possible. Often context also provides guidance on how to interpret certain terms or phrases. (See the discussion on topic maps in Chapter 5 for one approach to providing context.) Another obstacle to classification is the many forms that information can take and that it can relate to any subject. This means that most collections of information are not uniform, but heterogeneous. The heterogeneity does not only refer to the content, but also to the formats available on the Internet and the different levels of granularity. This heterogeneity means that the “one-size fits all” approach or highly structured organization schemes do not work well on the Internet or in organizational collections that can include documents as diverse as agendas, reports, news articles and staff evaluations. Classification can also be guided by the creator’s perspective on a topic and on what is important and/or relevant. In an open structure like the Internet there are multiple perspectives and interpretations of language. All the factors mentioned above contribute to the challenge of arriving at a cohesive classification (Rosenfeld & Morville, 1998:22-26). Despite these challenges, classification schemes enable users to find information.

4.4 Metadata

Metadata is intrinsically related to and dependent on classification. Metadata records (or information about information) generally record information such as the author, publication date and subjects about an information resource. The author is an example of an exact organization scheme, while the subjects will be in an ambiguous classification scheme. The focus of the metadata discussion will be the role it plays in the refinement of information design and structure through identifying information resources and providing easier access to the correct resources. In this way metadata works as an information retrieval tool (Evernden & Evernden, 2003, 95; Garshol, 2004).

Metadata is important for information retrieval because it captures and stores information about information resources that enables the document, file, picture, etc to be found within a collection. Metadata can also be applied for content management (Garshol,

2004) (However the discussion about content management is beyond the scope of this dissertation.)

The theory behind metadata as a tool for information retrieval is to record the type of information that would assist the end user in finding the specific information resource (Hanisch, 2005: 1929). There are two general types of searches conducted by end users. The first is where they have used or read a document or other information resource previously and want to retrieve it. The second is where they are looking for information about a certain topic. This type of search will be discussed in chapter 4.5 – Subject indexing.

With the first type of search users might have some of the information relating to the information resource like a title or author and it is much easier to identify the correct material. This type of search may be well supported through an exact organizational scheme that records specific information. One example of such an organizational scheme is the Dublin Core, which defines vocabulary to be used in metadata records (Garshol, 2004).

4.4.1 Dublin Core

The Dublin Core Element Set was developed in 1995 as a standard for the organisation of electronic documents. The volume and diversity of electronic documents have made describing them in a uniform way a challenge and the Dublin Core has its own challenges. The attributes of electronic documents are called elements and the Dublin Core specifies these elements. Elements are divided into those that describe content items and those that describe information that relate to “rights” of the resource, i.e. author and publisher. Elements can be qualified further or refined by specifying format or language used. In an attempt to allow for the diverse range of electronic documents available and to allow for flexibility, the Dublin Core does not specify formats to be used explicitly. Best practices are recommended, but no definitive standard exists. The advantage of the flexibility is that each application can specify its own vocabulary for its specific use. The disadvantage of the flexibility is that interoperability between applications is limited. This

supersedes the benefit and objective of a common standard. Furthermore the Dublin Core does not specify which of the elements are compulsory, all elements are optional. Certain applications, for example the Dublin Core Library application, specify the selection and semantics for that application (Hanisch, 2005: 1928).

There are two different schools of thought on how to improve the Dublin Core. Firstly, there is an argument for extending the element set by allowing for "a more detailed description of electronic documents and their media specific properties" (Hanisch, 2005: 1929). The disadvantage of the extended approach is that it could increase the cost of compiling the metadata record and thereby also increase the cost of developing the applications. The second suggestion for improvement to the Dublin Core is to limit the set of "core" elements to describe the basic properties of electronic documents. For example, by only describing each document in terms of who, what, where and when. These elements would also be extendable or qualified if necessary. In order to improve interoperability between applications certain elements could be made compulsory while maintaining the flexibility in specifying the other elements. The subset of elements would then be used to exchange information between applications (Kunze, 2001: 4; Hanisch, 2005: 1929). The main shortcoming of the Dublin Core is the abundance of options available which reduces the interoperability of different applications. Similar problems are being experienced in topic map development, but different solutions are being suggested in that arena. (See chapter 5 on topic maps.) The other shortcoming is that the Dublin Core rarely describes the content of electronic material adequately because it only lists a few keywords. The Dublin Core is better suited to exact specification such as titles, authors, etc.

4.4.2 Concluding remarks

Another and more common scenario in information retrieval is that users are looking for information relating to specific topics. This means that they are looking for documents and other records that would enrich their knowledge about a specific field or area of interest. Such enrichment can take many forms. For example, the user may be doing an exploratory search to expand on an interesting subject they heard about or they might be

interested in an in-depth examination of the subject matter. However, a standard metadata record only provides administrative information and very little information in relation to what the document is about. A title might contain some indication, but often titles are chosen to attract a potential reader's attention and not to convey information about the content. Most metadata systems contain a "subject" field; however the field is often populated with only a few keywords. The problem with keywords is that they are often assigned by the author, so spellings or form may differ. In addition, it is often very difficult to capture all the relevant topics covered by a document in only a few keywords. In metadata records subject keywords are often the most useful field to identify information related to the content of a document or information resource. Yet, as a tool for information retrieval these keywords are inefficient, especially if the few text fields have no restrictions or guidelines (Garshol, 2004).

4.5 Subject indexing

In order to address the problems associated with the ambiguity inherent in classifying documents by subject, various indexing languages and subject-based classification schemes have been developed. Examples of these subject-based classification schemes include controlled vocabularies, taxonomies, thesauri and faceted classification. Some authors (Rosenfeld & Morville, 1998: 29) relate subject indexing as it is practiced in Library and Information Sciences to ambiguous organizational schemes, because of the ambiguity and subjectivity that is involved in subject classification in libraries. Another term for subject indexing is concept indexing.

There are two steps in subject indexing. The first step is to understand the record's subject matter together with knowledge about the needs of the users. This is known as doing a conceptual analysis. The second step involves translating the understanding developed through the conceptual analysis into a particular vocabulary (Lancaster, 1986: 3). Some authors (Rowley, 1992: 165-167) explicitly split the subject indexing process into three steps. The first step requires the indexer to familiarize themselves with the content of the document. In the three step process, the second step is the analysis of the

concepts represented in the document. This step is made easier through competent familiarization of the document. The final step is the translation of concepts into the indexing language or vocabulary that is applicable to the specific classification scheme. The vocabulary also has to be linked to the vocabulary a user would feel comfortable using. There are two main types of vocabulary that an indexer can use. The first is a controlled vocabulary and the other is a natural language vocabulary.

4.5.1. Controlled vocabularies

Controlled vocabularies used in indexing represent a limited set of terms that should be used to represent the subject matter of a record in the database (Lancaster, 1986: 3; Garshol, 2004). The aim of controlled vocabularies is to prevent indexers from using obsolete or meaningless terms and to encourage conformity in terms of spelling and word form. However, as other authors have pointed out (Fischer, 2004: 9) the relationships between terms are not defined in controlled vocabularies and should therefore be self-evident. This may not always be the case and is why specializations such as taxonomies, thesauri and ontologies exist.

Controlled vocabularies were developed in the traditional library profession in order to assist indexers to control for synonyms and homonyms. Control was also introduced into vocabularies due to the variety that exists in natural language. When an indexer recognizes and understands the concepts that a document or record deals with s/he is able to use not only the terms used in the document but to provide synonyms for these terms in order to improve the probability of the document being recalled in the appropriate search. The specification of synonyms in the vocabulary also ensures that the classification terms include terms which are used more often and therefore increase possibility of correct recall of documents. Another objective of a controlled vocabulary is to specify the word form that will be most useful, i.e. 'heat' rather than 'hot'. Homonyms are controlled for by providing alternative terms and by not classifying a document under the incorrect homonym (Buchanan, 1979: 13; Rowley, 1988: 52, 57).

Controlling for synonyms and homonyms as described above was done with the intention of assisting the finding of information. The different variations of controlled vocabulary described below relate to searching and browsing in different ways. Browsing is supported through simple lists of controlled terms that are used as labels for the navigation system or through taxonomies that are also used to create hierarchical navigation systems. Synonym rings and thesauri are applied in search engines. However, the users, to inform their search terms, can also use terms from thesauri. The different strengths and weaknesses of the different controlled vocabularies need to be taken into consideration when the navigation to support browsing or the search engine is developed in order to optimise the use of the correct controlled vocabulary (Warner, 2004:182).

The other element to consider when designing the vocabulary is the size of the collection. A collection with 5,000 items in a specific subject area would need the same number of descriptive terms to fully cover the subject area as a collection with 50,000 items in the same subject area. However, the aim of a search is not to produce an exhaustive list of all available material but to produce useful sets of items. With useful sets of items as the goal, the larger collection would require a larger number of useful sets of terms than the smaller collection. In practice, the smaller collection will have fewer terms of a more general nature, while the larger collection will have more terms which are also more specific. A technique for making a vocabulary more specific is to increase the number of syntactically complex phrases. By combining words in more complex topics the search becomes more refined. The disadvantage of a very specific vocabulary is that it very quickly becomes very large. On the other hand, a vocabulary consisting of very general terms runs the risk that the terms will combine in unpredictable ways. Developing the right level of specificity within the vocabulary will ensure that the items retrieved maximize precision (Warner, 2004: 186-187).

4.5.1.1 Taxonomies

A taxonomy is a controlled vocabulary where the terms have been arranged in a hierarchy. This takes the concept of controlled vocabulary one step further by grouping related terms together. Taxonomies are often applied to hierarchical navigational systems

and inform their composition. This provides some context to the indexing terms and the user's position within the website, depending on where the taxonomy is applied. However, taxonomies are still very limited as they generally do not describe the nature of the relationships between terms and users have to use the exact term to find the correct information resources (Garshol, 2004; Warner, 2004:180).

4.5.1.2 Thesauri

A thesaurus is another example of a specialized controlled vocabulary. Rowley (1988: 57) defines a thesaurus as "an organized list of terms from a specialized vocabulary, which has been arranged to facilitate the selection of index terms". In more basic language, thesauri expand on the capabilities of taxonomies by providing additional statements to be made about the terms in a hierarchy. Therefore, thesauri provide a greater terminology to describe terms and subjects and are more user friendly because of it (Garshol, 2004).

The thesaurus is not only useful to the indexer, but it can also be useful to the searcher. Persons who are familiar with the structure of the thesaurus can use it to broaden or narrow their search within a particular database. By using more specific or more general terms or even just alternative terms, a searcher will be able to retrieve more documents that are relevant to their requirement. In order to know which terms the indexer used, the user will also be interested in the thesaurus. This is possibly the most important advantage of a controlled vocabulary: that it brings "the language of the indexer and searchers into coincidence" (Lancaster, 1986: 8).

Synonym rings group together terms that can be considered equivalent as search terms. If a user enters one of the terms in the synonym ring all terms in the cluster will be used to retrieve information. The difference between thesauri and synonym rings is that a synonym ring treats all the terms as equals and does not indicate a preference. Furthermore, synonym rings are used in collections that are not tagged. Thus, control is exercised at the input stage (tagging or labelling) with thesauri and at the output stage

(searching) with synonym rings (Rowley, 1988: 60-63; Lancaster, 1986: 8; Warner, 2004:181; Garshol, 2004).

4.5.1.3 Ontology and faceted classification

Faceted classification describes documents by choosing terms from different axes, for example personality, time and space. Faceted classification can be viewed as a very disciplined way of compiling a thesaurus. A variation of faceted classification generalizes each facet until it becomes a general property (Garshol, 2004).

Similarly, ontologies, as used in computer science, create a model to describe the real world defining different types, properties and relationships between types. The main difference between the different classification schemes described above and ontologies is that an ontology has an open vocabulary while the others have closed vocabularies (Garshol, 2004).

4.5.2. Natural language indexing

The opposite of controlled vocabulary is natural language indexing. The term 'natural language' refers to the language used in every day discourse by the authorities within their fields of expertise. A natural language system uses the words and terms as they appear in the text to develop an index. The big difference between controlled vocabularies and natural language indexing is that the subject matter used in the latter is represented in open-ended vocabularies. Because of the infiniteness of an open-ended vocabulary a controlled vocabulary will always be a subset of a natural language index. Although the term free text is sometimes used as a synonym for natural language, free text implicitly refers to using terms and phrases as they appear in the text which is more limiting than the terms used in natural language (Lancaster, 1986: 159; Rowley, 1988: 90; Rowley, 1992: 240).

The advantages of a natural language system are that individual interpretation and capturing errors are eliminated and in a computer-based system new terminology can be

added, as they become known. This is a very important feature in a world where new technological advances and available information increase continually. The new technologies and increased information result in new terms and phrases being developed continually. On the one hand, a natural language index might represent more closely the terminology that the searcher is familiar with. While on the other hand, one of the disadvantages of not controlling a vocabulary is that a user might not be able to find information or documents that are available or they might find many irrelevant items due to the lack of control in a natural language vocabulary (Buchanan, 1979: 13; Rowley, 1992: 240,272-273).

There are various criticisms of natural language indexing. Before the introduction of technology, lists of natural language indexes were rarely produced. Applying natural language indexing manually is very time consuming and the list would need to be updated every time new material is added to the collection. Computer based systems can print an updated version at any time, but this does not mean that different computer based systems will have exactly the same lists. Different versions will occur if different records are used as inputs, even if they represent the same documents (Rowley, 1988: 90). For example, abstract indexing and full text indexing will produce different results, even when referring to the same material.

One method of overcoming the disadvantages of a natural language indexing system is to modify the search strategy employed. A controlled vocabulary index requires greater input and incurs more cost at the time of compilation, while the natural language index shifts the costs and effort to the time when the information is retrieved from the database. To overcome potential recall and precision problems associated with homonyms, potential homonyms should be used in conjunction with other terms that will provide a better context to the search query. When dealing with synonyms the searcher needs to identify possible alternative words that would improve search results. Essentially the searcher needs to recreate the same results as a controlled vocabulary would achieve, through creative combination and use of search terms (Lancaster, 1986: 161-163).

Natural language indexes are more useful when the search term is more specific, while controlled vocabularies are more useful when the search is more general.

An elementary example of natural language indexing is using titles for subject indexing. The premise is that the terms in a title will accurately reflect the subject content of a document. Key terms from the title are identified and then the document is classified under each of those terms. The major attraction of this method of indexing is that it is quick and cost effective to produce. However, the disadvantages are that the title may not accurately reflect the content of the document and that the title only highlights the main themes of a document and sub themes will not be indexed with this method (Rowley, 1992: 151, 153-154).

A more complicated situation for applying natural language indexing is in abstracts or full-text documents. The first difference between title indexing and abstract or full text indexing is that the number of potential index terms increase dramatically and is usually linked to the length of the source. Secondly, the greater number of terms also means that it is more likely that terms are included as index terms that do not accurately represent the content of the document. In order to reduce the likelihood of irrelevant terms a stop list can be created. A stop list includes all the terms that are not to be considered keywords, for example also, than, the, etc. In conjunction, a go-list can also be created. A go-list includes all the terms that would create useful index terms for that specific subject area. The go-list can be based on a thesaurus. Although updating of go-lists is not automatic, indexers can periodically request all terms from the database that are neither on the stop list or go-list. This will provide a new set of potential index terms. This will give the indexer some control as to which terms are included in the go-list, but the allocation of index terms and the forms allowed for concepts are not controlled. If a natural language system only uses a stop list the indexing language is truly open while a system that includes a go-list is more controlled (Rowley, 1992: 271-272; Rowley, 1988: 100-101).

Usually words in abstracts and full text are used to search electronic databases or in an online environment. The development of hypersearch means that indexes do not have to

be created as the computer performs a very fast search of the complete database when the request is submitted (Rowley, 1988: 99-100). Due to the nature of classification, it has long been hoped that at some point it would become completely automated and no longer require human intellectual input. Thus far, this has not happened but many innovative methods have been created that do increase the automatic component and reduce human involvement in repetitive tasks. The great advantages of computers are that they can count, compare and match terms, which does go some way towards grouping information automatically. One comment is that even automated systems make use of thesauri, which is a controlled and structured vocabulary, created through human intellect (Buchanan, 1979: 123-124).

4.6 Conclusion

Recall is the measure of the number of documents retrieved for a given search request. Precision measures the relevance of the retrieved documents based on user requirements. Although both measures are difficult to calculate numerically, both concepts have relevance in determining the quality of results generated by an information retrieval system.

Classification groups items together based on common characteristics and indicates the relationships between items or groups of items. Classification can be based on exact organization, i.e. alphabetically or geographically, or on ambiguous organization, i.e. by subject or topic. Ambiguous classification is subject to different interpretations and differences in word form. One of the main concerns with classification is to provide context to terms or information.

Metadata is a collection of classification schemes, recording information about the information resource. The concern with metadata records is that it is difficult to specify fields that should be included for all information resources, especially for heterogeneous collections such as the World Wide Web. The Dublin Core is an example of a metadata standard developed as a classification standard for electronic documents. However,

Dublin Core metadata records are difficult to merge. This is because conformity on a number of fields is required to successfully merge two Dublin Core records.

Subject-based indexing is a subset of metadata and focuses on classifying information resources based on content. Two types of vocabularies are used by indexers for subject-based classification: controlled vocabularies and natural language indexing. Controlled vocabularies (such as taxonomies and thesauri) use a predetermined list of terms to index information resources. This has the advantage of controlling for synonyms and homonyms and word form. Natural language indexes use the words and terms as they appear in the text for indexing purposes. The advantage of this method is that it is easier to automate and can accommodate new terms as they are taken into use in the literature.

The above discussion highlighted a number of shortcomings in traditional subject classification schemes. Firstly, most schemes have a closed vocabulary. Although a number of positive aspects were described above, it is still considered a shortcoming because it limits the terms that can be used to retrieve information resources. Knowledge can be expressed with an almost limitless vocabulary and by using closed vocabularies we are unable to express the full range of ambiguities of knowledge. Secondly, a limited vocabulary is often unable to describe how the different terms relate to each other. This results in a lack of context and this makes it more difficult for users to modify their search to find additional useful information. It also means that traditional classification schemes are unable to reflect the intricate relationships between concepts, as they are understood by the human mind. Thirdly, such classification schemes also have limited ability to give additional information about the terms, i.e. describing the properties of the concept being classified. Therefore, the shortcomings relate to three areas: naming of concepts and subjects, clarifying the relationships between the concepts and describing the attributes of the information resource that relates to the concept. Alternatively, the shortcomings can be summarized as the inability of traditional classification schemes to accurately reflect the rich context and associations people identify between different components of their knowledge. Topic maps, as discussed in the next chapter, aim to address the shortcomings and better reflect the complexities of human knowledge.

Chapter 5 – Topic maps

5.1 Introduction

The discussion thus far has defined knowledge management and information architecture and the issues involved in both fields. In order to achieve the objectives of this dissertation a conceptual model still needs to be developed that clearly explicates the relationship between knowledge management and information architecture. However, this is a difficult task due to the complexity of the relationship. As part of the investigation into the relationship, it was realized that topic maps could be used as a method to clearly explicate the relationship between knowledge management and information architecture.

Topic maps are a relatively new method of leveraging subject classification methods with the additional storage capacity provided by technology. The additional storage does not only relate to the size of a classification scheme but also the complexity of it. Traditionally subject classification schemes have been two-dimensional because they were paper based. Technology allows the classification scheme to become three-dimensional and in effect become an information resource in itself. This means that the topic map as subject-based classification scheme evolves into a metadata record. It conveys information about the information resources; it can be used to classify material and is a valuable information retrieval tool.

These facets of topic maps are also highlighted in the following quotes:

"A topic map is essentially an index to the content of some collection of information resources, usually but not necessarily in digital form." (Bater, 2004: 133)

"With topic maps you can create an index of information which resides outside that information... The topic map describes the information in the

documents ...and databases ... by linking them using URIs³ (Garshol, 2002)

Although topic maps can be used to present, manage and search complex collections of heterogeneous information resources the concepts underlying a topic map is very simple. A topic map consists of topics with names, associations between topics and the occurrences of the topics. Each of the features of the topic map will be discussed separately.

This chapter will deliberately aim to explain the concept of topic maps without becoming too technical. The aim is to introduce the philosophy behind topic maps and its application for this research which is to clearly explicate and map features of knowledge management and information architecture and to improve understanding of the underlying principles rather than discuss the relative merits of different operational languages. That said, it does seem that XML⁴ provides the best platform for the development of topic maps. The other advantage of XML is that it is very compatible with web based applications and other database storage applications.

5.1 Topic maps in general

Topic maps originated from the attempts to merge traditional indexes and to represent the knowledge present in those indexes (Pepper, 2002). After the initial development it was realized that a more general application could be created with much greater possibilities. Topic maps also aimed to address the related information management problems such as creating, maintaining and processing indexes for heterogeneous collections. Topic maps further evolved to include navigational aids (Pepper, 2002; Ahmed & Moore, 2005).

Topic maps can be created through three methods. The method that provides the highest quality is those created by humans, as people can process the ambiguities and more clearly represent the associations between topics. However, this is very labour intensive

³ Unique Resource Identifier: See chapter 5.4

⁴ Extensible Markup Language

and can become very expensive and unpractical if applied to large projects. Another method is to automatically generate topic maps using available software from existing source data that is ideally but not necessarily in XML. As with other automated classification systems this only works well if the information being mapped is well structured. The third method is to automatically create a topic map with source data that is already in XML or another specialized application (Garshol, 2002).

Subject-based classification systems aim to classify information resources by their content – what they are about. Topics are concepts or subjects that are contained in the material being indexed. Because indexed material can be about anything, anything can be a topic. The ISO 13250 standard is specifically vague about what constitutes a topic because the options are limitless and therefore beyond definition. Each subject has three characteristics: a name, an association to other topics and information resources that expand on the knowledge about that topic. These characteristics are also the features included in a topic map (Garshol, 2004).

5.3 Topic names and types

Each topic has a name. The name can take any form (i.e. nickname, formal name, login name, etc) and is usually explicit but not always. An example of a non-explicit name is a cross reference such as “see also page 97”, because it is considered to link to a topic, even if it has no name (Pepper, 2002). Just as indexers can use any of a variety of terms to denote a particular subject, so too can those who create names for topics allocate any name in topic maps. Topic map structure allows a topic to have more than one name. The same topic might have a different name in different contexts, such as different languages or geographical locations. These differences can be specified in the scope notes (see the discussion on scope notes in chapter 5.6). Variants of names can also be provided for specific processing contexts. Traditional classification schemes avoid duplicate names for subjects as it can be confusing when out of context. However, the types of names, associations and occurrences act to lessen the chances of ambiguity surrounding topic names within a topic map. The scope facility allows the creator of the topic map to define

their own language for describing topic names. An example of how scope notes can be applied to different topic names that refer to the same subject is if different regional offices refer to the same subject by different names. Traditional classification schemes would indicate this with a “see also” cross reference but the scope facility allows a topic map to show different views to different users depending on their preference, while also indicating which regional offices use which terms (Pepper, 2002; Garshol, 2004).

It naturally follows that no two subjects can have the same name and scope notes, as this would indicate that they constitute the same topic (Pepper, 2002; Garshol, 2004).

The advantage of topics as they are contained in topic maps is that they can be grouped together by type. In controlled vocabularies or other traditional classification schemes different kinds of terms are grouped together, i.e. people, places and organizations, without the mechanisms to tell them apart. By assigning topics to different classes they can be searched and different characteristics can be defined for them (Garshol, 2004). For example, a traditional classification scheme might group “Idi Amin” and “Uganda” together as related terms without indicating how they relate to each other or that the one is a person and the other is a country. In a topic map “Idi Amin” would be grouped with persons and “Uganda” would be grouped with countries. The connection between the two terms would be clarified in the association.

5.4 Associations

The second feature of topic maps is the associations between topics. This is similar to the relationships presented in thesauri and taxonomies. However, the associations in topic maps are more dynamic and provide richer context for the topics. On the issue of context: Rosenfeld & Morville (1998: 58) state “In designing navigational systems for the Web, context is king.” Context is a very important component to information management, but is usually difficult to display in paper-based systems or for computers to correctly identify automatically. For example, thesauri do express relationships but only to the extent of broader/narrower, etc. There is very little semantic content in those types of

relationship indicators. In contrast, topic maps aim to make associations as specific as possible. Associations are also not limited in terms of the number of topics that can be represented. Three or more topics can occur with the associations between them described (Pepper, 2002; Garshol, 2004, Garshol, 2002).

As with topic names, associations can also be grouped together according to type. This means that a subset of topics can be grouped together based on similarities in relationships to other topics (Pepper, 2002; Garshol, 2004). For example if an author's name is the main topic, selecting other topics that have the association "written by" to the author's name will result in a list of documents and books written by that author.

The associations between topics are independent of the information resources underlying the topics. Therefore, the hyperlinks are between topics, which is different from other, "normal", cross-references where the hyperlink occurs within the information resource. In the traditional system, the cross-reference is only effective if the information resource is operational. Because the association hyperlinks in topic maps are independent of the information resource, the topic map itself becomes an information resource. The independence of topics and information resources also means that topic maps can be applied to different collections and still be relevant. As the associations between topics remain the same it would only be the occurrences for that collection that would have to be updated. It is also true that different topic maps can be applied to the same collection or information set to provide different "views" for different users (Pepper, 2002).

In addition to recording the relationships between topics, topic maps also records the role each topic plays in the relationship. Associations between topics are inherently multidirectional. Association types do not indicate direction and that is why association roles are very important. This means that the relationship, the direction of the relationship and the role of each concept relative to the other are captured (Pepper, 2002). This is another illustration of the depth of metadata that can potentially be captured in topic maps.

5.5 Occurrences

The third feature of topic maps is occurrences. Occurrences refer to the resources that relate to the specific topic and where information about the topic can be found. The resource can take any form (i.e. written document, picture, sound, etc) in which it will have relevance to the specified topic. In this way topic maps move in the opposite direction of metadata schemes such as the Dublin Core that links an object to the subjects that relates to it. A topic map links subjects or topic to objects that are about the topic. The information resources usually remain outside the topic map and are only pointed too. Another major difference between traditional indexing languages and topics maps is that the information resource does not have to be present in the collection being indexed for the topic to be presented in the topic map. The topic can be mapped with no occurrences if that is the case. Therefore, there is a separation between topics and their occurrences (Pepper, 2002; Garshol, 2004).

This has the potential to be very important for knowledge management. When used as a knowledge mapping tool, the topic map can indicate areas of knowledge even if that knowledge is not recorded or currently present in the organization. Therefore, all areas of a subject field can be indexed even if physical resources (including electronic versions) are not available at that point in time.

As with topic names, each occurrence can be typed (classified) depending on the nature of the occurrence. Occurrence types also allow distinctions to be made between the various types of relationships there are between different information resources. Examples would include bibliographic references and influences one occurrence might have had on another. Material can be further distinguished by using the scope feature on occurrences (see chapter 5.6). Both methods facilitate searching and managing of information resources that reside in a heterogeneous collection. The other advantage of typing occurrence categories is that it allows the creator of the topic map to define the language for describing occurrences. As was seen with the Dublin Core it is often difficult to specify elements that should be defined for all electronic resources. Topic maps in effect provide for an open vocabulary for the properties of occurrences, a feature

few traditional techniques allow. However, the open vocabulary only provides real benefits when the occurrences can be typed. Specific properties can be added for specific types, i.e. date of birth applies to people but not to cities or countries (Pepper, 2002; Garshol, 2004).

Most occurrences have Uniform Resource Identifiers (URIs) that identifies the resource being linked to the topic. This means that any resource, for any location can be linked to the topic. Some occurrences can be stored as strings inside the topic map, for example date or birth or phone numbers. In such instances the occurrence will not have a URI (Garshol, 2004). (See further discussion on URI's in chapter 5.7)

5.6 Scope

The three features of a topic in a topic map: the name, occurrences and associations are collectively known as "topic characteristics". Topic characteristics are always made in a specific context. The context may be implied or explicit. However, the topic characteristics are only valid within that context. The limits of validity of topic characteristics within a topic map are defined in the scope. The scope consists of a set of topics, members of which are called themes, which represent the context within which the topic characteristics are valid. The ISO standard on topic maps 13250 (International Standards Organization ISO/IEC 13250 1999:3-4) standard does not require that a scope be specified explicitly. If no scope is specified, the topic characteristic is considered to have unlimited validity (Pepper, 2002; Garshol, 2004).

By defining the different validities of different topic names, ambiguity in homogeneous terms is removed. This is not only true for words that are homogeneous; it is also true for topics that have the same name but different contexts. For example, place names that could also refer to food or be a title of a book. Furthermore, scope can also aid navigation in topic maps if it is used to specify the view depending on the user's profile. Scope can also be used to determine which of the topic names for a specific subject should be

displayed. Scope acts as a filtering mechanism based on the properties of the topic (Pepper, 2002).

5.7 Subject identity

Although the goal of topic maps is to represent each subject with one topic, this is not always the case. When topic maps are merged the same subject may have different allocated topic names, for example if topic maps from different languages are merged. In order to enable a user to retrieve all relevant material about a subject through one topic name, topics should be identified that relate to the same subject. One way to do this is through the concept of subject identity. Some subjects have URIs that can be used as its subject identity. However, this is only valid for subjects that are addressable (Pepper, 2002, Garshol, 2004).

Abstract subjects cannot be addressed directly, so subject indicators must be assigned. In ISO 13250 subject indicators are called “subject descriptors” and defined as “information intended to provide a positive, unambiguous indication of the identity of a subject” (International Standards Organization ISO/IEC 13250, 1999:4). The subject indicator will be given an address and become the subject identifier of the topic. Topics described by a common subject indicator will be automatically merged when topic maps are merged (International Standards Organization ISO/IEC 13250, 1999:3). The resulting topic map will have one node that includes the combined characteristics (names, occurrences and associations) of the topics that were merged (Pepper, 2002).

Naturally, subject indicators should be uniform in order to facilitate interoperability between topic maps. “Public subject indicators” are defined in the ISO 13250 (1999:3) standard as “a common referent of identity attributes of many topic links in many topic maps.” In practice the public subject indicators are published and maintained at advertised addresses so that it is available for use by all creators of topic maps (Pepper, 2002).

The challenges in this regard are similar to those faced by the administrators of the Dublin Core. The difference is that topic map technology allows topics to have different names, while the subject indicator enables the different maps to be merged. This is different from the collections described by the Dublin Core metadata scheme where conformity is needed on a number of elements in order for it to be effectively merged. Collections classified using topic maps only need conformity on subject indicators in order to be transferable or merged with other topic maps. The difference in “changeability” between topic maps and the Dublin Core is also partly related to the independence of topic maps and the collection or occurrences being mapped. Because the elements in the Dublin Core are so closely intertwined with the occurrences this makes merging or transferring Dublin Core metadata more complicated than merging or transferring topic maps, even for the same collections.

5.8 How topic maps promote information retrieval

The main components of topic maps (topics, associations and occurrences) in combination provide a far better overall view of a topic than traditional subject-based classification schemes. Firstly, topic maps present information in networks that more closely resemble reality than hierarchical systems used in thesauri and other two dimensional classification schemes. Secondly, the relationships are defined explicitly instead of being given in generic form. These two features combine to give a tool that facilitates navigation between topics and more efficient information retrieval, therefore acting as a more effective knowledge discovery tool (Garshol, 2004).

Topic maps can be used to support full text searching. Traditional full text searching technologies recall all documents that mention the search term. As an alternative, a topic map system can return topics that match the search term and include the relevant additional information (Garshol, 2004). This allows the user to browse through “a multidimensional topic space of knowledge” (Pepper, 2002) instead of scanning the recalled documents to find the most relevant information. In this way the user can refine the search strategy to achieve greater precision. As discussed earlier, the information in

the topic map, such as relationships and the different roles that topics play in the relationship will already provide users with information that could aid their understanding of the subject area. In effect, topic maps allow users to have the best of both worlds. Topic maps provide extensive lists of related terms and the use of differing terms to describe the same topic, such as would be produced through natural language indexing. However, topic maps also provide context so there is more structure and results are likely to have a higher precision than natural language indexing.

Using the query language of topic maps can also assist the user in refining their search and thereby speeding up the information retrieval process. The only predicament is that the user would need knowledge about the query language and the ontology of the topic map if they want to phrase their query correctly. A solution would be to design the user interface to include dropdown menus and other features that would assist the user to define their query. Natural language querying is still in the experimental phase for topic maps, but it has great potential (Garshol, 2004).

As discussed earlier, one of the main concerns about topic maps are that no official standard exists for the language used to define the structures used on the data set. A Topic Map Constraint Language (TMCL) is being developed as a solution to this problem. The aim of the constraint language would be to use the capabilities provided by the typing feature to create rules that can be recognized by software applications. Examples would include: "only persons and organizations can have telephone numbers" (Garshol, 2004).

Another way that topic maps support information retrieval is to act as a content management tool. Because more information is available to describe content more variables can be used to group similar items (Garshol, 2002).

5.9 Topic maps as an information architecture tool

The role of an information architect is to enable users to find relevant information. Topic maps provide a platform that information architects can use to enable users to find information (Garsahol, 2004). Topic maps were developed for a digital environment and can be used to build websites as well as manage electronic document collections. The technology can also be applied to print media and is especially useful in heterogeneous collections that contain print, electronic and archived material of various sources and uses.

When topic maps are used to construct websites they are used to provide the structure of the websites. The content of the website can be taken partly from the content of the topic map and partly from the occurrences. This structure can be applied to portals, catalogues and site indexes (Garshol, 2002).

Topic maps also improve information architecture by allowing heterogeneous collections (in terms of sources, type of information and even content) to be effectively mapped. Interoperability between topic maps, across information sets or providing different views of one information set, allows information as the resource to be supported as opposed to supporting the specific collection of information resources. Topic maps can also be merged, which is difficult to do with traditional indexes or metadata schemes (Garshol, 2002).

5.10 Topic maps as enabling technology for knowledge management

“Topic maps are a new ISO standard for describing knowledge structures and associating them with information resources. As such they constitute an enabling technology for knowledge management” (Pepper, 2002)

Knowledge management can be seen as the optimisation of knowledge organization within an enterprise, as discussed in chapter 2. Knowledge organization is very similar to

the goals of information architecture, i.e. optimising the organization and conceptual access structure of a given organization's knowledge repositories in order to support retrieval and creation and sharing of knowledge. When optimisation is achieved, all knowledge assets and knowledge flows are known, used and improved depending on their strategic position and potential for adding value within that organization.

"The ability to encode arbitrarily complex knowledge structures and link them to information assets indicates a major role for topic maps in the realm of knowledge management: Topic maps can be used to represent the interrelation of roles, products, procedures, etc. that constitutes corporate memory, and links them to the corresponding documentation." (Pepper, 2002)

This quote by Pepper highlights the importance of topic maps as an enabling technology for knowledge management and as an information retrieval tool. One of the challenges of knowledge management is the diversity of forms of knowledge and information resources that exist within an organization. Traditional classification and information retrieval schemes were restricted to two-dimensional representation and were therefore insufficient to represent the complexity of knowledge available to an organization. An additional feature of topic maps is that it has the ability to map knowledge and general concepts even if it is not currently residing in the organization. This means that knowledge gaps can be identified, while still representing the context and linkages to other areas of expertise.

"In fact, with a richly descriptive classification system like a topic map the information in the classification system becomes a valuable resource in its own right...This can effectively turn the navigational structure into a knowledge management application, while allowing it to be used for classifying content." (Garshol, 2004)

One of the greatest strengths of topic maps is that it can become a source of information and learning. The navigational feature allows a user to explore the specific area of knowledge and how the different concepts relate to each other before studying the detail of the area of knowledge. This helps the user to avoid information overload, by giving

them the freedom to explore different levels of detail as they become more familiar with the subject matter. It also guides the user through knowledge discovery.

5.11 Why topic maps are not more widely applied

Despite the various potential advantages mentioned in relation to topic maps, the application of topic maps is not as wide spread as expected. The first possible reason for the lag in interest is that it takes some time for new technologies and the accompanying innovation to be shared across different fields. A further factor, that is related to the first, could be that the standard specifications for topic maps is very technical and not intended to be understood by non-technical parties such as librarians, knowledge officers or cataloguers and has been difficult to apply in non-technical environments. Another hindrance is the different terminologies used by technical and non-technical practitioners of knowledge management, information architecture and designers of topic maps. Finally, the lack of tutorial material that is widely available (and affordable) means that few people are investing the time to understand the new technology and how it can be applied. The topic map “movement” would also benefit from the review of different real-world applications in order to make the necessary improvements (Siegel, 2000).

Another hindrance to the implementation of topic maps is that they are difficult to represent two-dimensionally. The result is that their potential can only be shown in a digital environment. Although the theory is relatively simple, potential creators of topic maps will have some difficulty in conveying their plans to non-technical people. See further discussion in chapter 6.

5.12 Conclusion

Metadata was described in chapter 4.4 as a tool that records information about objects (such as books or electronic documents) in order to facilitate document management and information retrieval. One aspect of metadata is that it describes the subject matter

covered by the information resource. Subject-based classification schemes are a specialization which focuses specifically on description of content in order to facilitate searches related to specific subjects, rather than specific documents. Topic maps are able to fulfil the function of both metadata and subject-based classification. As the name suggests, topic maps revolve around topics or subjects. Since objects, such as books and documents, can be treated as topics in a topic map the properties of the object are also recorded within the topic map. Topic maps use available technology to address the shortcomings of traditional two-dimensional classification schemes. Topic maps originally evolved out of an attempt to merge traditional indexes and to present the information available in those indexes.

Topic maps assign names to each topic, but there is no limitation of what can constitute a topic. Topics can also be assigned a 'type', which allows it to be grouped with other terms of the same type. Associations represent the relationships between topics and include the association roles. The roles indicate the direction of the relationship and provide more context. Associations can also be grouped together based on type. Occurrences refer to the instances where a reference to a topic or the topic itself can be found. Most occurrences are identified through Unique Resource Identifiers. Topic names and associations are independent of occurrences, which gives topic maps great power to be used on different collections and for different topic maps to be merged. The scope feature on a topic maps allows the indexer to define the context within which the topic name and characteristics will be valid.

Another way of describing the difference between traditional metadata applications and topic maps is that topic maps provide a means of managing the meaning embedded in the information resource, rather than just the information resource (Garshol, 2002). Topic maps provide a better overall view of a topic because they are not limited to the hierarchical structure of thesauri and associations can provide an explicit indication of the relationship between topics. This gives the user an overview of related information resources and thereby better enables them to find the information they are looking for.

Topic maps is a platform that can be used by information architects to enable users to find information. Topic maps can be used to construct websites and to map heterogeneous information collections. Topic maps also act as an enabling technology for knowledge management because it provides the ability to describe knowledge structures and to link it to information resources. Topic maps have the ability to capture complex relationships between knowledge flows and to represent it in some detail.

It is unlikely that topic maps will solve all classification problems. The first obstacle is that it is a relatively new technology and expertise might be scarcer than traditional classification skills. Most of the literature on topic maps is situated in the computer science field and only a few references are made in knowledge management websites. Very little information about topic maps is available in resources that deal with “traditional classification specialists”. In addition, the application of topic maps requires some programming or XML skills, which might also not be within the ambit of knowledge workers. The second concern regarding topic maps are that the additional structure created by topic maps requires greater effort to construct. This could have cost implications. The true measure of the potential cost implication can only be assessed once it can be compared to the benefit created through the additional capabilities (Garshol, 2004).

Chapter 6 – Discussion and conclusion

6.1 Introduction

As discussed in chapter 1, while reviewing the existing literature on knowledge management it was noticed that there is ambiguity surrounding the role that information technology and specifically information architecture can play in the promotion and functioning of knowledge management initiatives. Despite this ambiguity very little has been written about the specific role that information architecture could play to promote knowledge management. In the course of the investigation into the benefits that information architecture may have for knowledge management, the relatively new phenomena of topic maps was found to have special significance. Although topic maps hold promise for both fields, most of the academic literature on topic maps is in the field of computer science. One of the contributions of this dissertation is to make topic maps and the surrounding discussion more accessible to practitioners in the traditional information sciences and knowledge management fields.

6.2 Summary of main conclusions derived from the literature

6.2.1 Information age and knowledge

Information is becoming more important in the information age (cf. chapter 1.1). The information age refers to social, political and economic changes that have resulted because of the increased role of information and knowledge in society. One of the most important changes is the value that information contributes to economic activity, both in terms of inputs and outputs (cf. chapter 1.2).

Information and knowledge are often represented in a hierarchy with knowledge at the top and data at the bottom. To arrive at the concepts higher up in the knowledge hierarchy require greater human involvement. For this reason, some authors argue that knowledge is a uniquely human ability (cf. chapter 1.3). In epistemological terms beliefs

should ideally be true, justified and rational. Importantly, beliefs are based on evidence. And by analogy, information is seen to be the evidence that informs knowledge (cf. chapter 1.4).

Although the information age has brought about certain changes, organisations still aim to gain strategic advantage through effective use of the resources available to them. The resource-based view of the firm is a management theory that considers all the assets that are used in the organisation to meet their strategic objectives. In this theoretical framework, assets are defined as resources when they enable the organisation to either develop opportunities or eliminate threats. In this way the resource also assists the organisation to create a competitive advantage, which should be sustained. The competitive advantage can be sustained through at least having resource heterogeneity and through having resources that are valuable, rare, imperfectly imitable and without strategic equivalents. Taking these requirements into consideration, the researcher evaluated whether knowledge and information have the potential to provide a strategic competitive advantage to an organisation (cf. chapter 2.1).

6.2.2 Strategic importance of knowledge

The first requirement for resources in the resource based view of the firm is that it should have value. The researcher compared knowledge to traditional factors of production to determine if knowledge had the characteristics to add value. Factors of production could be viewed as representing the main classes of resources. The assumption is that because traditional factors of production add value, knowledge will also add value if it compares favourably with the traditional factors of production. Traditional, neo-classical economic theory identifies three factors of production: Land, labour and capital. In this model economic output in the short term is given as a function of labour and capital. The usual assumption is that knowledge and land or natural resources are unchanging in the short term, i.e. unchanging or constant without major investments of time or money. However, knowledge is playing an ever-increasing role in the modern economy and can have a more immediate impact on economic output. Therefore, knowledge can be considered a new factor of production and an essential, valuable resource. Another advantage of

knowledge is that the law of diminishing marginal return does not apply to knowledge. Knowledge, like other factors of production, is influenced by the availability and quality of the resource (information) supporting it. The methods used to manage the factor of production also determine how effectively the resources will be applied to the strategic plans of the organisation (cf. chapter 2.2).

6.2.3 Knowledge management

When knowledge is considered as a resource it is not an object, but a process that enables knowledge to add value to an organisation. The process involves interaction between people and their environment. Therefore, there is a need to move beyond the traditional definition of knowledge from an epistemological point of view (that focused on the individual) to social epistemology (that focuses on group dynamics) in order to understand the social process of knowledge in an organisation. It is also this social nature of knowledge that causes it to be unique in each situation and thus enables it to potentially provide a sustainable, competitive advantage according to the resource-based view of the firm (cf. chapter 2.3). A very effective method that has been developed to build on the social aspects of knowledge is that of the communities of practice approach. Encouraging interaction between individuals and the encouragement of the sharing of knowledge are seen as a method of supporting knowledge management (cf. chapter 2.3.1).

In chapter 2.4 the debate regarding the validity of knowledge management as a management practice is discussed. Due to the lack of a generalised and universally accepted definition of knowledge management, certain information technology programmes were incorrectly labelled as knowledge management initiatives. This caused confusion about the focus of knowledge management and some authors questioned the feasibility of attempting to manage knowledge. Currently, however, it would appear that a common understanding of knowledge management is emerging. The consensus is moving towards knowledge management as an organisational tool that attempts to develop knowledge and information as resources, with the goal of attaining a strategic advantage from them. This researcher, concluded that knowledge management could be

defined as a programme consisting of social, organisational and technological tools that aim to enable employees and agents to share their knowledge to increase productivity and for the benefit of the organisation.

In the initial development stages of knowledge management, emphasis was placed on the tacit-explicit distinction of knowledge. Tacit knowledge was defined as difficult to articulate and capture electronically, while explicit knowledge was defined as easy to articulate and capture. The discussion of tacit and explicit knowledge was heavily influenced by Nonaka's concept of a knowledge spiral, where tacit knowledge can become explicit if taken through a number of steps. This created the false impression that all tacit knowledge can be made explicit and that this transformation should be the focus of knowledge management initiatives. The researcher contends that knowledge, which is truly tacit, cannot be made explicit, because it is "more than we can tell" (Polanyi, 1967: 24). An added dimension is provided to this two-fold distinction by the introduction of the term, implicit knowledge. This term refers to both categories since it encapsulates some characteristics of tacit knowledge, but also includes those aspects of knowledge which can be expressed if needed. The result is a knowledge continuum from tacit knowledge to explicit, with implicit somewhere in between (cf. chapter 2.5).

Because of the intrinsically human nature of tacit and implicit knowledge they are best managed through social structures, like communities of practice. The explicit side of the continuum is very close to information because explicit knowledge shares many of the characteristics of information. Therefore, the explicit side of the continuum will benefit from information management techniques, which can help to improve the whole knowledge continuum. Explicit knowledge and information provide the evidence that supports tacit and implicit knowledge.

6.2.4 Information architecture

The information age led to an exceptional increase in the information resources available. However, more information (or evidence) does not necessarily lead to more knowledge or better decisions. Part of the problem is information overload. The considerable

expansion in technology has led to the increase in information resources and this in turn has resulted in information overload. Technology, on the other hand, is also assisting in alleviating the problem by creating tools with which to manage information and to improve information retrieval. This however should not lead to the conclusion that technology is the only solution to the problem. Traditional Library and Information Science techniques can also be applied (cf. chapter 3.1).

The traditional concerns of Library and Information Science are reflected in the definition of an information architect. An information architect was defined as a person that focuses on making information more accessible through classification and by providing guidelines to indicate how to find information. Therefore, an information architect can be defined as a person who is concerned with alleviating the problems created by an information overload and who develops the process to improve the information resources that support knowledge and knowledge management (cf. chapter 3.2). Initially the focus of information architecture was on using technology to automate existing practices. The Zachman framework (cf. chapter 3.4) is a pioneering work in information systems architecture. It is however very technology oriented and recently, there has been a shift in focus in information architecture to the design and structure of information, away from the technology. This shift is difficult to describe because it is often difficult to distinguish the difference between information design and the technology that supports it. Currently the view is that technology provides the means in information design to ensure that the end goal, effective retrieval, is achieved. The shift in focus has brought the intentions of knowledge management and information architecture closer together. Improved information design leads to better access to explicit knowledge and therefore improves the ability of other knowledge management initiatives to build on explicit knowledge (cf. chapter 3.5).

Information design and information structure are especially important in the field of information retrieval. Designing information retrieval systems is a fundamental aspect of information architecture. It is a way to make information available to users and to guide them to the relevant information resources. It is also the way through which information

as a strategic resource is made more accessible. The evaluation of an effective information retrieval system is based on the quality of the results retrieved as well as the time and cost associated with implementing and using the system. The time and cost aspects can be divided between indexers and users. Certain systems will require greater effort at the input stage (labelling or indexing the information resources), while others will require greater effort at the output stage (querying or browsing the available resources). Typically systems that require greater effort at the input stage will require more of the indexer, while users will be required to do more work with systems that require more at the output stage (cf. chapter 3.6).

The output stage of information retrieval and information seeking behaviour are important factors that should be considered when designing an information architecture. User requirements, experience with information retrieval systems and familiarity with subject matter are all crucial parameters that will influence the development of the information retrieval component of an information architecture. User interaction with the information architecture often occur at the output stage through information seeking behaviour such as searching and browsing of the information resources, which the information architecture supports. Searching and browsing, although similar in meaning, are not synonymous. Searching is defined as the process used to identify the location of a specific item, i.e. the ultimate goal is known. Browsing, on the other hand, involves a less defined process of identifying material that has the potential to answer the information need of the user. In an online environment navigational tools are used to assist users to search or browse for information. The navigational tool should provide context and assist users to move between information resources or web pages. Navigational systems also help to fulfil the goals of information architecture, by ultimately helping users to find the necessary information (cf. chapter 3.7 and 3.8).

Recall and precision are two methods used to evaluate the results retrieved by information retrieval systems. The recall ratio is the number of items retrieved as a percentage of the total number of applicable items in the collection. Precision measures the relevance of items retrieved based on the requirements of the user. These measures

are traditionally used in library environments but the concepts can be applied to website and intranet searches (cf. chapter 4.1).

The grouping of similar things (classification) is one of the techniques used to improve information retrieval. Beyond grouping of like items, classification schemes also indicate the relationships between items. Classification schemes can be either exact (i.e. for geographical or alphabetical classification) or ambiguous (i.e. for subject classification). In both schemes context needs to be provided to the index terms, but this is often difficult to achieve in paper-based or two-dimensional schemes (cf. chapter 4.3). Metadata is information about information and consists of schema that record details about information resources. One of the concerns with metadata is to determine which fields should be included in all metadata records in order to facilitate the merging of different metadata schemes. In heterogeneous collections not all fields will be applicable to all resources. The Dublin Core is an example of an attempt to define a metadata scheme for electronic documents, but which experienced some difficulties in creating metadata fields that would be applicable to all resources (cf. chapter 4.4.1).

Ambiguous classification or subject-based classification is an example of a classification scheme that could be included in a metadata system. Two types of vocabularies are used to classify information resources based on content: controlled vocabularies and natural language indexing. Controlled vocabularies refer to a predetermined list of approved terms that are used to classify content and index information resources. Natural language indexing uses the terms as they appear in the text to classify the document. The advantage of a controlled vocabulary is that synonyms and homonyms are controlled for and that duplication is minimised. The disadvantage is that there might be a time delay in the acceptance of new terms into the controlled vocabulary, so information resources might not be indexed using the most recent terminology. Furthermore, the controlled vocabulary might not use the terms that the user is most familiar with. The advantage of natural language indexing (which is usually an automated process), is that it can include the latest terms and include terms exactly as they are used in the literature, which might be closer to what the user is familiar with. However, some information resources might

use different spellings, word form or other small difference that might not be picked up by the automated information retrieval system, which then excludes resources that do not have the exact search term (cf. chapter 4.5).

6.2.5 Topic maps

Another shortcoming of traditional subject-based classification is that it provides a limited capacity or vocabulary to describe the relationships between concepts. Although this aspect of indicating relationships is one of the main features of classification schemes, it has traditionally been a difficult aspect to depict explicitly. The researcher, therefore, suggests that these shortcomings of traditional classification, i.e. naming of concepts, depicting relationships between concepts and describing the attributes of the information resources that relate to the concepts, can be addressed by using topic maps and its available technology. Topic maps consist of topic names, association and occurrences that address the three shortcomings mentioned above. Topic maps developed out of attempts to merge traditional indexes and therefore they address many of the issues that occurred when attempts were made to merge metadata in the Dublin Core format.

Firstly, topic maps assign names to each topic. There is no restriction on what a name can be, because anything can be a topic. The topic name can be qualified using the scope feature in order to provide context or to indicate under what circumstances the topic name is valid. Topic names can also be allocated to a certain “type”, as defined by the creator of the topic maps, which allows it to be grouped with other topics of the same type or to enforce certain criteria for the relationship between different types (cf. chapter 5.3 and 5.6). Association is the term used in topic maps to refer to the relationship between different topics. Associations are defined as explicitly as possible by the creator of the topic map and association roles help to indicate what the function of each topic is in the relationship and what the direction of the relationship is. This means that the relationships between topics are indicated more clearly and this enables the user to better understand how different topics relate to each other. Associations also place topics in context (cf. chapter 5.4). Occurrences refer to the instances where a reference is made to a topic or where the topic can be found. Topic names and associations are mapped independently of

the occurrences and this means that the detail of the topic names and associations are not limited to available resources but can depict an area of knowledge rather than a collection of material. The other advantage is that one topic map can be applied to multiple collections while only the new occurrence indicators have to be changed. Occurrences are usually identified with URI's. This also makes the merging of topic maps easier as it is the only feature that has to be merged, which is different from the Dublin Core system where multiple fields are required for a successful merger (cf. chapter 5.5 and 5.7).

Topic maps promote information retrieval because it gives a better overview of an area of knowledge. It is not limited to the hierarchical structure of traditional index languages and associations give a more detailed description of the relationship between topics. This allows the user to obtain an overview of related information and information resources and this enables them to broaden or narrow or redirect their search based on their improved understanding of the specific field of knowledge (cf. chapter 5.8). Topic maps can be considered an information architecture tool because it assists users to find information and it can be used to construct and populate websites. Topic maps are especially relevant because they have the capability to map heterogeneous collections that have previously been difficult to map (cf. chapter 5.9). By supporting the objectives of information retrieval, which is an objective of information architecture, topic maps support the retrieval of information from knowledge repositories and thereby supports the creation and sharing of knowledge. More directly, topic maps also act as an enabling technology for knowledge management because it can be used to map and describe knowledge structures, while linking them to information resources. Topic maps also have the ability to capture complex relationships in knowledge flows. Finally, topic maps also act as a learning tool for knowledge management practitioners by enabling users to become familiar with the concepts in an area of interest and the relationships between them. The different views that can be created and the different levels of information that can be represented acknowledge that different users learn differently and allow them to do so at their own pace (cf. chapter 5.10).

6.3 Conceptual Model

The discussion above summarised the important concepts, conclusions and arguments made in the process to answer the primary research question that referred to the relationship between knowledge management and information architecture and the subsidiary question of how such a relationship can be conceptualized. In this section the researcher thus develops a conceptual model to draw together and logically categorise all the relevant aspects that evolved from the study and which explicates the relationship between the two concepts. This is further in line with the goals and objectives of the research project as outlined in chapter 1.7. The conceptual model that evolved in this way is graphically represented by means of a topic map in Figure 1 below.

Figure 1 - Conceptual Model – Phase I

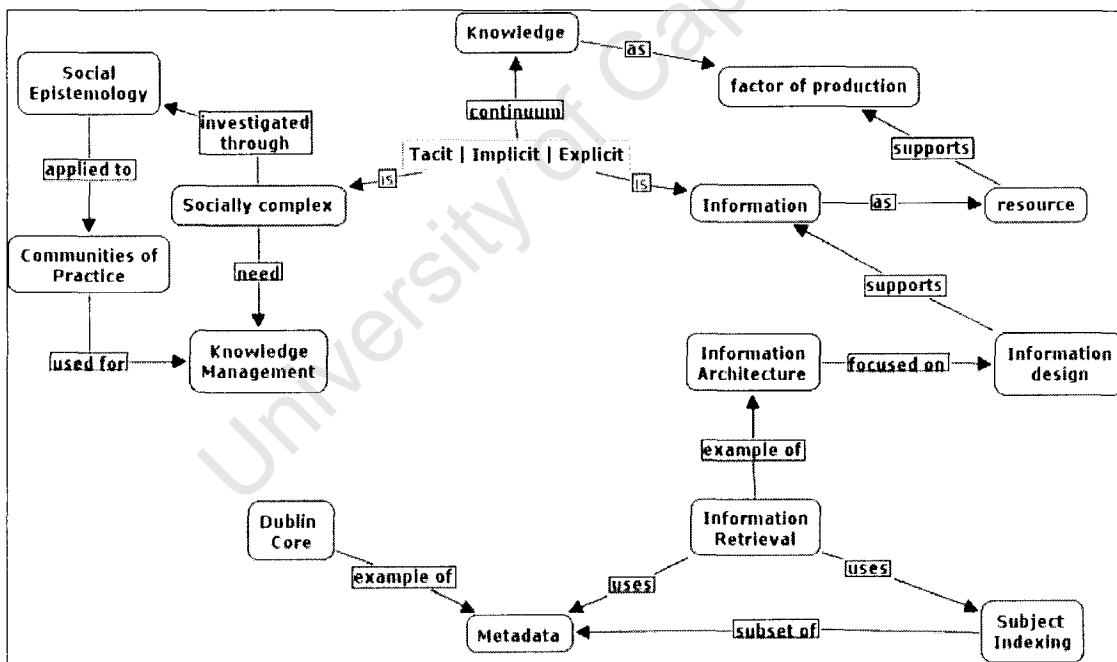


Figure 1 depicts the relationship between knowledge management and information architecture as it emerged during the developing stages of the dissertation. In concise terms the researcher suggests that the relationship is based on the premise that information architecture, through information design and structure (i.e. classification), improves access to explicit knowledge that, as part of the knowledge continuum, is the

object that knowledge management aims to manage. The other segments of the knowledge continuum, tacit and implicit knowledge, are managed through initiatives such as communities of practice. Information architecture, however, also has a role to play in rendering these less tangible knowledge forms more accessible by creating logical referents to the embodiments of such knowledge (i.e. the people with the knowledge, know-how, etc.) in electronic systems such as intranets, etc.

The expanded description of the relationship between knowledge management and information architecture is seen in the light of information as a resource that supports knowledge as a factor of production. In this conceptual model knowledge management is defined as “a set of systematic and disciplined actions an organization takes to obtain the greatest value from the knowledge available to it” (Marwick, 2001: 814, also chapter 2.4.). It was seen that knowledge management evolved in order to harness the power of knowledge, which is socially complex and exists on a continuum that includes tacit, implicit and explicit information. Tacit and implicit knowledge is best managed through attempts that recognise the social complexity of sharing and using these types of knowledge in an organisational framework. The social nature of knowledge can be investigated through the field of social epistemology. Communities of practice provide an example where the social nature of knowledge and the sharing of tacit and implicit knowledge are actively encouraged. Explicit knowledge has many of the characteristics of information and therefore is best managed using the available technologies and tools that have been developed to manage information. In order for information to be used as a resource, attention needs to be paid to the design and structure of information.

The focus of information architecture has also recently shifted from the previous overemphasis of technology to the design and structure of information. Information retrieval systems are an example of the type of information architecture that uses the structure of information resources to assist users to find what they are looking for. Metadata records and subject-based indexes are examples of information retrieval techniques used to organise information. Previous examples of metadata systems that attempted to provide a coherent strategy for the classification of electronic documents

include the Dublin Core system. However the main shortcomings of the Dublin Core are that different versions are difficult to merge and that not all the fields specified are applicable to each information resource. At this stage, the researcher questioned if alternatives exist to the Dublin Core and other traditional metadata and subject-based indexing systems that would be able to address the shortcomings and which would be able to better represent explicit knowledge as it is understood by humans.

As was mentioned in chapter 1.6 and 1.8.4 the exploratory and inductive nature of the research lead to additional questions arising as the research progressed. The first related to the problems encountered with traditional information retrieval techniques, especially in heterogeneous collections. Given the expansion in the capabilities of information technology, the question thus arose whether other technologies or tools existed that would improve traditional information retrieval techniques? This investigation led to the identification of topic maps as a relatively new technological application that shows great promise for addressing the concerns of traditional metadata systems and subject-based indexing. Further investigation into topic maps led the researcher to consider the possibilities of topic maps in the knowledge management framework based on her understanding of knowledge management. Understanding the different features of topic maps also raised the question if topic maps could be used to clearly explicate the relationship between knowledge management and information architecture. The conceptual model represented in Figure 2 is a visual representation of the culmination of the answers to each of the questions mentioned above.

The researcher thus suggests that topic maps could fruitfully be used as an information architecture tool because it facilitates the finding of information, by providing the ability to group related items and by mapping the possible linkages to other related information. The depth of information recorded in a topic map further also makes it an ideal tool for knowledge management applications. Different knowledge flows can be recorded and the dynamic nature of explicit knowledge can also be captured through detailed associations and multiple levels of information about the same topic or information resource. In this way topic maps can be seen as providing a direct link between all the facets of knowledge management and information architecture. Topic maps enable both fields to develop their own agendas, while at the same time assisting both disciplines. Topic maps provide a particularly skilful use of technology that can bridge the gap between knowledge management and information architecture by making explicit knowledge more accessible and by depicting the complex knowledge flows that result from the social nature of knowledge. Topic maps not only address all aspects of knowledge (i.e. the entire continuum), but also link the manifestations of knowledge to the information resources that supports the knowledge. It is thus clear that topic maps hold substantive potential for the field of knowledge management and that they can effectively be used to clearly explicate the relationship between knowledge management and information architecture.

The conceptual model as outlined above and depicted in Figure 1 clearly indicates the intrinsic relation that exists between knowledge management and information architecture. It was further seen that topic maps, although not an essential or fundamental element of this relationship, has the potential (by addressing the requirements of both fields) to play an important role as a tool that that can improve the functionality of both knowledge management and information architecture. Topic maps further also depict the relationship between the two concepts more explicitly by creating a shorter and more direct visual link between knowledge management and information architecture. These features are discussed above and graphically depicted in Figure 2.

The researcher therefore suggests that the goal and objective of this research project to develop a conceptual model that investigates the linkages between knowledge

management and information architecture has been accomplished by the conceptual models as represented and discussed above.

6.4 *Suggestions for further research*

Chapter 1.8.7 described the role that conceptual models can play in the development of new theories in an area of research. Certain requirements for the development of rigorous conceptual models were also identified. Firstly, a conceptual model should specify the essential components and the relationships between them. This was achieved in the conceptual model presented above. The next step suggested by Engelbart (1962) is to specify how changes in the components or in the relationships between them, would influence the conceptual model. This step is beyond the scope of this research, but the researcher suggests that a study of the potential impact of changes in the components or their relationships be the focus of future research. For example, how does the increased focus on social epistemology in knowledge management affect the perceived role of information architecture in organisations?

One of the areas highlighted in chapter 2 is the social nature of knowledge and the dynamics that this feature adds to the process of managing knowledge. Unfortunately, this dissertation focused more on the information resources that support the development of knowledge within individuals or groups than on how those people would interact with information to create knowledge. Topic maps aim to represent information in ways that more closely resemble the way humans understand and link different concepts. A possible area for research would be to investigate how different representations assist users to make sense of new information or to draw new conclusions from known sources.

Although the emphasis of this research project was on the relationship between knowledge management and information architecture, it further also exposed the value of topic maps as a tool that could benefit both fields. However, very little is written about the application of topic maps in these disciplines. For this reason the researcher suggests the following potential areas for future research:

- The transferability of topic maps. Although topic maps can be transferred between collections, the question is whether they ultimately add value for different user groups and if the understanding of the associations between topics can be transferred. For example, a potential area of research would be to investigate if different communities of practice could understand one other's topic maps in the same field of interest.
- The practical implementation of topic maps at South African Academic libraries. The suggested focus would be on the knowledge sharing features of topic maps and what specific features are required for a topic map to be successfully transferred from one collection to the other. This investigation could also link to the question of how topic maps could support knowledge management in an academic library environment.

Topic maps also raise other philosophical questions. For example, what knowledge is universally true in all circumstances? Moreover, on a practical level: how often does knowledge change and need to be updated? The first question is important because it addresses the fundamental positive assumptions about topic maps. If knowledge is completely situation specific, as some authors believe, then topic maps would have very little use. Although topic maps are not limited in the amount of information they can encapsulate, creating a topic map that includes all known and possible known information would be very difficult. This leads to another philosophical question: is it possible to map all knowledge/information, for the world or even for a given field of interest? It is unlikely that one person would be able to do so. However, this creates opportunities for communities of practice in different fields to collaborate and expand the sharing of knowledge.

The next question is, whether all knowledge/information should be mapped? Would this mean a return to the days of trying to codify all knowledge? As was discussed in chapter 2, the process of attempting to codify knowledge has certain drawbacks and can limit the

effectiveness of knowledge management initiatives. An assumption is that some areas would be easier to map than others, for example geography vs. religion.

The above brief discussion of questions relating to topic maps and their role in the world has highlighted some topics that can be explored in further research. The theory of topic maps is very interesting and it provides the possibility for a multitude of interesting projects. As with other technologies that came before it, topic maps should not be seen as a panacea, but should be used in its fullest capacity in the areas where it can add value. Testing the abilities and functionality of topic maps in the real world will also indicate if the theoretical possibilities embedded in topic maps, can be translated into tangible organisational benefits.

6.5 Conclusion

As mentioned in Chapter 1, this research project followed an exploratory, inductive research model that utilised an adapted grounded theory approach to analyse the concepts and themes that were derived from an analysis of the literature in the field. The structure of the dissertation thus reflects the iterative approach adopted. The various concepts and categories that evolved during the entire process were finally synthesised into the two conceptual models. During this process the researcher has attempted to answer the research questions (cf. chapter 1.6.) in an integrated way.

The researcher would finally like to conclude the dissertation by providing a concise, synthesised overview of how the dissertation has addressed these research questions

The primary research question asked

- What is the relationship between knowledge management and information architecture?

This was answered by identifying the positive role that information architecture plays in enhancing the retrieval of information/explicit knowledge by creating logical structures or classifications in the design of information retrieval systems. Through the clear

categorization of explicit knowledge, information architecture supports access to the knowledge continuum, an important focus of knowledge management.

The secondary question asked

- How can this relationship be clearly conceptualised?

This question was addressed by the conceptual models that were presented in 6.3. and graphically depicted by using topic mapping techniques in Figures 1 and 2.

The inductive process exposed the use of topic maps as an enabling technique and this led to the conceptualisation of subsequent research questions. The first question that arose from the ongoing investigation asked

- What technologies or tools exist that would improve traditional information retrieval techniques?

This was answered through the discussion of the relative advantages that topic maps have as a metadata and subject-indexing tool. This relates especially to the ability of topic maps to accurately name concepts, reflect the complexity of the relationships between concepts and to record the necessary information about the resources that refer to the concepts. The ways in which topic maps record information also make them transferable between collections and easier to merge. This is a great improvement on previous metadata capabilities.

However, topic maps do not only improve information architecture techniques but they also hold potential for the field of knowledge management. This is embodied both indirectly through the improved access to explicit knowledge and directly through the ability to map knowledge flows. They further also have the potential to assist learning in organisations by providing a better understanding of the various areas of knowledge by means of their explicit mapping of the fields of knowledge and their relationships. This clearly also answered the question

- What potential do topic maps hold for the field of knowledge management?

The last question that was posed asked

- Can topic maps be used to clearly explicate the relationship between knowledge management and information architecture?

The final conceptual model shows that topic maps can be used to directly link information architecture and knowledge management.

It is suggested that while this dissertation has clearly shown that topic maps have the potential to enhance the supporting role of information architecture to knowledge management many of the shortcomings inherent in traditional systems remain despite the alternatives provided by topic maps. One of the stumbling blocks is the amount of knowledge human beings possess, the multitude of ways they express it and the unlimited language possibilities that can be used to describe and convey what they know. Having said that, the researcher is still of the opinion that topic maps can facilitate the process of knowledge sharing by enriching the support offered by information architecture to knowledge management.

University of Cape Town

Reference list

- Aarons, J. 2006. Epistemology and knowledge management. In *Knowledge management encyclopaedia*. [Online] Available at www.igi-online.com [27 March 2007].
- Ahmed, K. & Moore, G. 2005. An introduction to topic maps. *Microsoft architect journal*. [Online] Available at <http://msdn2.microsoft.com/en-us/library/aa480048.aspx> [27 March 2007].
- Al-Hawamdeh, S. 2002. Knowledge management: Rethinking information management and facing the challenges of managing tacit knowledge. *Information research*. 8(1): paper no 143. [Online] Available at <http://informationr.net/ir/8-1/paper143.html> [27 March 2007].
- Ambert, A., et al. 1995. Understanding and evaluating qualitative research. *Journal of marriage and the family*. 57(4): 879 – 893. [Online] Available from JSTOR at www.jstor.org [27 March 2007].
- American heritage dictionary. 2007. [Online] Available at www.answers.com [26 March 2007].
- Babbie, E. 1995. *The practice of social research*. 7th Ed. California: Wadsworth.
- Babbie, E. & Mouton, J. 2001. *The practice of social research*. Cape Town: Oxford Press.
- Baek, S., Liebowitz, J., Prasad, S.Y. & Granger, M. 1999. Intelligent agents for knowledge management – Toward intelligent web-based collaboration within virtual teams In *Knowledge management handbook*. Ed. J. Liebowitz. Boca Raton: CRC Press. 11:1-11:23.

- Barney, J. 1991. Firm resources and sustained competitive advantage. *Journal of management*. 17(1): 99-119 [Online] Available from EBSCOHost: Business Search Premier at <http://web.ebscohost.com> [27 March 2007].
- Bater, B. 2004. Topic Maps: Indexing in 3D. In *Information architecture: Designing information environments for purpose*. A. Gilchrist & B. Mahon, Eds. New York: Neal-Schuman 132-144.
- Beckman, T.J. 1999. The current state of knowledge management. In *Knowledge management handbook*. Edited by J. Liebowitz. Florida: CRC Press. 1-22.
- Bell, D. 1974. *The coming of the post industrial society: A venture into social forecasting*. London: Heinemann.
- Blaikie, N. 2004. Induction. In *SAGE Encyclopaedia*. Ed. M.S. Lewis-Beck, A. Bryman, T.F. Liao. Thousand Oaks: Sage Publications. 486-487.
- Bradley, J. 1993. Methodological issues and practices in qualitative research. *Library Quarterly*. 63(4): 431 – 449.
- Brown, J.P., Massey, A.P. & Boling, E. 2005. Evaluation of knowledge management systems: Insights form the study of a technical support knowledge base. *Knowledge management research and practice*. 3(2): 49-59 [Online] Available from SwetsWise at www.swetswise.com/eAccess/viewToc.do?titleID=119916&yevoID=1501534 [27 March 2007].
- Buchanan, B. 1979. *Theory of library classification*. London: Bingley.
- Burnett, R. & Marshall, P.D. 2003. *Web theory: An introduction*. London: Routledge.

- Butler, T. 2006. Anti-foundational knowledge management. In *Encyclopaedia of knowledge management*. [Online] Available at <http://www.igi-online.com> [27 March 2007].
- Charmaz, K. 2004. Grounded Theory. In *SAGE Encyclopaedia of social science research methods*. Ed. M.S. Lewis-Beck, A. Bryman & T.F. Liao. Thousand Oaks: SAGE Publications. 2: 440-444.
- Churchill, E.F. & Bly, S. 1999. Virtual environments at work: ongoing use of MUDs in the workplace. Proceedings of the international joint conferences on work activities coordination and collaboration '99. San Francisco: ACM Press. 99-108.
- Churchman, C.W. 1971. *The design of inquiring systems*. New York: Basic Books. Quoted by Malhorta, Y. 2000. From information management to knowledge management: Beyond the Hi-tech hidebound system. In *Knowledge management for the information professional*. Ed T.K Srikantaiah. & M.E.D. Koenig. New Jersey: Information Today. 35-61.
- Concise dictionary of library and information science*. 2000. 2nd Ed. Edited by S. Keenan & C. Johnston. London: Bowler-Saur.
- Davenport, T.H. 1994. *Coming soon: The CKO*. InformationWeek. 5 September 1994 [Online] Available at <http://web.lexis-nexis.com/professional> [27 March 2007].
- Davenport, T.H. & Prusak, L. 1998. *Working knowledge: How organizations manage what they know*. Boston: Harvard Business School Press.
- Day, R.E. 2005. Clearing up "implicit knowledge": Implications for knowledge management, information science, psychology and social epistemology. *Journal of the American society for information science and technology*. 56(6):630-635 [Online] Available from ProQuest at <http://proquest.umi.com> [27 March 2007].

- Dey, I. 1993. *Qualitative data analysis: A user-friendly guide for social scientists*. London: Routledge.
- Drucker, P.F. 1969. *The age of discontinuity*. Revised. New Jersey: Transaction Publishers.
- Drucker, P.F. 2001. The information executives need today. In *The essential Drucker*. 95-111. Oxford: Butterworth-Heinemann.
- Economist Intelligence Unit. 2005. *Foresight 2020: Economic, industry and corporate trends*. Economist Survey Unit [Online] Available at newsroom.cisco.com/dlls/tln/research_studies/2020foresight/pdf/2020foresight_full_report.pdf?CMP=AF17154 [29 March 2007].
- Engelbart, D.C. (1962). Augmenting human intellect: a conceptual framework. Menlo Park, CA: Stanford Research Institute. (Summary report AFOSR-3233) [Online] Available at <http://www.bootstrap.org/augdocs/friedewald030402/augmentinghumanintellect/ahi62index.html> [27 March 2007].
- Everden, R. & Evernden, E. 2003. Third generation information architecture. *Communications of the ACM*. 46(3): 95-98 [Online] Available from EBSCOHost: Business Source Premier at <http://web.ebscohost.com> [27 March 2007].
- Felderer, B. & Homburg, S. 1992. *Macroeconomics and new macroeconomics*. New York: Springer-Verlaag.
- Fischer, M. 2004. Developing an information model for information and knowledge based organisations. In *Information architecture: designing information environments for purpose*. Gilchrist, A. & Mahon, B. Eds. New York: Neal-Schuman Publishers. 5-26.

Foskett, A.C. 1996. *The subject approach to information*. 5th ed. London: Library Association Printing.

Garshol, L.M. 2002. What are topic maps? [Online] Available at www.xml.com/pub/a/2002/09/11/topicmaps.html [27 March 2007].

Garshol, L.M. 2004. Metadata? Thesauri? Taxonomies? Topic Maps! *Journal of information science*. 30(4) 378-391 [Online] Available at www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html [19 November 2006].

Glaser, B.G. 1992. *Basics of grounded theory analysis*. California: Sociology Press.

Goldman, A.I. 1999. *Knowledge in a social world*. Oxford: Clarendon Press.

Hammersley, M. 2004. Objectivity. In *SAGE Encyclopaedia of social science research methods*. Ed. M.S. Lewis-Beck, A. Bryman & T.F. Liao. Thousand Oaks: SAGE Publications. 2: 250-251.

Hanisch, T. 2005. Metadata for electronic documents using the Dublin Core. In *Encyclopaedia of information science and technology*. Ed. M. Khosrou-Pour. 1928-1930 [Online] Available at <http://reference.igi-online.com> [27 March 2007].

Huxor, A. 1998. An active worlds interface to BSCW to enhance chance encounters. In *Proceedings of Collaborative virtual environments 1998*. June 1998. Manchester, United Kingdom. Quote by Thomas, J.C., Kellogg, W.A & Erickson, T. 2001. The knowledge management puzzle: Human and social factors in knowledge management. *IBM Systems journal*. 40(4): 863-884 [Online] Available from Directory of Open Access Journals at www.research.ibm.com [29 March 2007].

International Standards Organization ISO/IEC 13250 Topic maps. 2000. [Online] Available at www.y12.doe.gov/sgml/sc34/document/0129.pdf [16 November 2006].

Jacso, P. 2003. It's the output, stupid! *Information today*. 20(8): 17-19 [Online] Available from EBSCOHost at www.ebscohost.com [29 March 2006].

Jastroch, N. 2001. The information age: Remarks and basic implications. Knowledge Board. [Online] Available at www.knowledgeboard.com [29 March 2007].

Jarvelin, K. & Wilson, T.D. 2003. On conceptual models for information seeking and retrieval research. *Information Research* 9(1) paper 163 [Online] Available at <http://InformationR.net/ir/9-1/paper163.html> [8 March 2007].

Kermally, S. 1999. *When economics means business: The new economics of the information age*. London: Financial Times Management.

Kunze, J.A. 2001. A metadata kernel for electronic permanence. *Journal of digital information*. 2(2) [Online] Available at <http://journals.tdl.org/jodi/article/view/jodi-38/46> [20 November 2006].

Lancaster, F.W. 1986. *Vocabulary control for information retrieval*. Arlington: Information Resources Press.

Lesser, E.L. & Storck, J. 2001. Communities of practice and organisational performance. *IBM Systems Journal*. 40(4): 831 – 841. [Online] Available from EBSCOHost at <http://web.ebscohost.com> [29 March 2007].

Malone, D. 2002. Knowledge management: A model for organisational learning. *International journal of accounting information*. 37:1-13 [Online] Available from Science Direct at www.sciencedirect.com [29 March 2007].

- Malhorta, Y. 2000. From information management to knowledge management: Beyond the Hi-tech hidebound system. In *Knowledge management for the information professional*. Edited by Srikantaiah, T.K. & Koenig, M.E.D. New Jersey: Information Today. 35-61.
- Marwick, A.D. 2001. Knowledge management technology. *IBM Systems Journal*. 40(4): 814-830. [Online] Available from EBSCOHost at <http://web.ebscohost.com> [29 March 2007].
- Morton, A. 2003. *A guide through the theory of knowledge*. 3rd Ed. Oklahoma: Blackwell Publishing.
- Mouton, J. & Marais, H.C. 1990. *Basic concepts in the methodology of the social sciences*. Pretoria: HSRC.
- Nonaka, I. 1991. The knowledge creating company. *Harvard business review*. Nov-Dec 1991: 96 – 104. [Online] Available from Business Source Premier at <http://web.ebscohost.com> [29 March 2007].
- Olsen, G.M. and Olsen, J.S. 2000. Distance matters. *Human-computer interaction*. 15(2/3): 107-137. [Online] Available from Business Source Premier at <http://web.ebscohost.com> [29 March 2007].
- Pepper, S. 2002. *The TAO of topic maps*. Oslo: Ontopia AS. [Online] Available at www.ontopia.net/topicmaps/materials/tao.html [29 March 2007].
- Polanyi, M. 1967. *The tacit dimension*. London: Routledge & Paul K.
- Ponelis, S. & Fairer-Wessels, F.A. 1998. Knowledge management: a literature overview. *South African journal of library and information science*. 66(1): 1- 9.

- Prusak, L. 2001. Where did knowledge management come from. *IBM Systems Journal*. 40(4): 1002-1007 [Online] Available from EBSCOHost at <http://web.ebscohost.com> [29 March 2007].
- Rosenfeld, S. & Morville, P. 1998. *Information architecture for the World Wide Web*. Beijing: O'Reilly.
- Rowley, J.E. 1988. *Abstracting and indexing*. London: Clive Bingley.
- Rowley, J.E. 1992. *Organising knowledge: an introduction to information retrieval*. England: Ashgate.
- Rubin, A. & Babbie, E. 1993. *Research methods for social work*. 2nd Ed. California: Brooks/Cole Publishing.
- Ruggles, R. 1998. The state of the notion: Knowledge management in practice. *California Management Review*. 40(3): 80-89. [Online] Available from Business Source Premier at <http://web.ebscohost.com> [29 March 2007].
- Segars, A.H. & Grover, V. 1996. Designing company-wide information systems: Risk factors and coping strategies. *Long range planning*. 29: 381-392 [Online] Available from Science Direct at www.sciencedirect.com [29 March 2007].
- Selltiz, C., Jahoda, M., Deutsch, M. & Cook, S.W. 1965. *Research methods in social relations*. Revised. London: Methuen.

- Sigel, A. 2000. Towards knowledge organization with topic maps. *Conference proceedings XML Europe 2000*. [Online] Available at www.infoloom.com/gcaconfs/WEB/paris2000/S22-02.HTM [29 March 2007].
- Smith, J.K. 2004. Reliability and validity in qualitative research. In *SAGE Encyclopaedia of social science research methods*. Ed. M.S. Lewis-Beck, A. Bryman & T.F. Liao. Thousand Oaks: SAGE Publications. 3:957.
- Sowa, J.F. & Zachman, J.A. 1992. Extending and formalizing the framework for information systems architecture. *IBM Systems journal*. 31(3): 590-615 [Online] Available at EBSCOHost at <http://web.ebscohost.com> [29 March 2007].
- Spool, J. et al. 1999. *Web site usability: A designers guide*. San Francisco: Morgan Kaufmann.
- St. Clair, G. 2003. Knowledge management. In *Encyclopaedia of library and information science*. 2nd ed. Edited by M.A. Drake. New York: Marcel Dekker. 1486-1494.
- Storck, J. & Hill, P. 2000. Knowledge diffusion through “strategic communities”. *Sloan Management Review*. 41(2): 63-74. Quoted by Lesser, E.L. & Storck, J. 2001. Communities of practice and organisational performance. *IBM Systems Journal*. 40(4): 831 – 841. [Online] Available from EBSCOHost at <http://web.ebscohost.com> [29 March 2007].
- Tesch, R. 1990. *Qualitative research: Analysis types and software tools*. New York: Falmer Press.
- Thomas, J.C., Kellogg, W.A & Erickson, T. 2001. The knowledge management puzzle: Human and social factors in knowledge management. *IBM Systems journal*. 40(4): 863-884 [Online] Available from EBSCOHost at <http://web.ebscohost.com> [29 March 2007].

- Wade, M. & Hulland, J. 2004. Review: The resource-based view and information systems research: Review, extensions and suggestions for further research. *MIS Quarterly*. 28(1): 107 – 142 [Online] Available at <http://web.ebscohost.com> [22 September 2007].
- Warner, A.J. 2004. Information architecture and vocabularies for browse and search. In *Information architecture: Designing information environments for purpose*. Gilchrist, A. & Mahon, B. Eds. New York: Neal-Schuman Publishers. 177-191.
- Wilson, T.D. 2002. The nonsense of “knowledge management”. *Information research*. 8(1), paper no. 144 [Online] Available at <http://InformationR.net/ir/8-1/paper144.html> [29 March 2007].
- Wurman, R.S. 1996. *Information architects*. Zurich: Graphics Press. Quoted by Rosenfeld, S. & Morville, P. 1998. *Information for the world wide web*. California: O’Reilly.
- Zachman, J.A. 1987. A framework for information systems architecture. *IBM Systems journal*. 26(3): 276-292 [Online] Available from EBSCOHost at <http://web.ebscohost.com> [29 March 2007].