A Framework for Computer-Based Knowledge Management Systems

A dissertation submitted to the
Department of Computer Science,
Faculty of Science, University of Cape Town
in partial fulfilment of the requirement
for the degree of
Masters of Information Technology.

By

Victor Katoma

February 2007

Supervised by

Sonia Berman.

Dedication

This thesis is dedicated to the people who supported me through the most trying times of my life: My parents Andrew and Victoria, my brother Francis, my friends Sam, Candice and Nuroo Ismail from the Knowledge Commons, UCT Libraries.

Acknowledgement

Thank you to my supervisor, Sonia Berman for her insight and invaluable contributions to this research. Thank you to James Gain for sparking my interest in this research.

Abstract

Human knowledge is now recognised as a key organisational asset. Different knowledge needs require different knowledge management systems (KMS). A KMS is an electronic system that facilitates management of knowledge and at the same time returns and leverages knowledge. This study aims at investigating a framework that utilises the latest advances in KMS technology. This research outlines the framework and describes a system developed around this. The framework is based on fundamental knowledge management concepts, Resource Description Framework (RDF) and basic modeling concepts and tools. It also incorporates notions of knowledge quality measure and a simple inference mechanism in order to probe, learn and adapt the KMS with the changing needs of its users.

In particular, a prototype was developed targeting the research training context of post-graduate education. An evaluation of the prototype showed that the framework was plausible and that users found it appropriate for their needs

Table of Contents

DEDICATION I

ACKNOWLEDGEMENT II

ABSTRACT III

CHAPTER 1 1

INTRODUCTION 1

1.1 INTRODUCTION 1
1.2 RESEARCH GOALS 1
1.3 PROTOTYPE SYSTEM 2
1.4 THESIS OUTLINE 3

CHAPTER 2 4

KNOWLEDGE MANAGEMENT 4

2.1 KNOWLEDGE 4
2.2 EPISTEMIC THEORIES 5
2.3 KNOWLEDGE MANAGEMENT 6
2.4 SHARING KNOWLEDGE 7
2.5 KNOWLEDGE PRESENTATION 8
2.6 KNOWLEDGE AUDITING 9
2.7 KNOWLEDGE ACCUMULATION CHANGES OVER TIME 10
2.8 WHEN SYSTEMS ARE LIKELY TO FAIL 11
2.9 CRITICAL REQUIREMENTS OF KNOWLEDGE MANAGEMENT SYSTEMS 11
2.10 SUMMARY 12

CHAPTER 3 13

SEMANTIC WEBLOGS 13

3.1 POSSIBLE KMS PLATFORMS 13
3.2 WEBLOGS 14
3.3 RSS 16
3.4 AGGREGATORS 20
3.5 RDF 21

3.6 SEMANTIC WEBLOGS 23
3.7 KM NEEDS THAT WEBLOGS DO NOT OFFER 23
3.8 SUMMARY 24

CHAPTER 4 25

MAXIMUM LIKELIHOOD ESTIMATION 25
4.1 SIGMA ALGEBRAS 25
4.2 MEASURE THEORY AND MLE 26
4.3 EXAMPLE 27
4.4 MLE REQUIREMENTS 27
4.5 SIGMA ALGEBRAS AND MLE IN A KNOWLEDGE MANAGEMENT SYSTEM 29
4.6 ADVANTAGES OF MLE 30
4.7 DISADVANTAGES OF MLE 30
4.8 SUMMARY 31

CHAPTER 5 32

A KNOWLEDGE MANAGEMENT FRAMEWORK 32
5.1 RESEARCH TRAINING 32
5.2 THE DATA MODEL 32
5.3 THE FRAMEWORK 34
5.4 SYSTEM OVERVIEW 35
5.5 SYSTEM USE CASES 36
5.6 USE CASES FOR THE KMS EDITOR 37
5.7 SUMMARY 38

CHAPTER 6 39

PROTOTYPE KMS DESIGN 39
6.1 KNOWLEDGE AUDIT 39
6.2 AUDIT ANALYSIS 42
6.2.1 Knowledge search 42
6.2.2 Knowledge Accessibility and sharing 43
6.2.3 Knowledge Restructuring 43
6.2.4 Audit Results Summary 44
6.3 KC ORGANIZATION 45
6.4 KC OBJECT LIFE CYCLE 47
6.5 OBJECT PROGRESS THROUGH KC 48

6.6 KNOWLEDGE PROPERTIES 48
6.7 SUMMARY 49

CHAPTER 7 50

THE MANILA CONTENT MANAGEMENT SYSTEM 50
7.1 THE FRONTIER DATABASE SYSTEM 50
7.2 DEVELOPING WEB PAGES 52
7.3 MANILA'S BUILT-IN MACROS 55
7.4 MANILA AS A BASIS FOR KMS DEVELOPMENT 57
7.5 FRONTIER TABLES USED IN THE PROTOTYPE KMS 57
7.6 SUMMARY 58

CHAPTER 8 59

STATISTICAL ANALYSIS IN KC 59
8.1 IMPROVING KC USING KNOWLEDGE PROPERTIES 59
8.2 USING R FOR MLE 60
8.3 USING MLE IN KC 61
8.4 MULTIVARIATE ANALYSIS 66
8.5 VISUAL KNOWLEDGE MEASURES FOR KC 69
8.6 SUMMARY 73

CHAPTER 9 74

KC EVALUATION 74
9.1 INITIAL EVALUATION 74
9.2 FOLLOW-UP EVALUATION 75
9.3 SUMMARY 78

CHAPTER 10 80

CONCLUSION 80
10.1 THESIS SUMMARY 80
10.2 CONCLUSION 80
10.3 FUTURE WORK 81

REFERENCES **83**

APPENDICES **93**

APPENDIX A (96): AUDIT QUESTIONNAIRE 94

APPENDIX B - SAMPLE KC PAGES 97

APPENDIX C (96): EVALUATION QUESTIONNAIRE FOR MSC. STUDENTS 107

APPENDIX D (96): USABILITY QUESTIONNAIRE 109

APPENDIX E (96): BLOG SOFTWARE FEATURES ANALYSIS CHART 114

Listof Figures

FIGURE 1: KNOWLEDGE CYCLE [BK06] 7

FIGURE 2: KNOWLEDGE ACCUMULATION, SHARING AND KM FAILURE 10

FIGURE 3: RSS STRUCTURE SAMPLE 16

FIGURE 4: DETAILED RSS SAMPLE 17

FIGURE 5: RDF SAMPLE CODE 18

FIGURE 6: SAMPLE WEBLOG [HTTP://BLOGS.LAW.HARVARD.EDU/TECH] 19

FIGURE 7: XML NEWS FEED [HTTP://WWW.COMPUTERWORLD.COM] 20

FIGURE 8: XML NEWS AGGREGATOR 21

FIGURE 9: RDF GRAPH 22

FIGURE 10: RDF EXTENDED GRAPH 22

FIGURE 11: DATA MODEL 33

FIGURE 12: SYSTEM PROCESS OVERVIEW 35

FIGURE 13: USE CASE 36

FIGURE 14: MANAGE KMS EXPLORED 37

FIGURE 15: ANALYSIS: SHOWS THE RESULTS OF THE SURVEY: 41

FIGURE 16: FILE (ITEM) CYCLE 47

FIGURE 17: SYSTEM STRUCTURE [BK07] 48

FIGURE 18: XML IN FRONTIER 51

FIGURE 19: DEVELOPMENT OF THE WEB PAGES 52

FIGURE 20: HTML GENERATION 53

FIGURE 21: INDEXING PAGES 53

FIGURE 22: RENDERED PAGE 54

FIGURE 23: ADJUSTING WEB PAGES PROPERTIES 54

FIGURE 24: R SOFTWARE 60

FIGURE 25: R MLE MODEL 61

FIGURE 26: CONTRIBUTION BAR GRAPH PRODUCED BY R 62

FIGURE 27: FITTING A NORMAL DISTRIBUTION GRAPH TO SAMPLE DATA IN R. 63

FIGURE 28: MLE FOR ITEM READS 64

FIGURE 29: BOX PLOTS: SPREAD OF THE RATINGS(X~AXIS) OVER NUMBER OF READS(Y~AXIS) 70

FIGURE 30: DOT CHART: NAMES (Y~AXIS) AND NUMBER OF READS (X~AXIS) 71

FIGURE 31: SNIPPET NUMBER (Y~AXIS) AND RANKING (X~AXIS) 71

FIGURE 32: SYNDICATED ITEMS AND STATS 72

FIGURE 33: HOURLY HITS 73

FIGURE 34: PARTITIONING OF ITEMS ACCORDING TO QUALITY RATING (OR RANKING) 73

FIGURE 35: INTRODUCTORY PAGE 97

FIGURE 36: DIRECTORY MANAGEMENT 98

FIGURE 37: PRIVATE DIRECTORIES OF PARTICIPANTS 99

FIGURE 38 : DEPARTMENTAL GROUPINGS AT UCT 99

FIGURE 39: ITEMS AWAITING RELEASE BY THE EDITOR 100

FIGURE 40 PUBLISHED ITEMS CAN BE ACCESSED THROUGH THIS PAGE 101

FIGURE 41: DISCUSSION FORUM 101

FIGURE 42: POSTED ITEMS TO BE RANKED AND OR COMMENTED ON. 102

FIGURE 43: BULLETIN PAGE 102

FIGURE 44 : IN-HOUSE NOTES FEEDBACK FROM LECTURERS OR EDITOR 103

FIGURE 45: VOTING PAGE. 104

FIGURE 46 : NEWS AGGREGATORS' PAGE 104

FIGURE 47: A COLLECTION OF RSS FEEDS FROM DEPARTMENTS 105

FIGURE 48: QUERY AND SEARCH PAGE 106

List of Tables

TABLE 1: USE CASES 37

TABLE 2: USE CASE 2 – THE EDITOR 38

TABLE 3: MANILA BUILT IN MACROS 56

TABLE 4: CONTRIBUTION FREQUENCY 61

TABLE 5: DATA ABOUT ITEM USEFULNESS 64

TABLE 6: DATA OBJECT FOR MULTIVARIATE ANALYSIS 67

TABLE 7: ANALYSIS SUMMARIES 67

TABLE 8: PLACING OF ITEM ON RATING SCALE 68

TABLE 9: USEFULNESS OF ITEMS ON ONE DISCUSSION TOPIC 69

TABLE 10: PARTICIPANTS (INITIALS) AND NUMBER OF USEFUL ITEMS 69

TABLE 11: RESULTS FROM THE SURVEY OF KC USAGE 74

TABLE 12: SECOND EVALUATION 76

Chapter 1
Introduction

1.1 Introduction

Knowledge management (KM) is complex. Differences in knowledge requirements and approaches to building knowledge management systems (KMS) abound. This study presents a framework for knowledge management. The framework is based on the use of a semantic weblog, the use of knowledge measures, and a mechanism to adapt the system to stay in accord with the changing needs of its users. The framework facilitates the storage, management and sharing of knowledge.

KM transcends many fields including technology, statistics and psychology [RT99, BLA95, DAV98, PRU98, JOH96, NON96]. Such work tends to spark controversies on what should constitute a good KMS. To begin with, this study earmarks three KM essentials, namely a well-structured system, semantic modeling and inference. Because institutional knowledge is predominantly reliant upon individual knowledge elicitation [CA69,Tus77], the system should encourage individuals to acquire, share and disseminate knowledge. KMS input would be human knowledge, best practices, quality measures and relevant technologies such as RSS. Output would be well-structured presentation of material, quality indicators, and suggested groupings of users with similar interests, etc.

1.2 Research Goals

This study looks at:

1. What should a knowledge management system consist of?
2. If management is the goal, then measure is a requirement. Can knowledge quality measures be found and meaningfully used to better manage a knowledge management system?

The high demand for better KMS is a strong motivator of this research; for instance, it is asserted that one trillion dollars has already been invested in KM over the last twenty years [Hyl02]. There are no standards or guidelines for structuring organizational memories; no framework exists for building sustainable KM systems [Sch06]. It is also estimated that 80% of KM systems have failed [Hyl04, SK00]. This means loss in both monetary terms as well as managerial vision. Obviously, it also raises questions as to whether such KMS were built on concrete and well founded principles.

The high dependence on the Internet for knowledge requires that special tools should be employed to garner the knowledge. New software technologies such as XML, RDF and RSS are paving the way to knowledge modeling with metadata harnessing, which in turn makes inference plausible. To address the research questions in this project, a KMS framework was designed that uses metadata and knowledge quality measures to evaluate and adapt the system. A prototype implementation of the framework was built in the form of a knowledge management system for post-graduate research students.

1.3 Prototype System

An institution that provides research training should offer some form of knowledge management software. A good KMS for postgraduates should give students practice in the skills needed to do research successfully: finding relevant information, organizing their knowledge appropriately, reading effectively, writing up their ideas, sharing their ideas with others, giving and responding to feedback and constructive criticism.

Much of world data/information and knowledge units are stored in unorganized and unstructured form, making it very difficult to manage [Hyl04, Per03]. Even though it is hard to state how much of this (data/knowledge units) is on campus repositories, universities are huge knowledge reservoirs by virtue of being academic centers. A KMS for research students was therefore built as a prototype to evaluate the framework and at the same time provide a useful academic tool.

1.4 Thesis Outline

This dissertation is structured as follows. Chapter 2 introduces the fundamental concepts – knowledge and knowledge management systems. Chapter 3 discusses Weblogs and related technologies, and chapter 4 outlines sigma-algebras and maximum likelihood estimation which we use to make inferences about a knowledge base. The remaining chapters cover the contribution of this project. First the proposed framework is presented, and then our prototype realization of this framework, called KC (Knowledge Centre). Chapter 7 describes the use of the Manila [Fro] weblog construction tool in building KC, and the next chapter outlines the use of the R statistical software package [FSC05] for making inferences about KC content. Chapter 9 discusses KC evaluation by student users. This is followed by the conclusion, which presents a summary of the work and some ideas for future research.

Chapter 2

Knowledge Management

This chapter describes what knowledge management is and the concepts around it. These include knowledge itself, epistemic theories, sharing, knowledge auditing and knowledge accumulation trends. It also gives an account of knowledge presentation, critical knowledge management requirements and some causes of knowledge management system failures.

2.1 Knowledge

Knowledge has been defined as "a fluid mix of framed experience, values, contextual information and expert insight that provide a framework for evaluating and incorporating new experiences and information" [DP97]. Earlier research unveiled two forms of knowledge, namely, tacit and explicit [Pol67, Non98, Sam05, Har03].

Tacit knowledge is unstated and is normally hidden in the human mind, while explicit is codified and can easily be extracted, transmitted and shared. Further research yet shows that the two states actually form the distinction between knowing and knowledge [BD98, BC99]. Knowing is taken as some hidden sense that could manifest through action in a particular context, while knowledge is facts that can be easily codified and shared.

Knowledge is asserted to emanate from processed information and to reside in the mind of the knower [Pol92, Chu72]. Knowledge cannot entirely be embedded in resources such as document repositories, routines, processes, practices, and norms [DP98, FP98, Tee98a]. It is hard to isolate and represent objectively, and in its entirety, "a fluid mix of framed experience" or "expert insight" [DP97]. Instead, people independently use knowledge sources to explicate, interpret and ultimately construct knowledge. Knowledge items are prone to different interpretation; different

2.2 Epistemic Theories

Epistemology is a branch of philosophy concerned with the nature and extent of human knowledge [Kle98]. Epistemology dates back to the works of Plato and Aristotle [Ari45]. In recent years, research in epistemology has provided insights [Sve94, 97, 01, Non94, NT95] from which concrete KM elements can be derived. Contemporary investigations have further attempted to specify what legitimizes knowledge, and how humans acquire knowledge [Pap98]. Epistemic theories relevant to knowledge management are *coherentism*, *reliabilism* and *consequentialism*.

By coherentism [Bon85, Leh86] it is asserted that information is assimilated as knowledge through the ability to *integrate* it with existing knowledge in a logical fashion. In reliabilism [Gol79, Sos91, Set96], it is believed that people attach a certainty value to everything they know. As a result, people accept information as knowledge with respect to the *confidence* in the source and the way in which it was obtained. Consequentialism [Gol99] states that information is considered as knowledge only if enough people are known to qualify it as sufficient and *trustworthy*. This concept places emphasis on society at large as the ultimate determinant of the acceptance of knowledge.

From the above epistemic theories it can be asserted [BPA03] that an effective knowledge model should be based on:

1) Easy integration of ideas from organizations and society.

2) Agreement on terminology in the particular domain.

3) Reliability of the knowledge source - and thus the origin of knowledge should also be known.

4) The usefulness of the concept. This is what is known to distinguish knowledge from other beliefs.

5) Freedom to relate knowledge items flexibly and not only according to a particular scheme or structure.

6). Confidence values must be attached to every knowledge item, to capture the reliability associated with it by each person who knows it, and from which the community's acceptance or trust in the item can be ascertained [BPA03].

Epistemic theories provide an account of knowledge production and its nature, but little about knowledge states and processes. They do not suggest a dynamic view (*ongoing interpretation*) of KM that embodies the processes of *storage*, *usage*, *sharing*, and knowledge *dispersal* hence missing the practical aspect [Aar06]. Even though recent studies in epistemology have included sharing (social theory), they cannot be directly applied to knowledge *system processes* [Sch98, Gol99, Lau01, Sol01, Tur02]. The next sections discuss such knowledge management processes.

2.3 Knowledge Management

Knowledge management, can be defined as the "Explicit and systemic management of vital knowledge and its associated processes of creating, gathering, organizing, diffusion, use and exploitation. It requires turning personal knowledge into corporate knowledge that can be widely shared throughout an organization and appropriately applied" [Sky03].

Wherever KM exists or is envisaged, some form of a knowledge management system (KMS) must be present. Thus, KMS adopts the definition of KM as stated and extends this with a managerial system or tool. Devanport pointed out that KMS are often noticeable in implementations such as document repositories, expertise databases, discussion lists, and context-specific retrieval systems incorporating collaborative filtering technologies [Dev98]. The difficult task eluding KM today that has led to many KMS failures is how to formulate and implement sustainable systems embodying and encapsulating the needs and expectations of users. Tapp and Hughes [TH04] have further pointed out that even though KMS have increased the supply of knowledge objects, the usage of these same objects by other workers, "the crucial added value", remains elusive.

It is argued that what people usually termed KMS are simply document management systems [Azb01]. Others have in this light believed that very few enterprises have indeed implemented KM systems [Kin02, Des03], because systems have been developed without an understanding of the problems they are meant to address.

Expectancy theories state that individuals contribute to knowledge activity based on the expectancy of certain benefits. The perceived value from knowledge seeking depends on *contributor's expertise* and *credibility,* while the perceived expectation of value depends on *trust*, *obligation* and contributor *willingness* [Vro64, Kal99].

The figure 1 below depicts the flow of knowledge elements and how humans can interact with it. This is a rudimentary structure signifying the beginning of what can be achieved over time. The flow may not strictly follow the indicated paths.

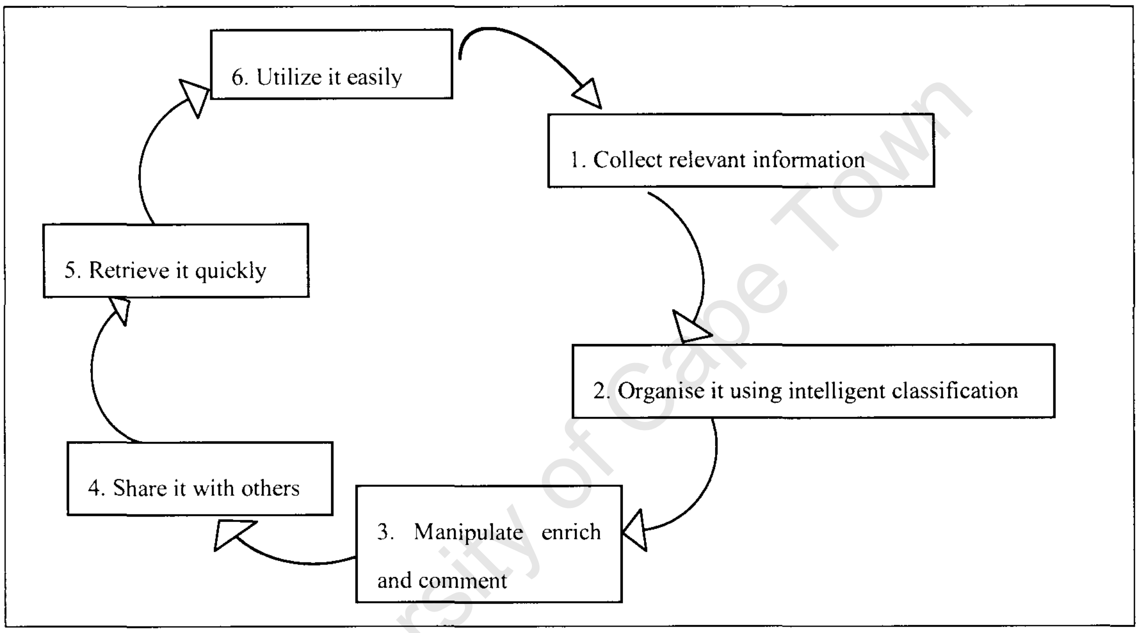


Figure 1: knowledge cycle [BK06]

2.4 Sharing knowledge

In the knowledge-intensive economy sharing and KM become inseparable [NT95, Hyl03, Bix02, Rom02]. Sharing of electronic material now ranges from merely document sharing to syndication [Wil05] (posting summaries and links to Web items so that they can easily be shared). This can complement the natural ways institutions have of creating, sharing and transferring knowledge [SK06, IN00]. *Externalisation* and *internalisation* convert tacit to explicit knowledge and vice versa. In attempting to externalise tacit knowledge, two methods are common, *intervention* and *representation.* By intervention, the system identifies and connects the needy to the knowledge owners. Representation is what technologies seek in automating knowledge processes through forums, networks and other internet-enabled interactions.

Sharing of knowledge should therefore be structured in systematic ways; especially when teams have common ground or want to establish mutual knowledge [Gla02].

2.5 Knowledge presentation

Important to knowledge presentation especially in academia are media functions. These are basically the methods of content delivery. There are three media functions, namely primary, secondary and tertiary [Hem01]. Primary is text without predefined order and therefore the reader painstakingly extracts what is necessary. Secondary requires that educational materials are arranged coherently in a specific structure or layout with associated ordering. Tertiary exhibits artificial intelligence in the sense that the system learns and *adapts* to the user's requirements.

Today, hypermedia (text, audio, video) are the main media used for electronic communication [GM96]. *Hyperlinking* is an easy way of engaging with knowledge resources [ZRC99] and provides ideal avenues for constructing knowledge relationships [Shu98, Rom02]. Researchers have different opinions on whether high quality hypermedia leads to improvement in training [HRK01]. For example, The Web and hyperlinking have been noted for contributing to management information overload. This problem is primed to become worse unless effective *filters* such as *user profiling* are developed [FD02]. Although e-mail is one of the most utilized features of the Internet, it has not recorded significant impact as a KM tool due to the rapid accumulation of in e-mail accounts.

2.6 *Knowledge auditing*

Designing and implementing a sustainable KM system first requires assessment of current knowledge and needs. This process is called knowledge auditing [Mic01]. Basic knowledge accounting begins here because, what is audited gets included and what is included essentially gets counted [Hyl01]. Knowledge auditing can in fact provide information about the quality of the KM process as a whole.

Knowledge mapping is a process of graphically defining related knowledge areas [Alf03]. It includes the identification and linking of critical information and knowledge competencies of an organization. These stop short of assigning knowledge measures; but measures can be obtained by analyzing answers to audit questions.
Knowledge auditing is normally done through a questionnaire followed by interviews with the potential system users. For a high degree of system acceptance, users therefore have to be part of the process from start to end. [Hyl04] and [SF00] propose the following as vital questions to be addressed.

1) Who does one go to when one has a problem? 2) Do knowledge maps exist in the institution? If not, why not? If yes then are they used effectively? 3) With whom do knowledge people collaborate and share information and knowledge? 4) What are the barriers to knowledge sharing? 5) Do people know what they should know but don't in fact know? 6) Do people feel that their knowledge is valued? 7) Do people routinely document knowledge for repeated use? 8) What do people do with knowledge that is accrued from completed tasks or projects? 9) How do people get the information and knowledge they need? 10) Do people get the knowledge they need? 11) Who are the providers of the most critical knowledge and how do people get it? 12) Does the institution reward people for sharing? If yes, are the rewards appropriate? 13) What are people doing to tap into existing and potential knowledge sources? 14) Where is the knowledge hiding? 15) Are people deriving value from the knowledge system? If not, why not?

2.7 Knowledge accumulation changes over time

The figure 2 below illustrates how knowledge accumulation occurs in an institution using a knowledge management system.

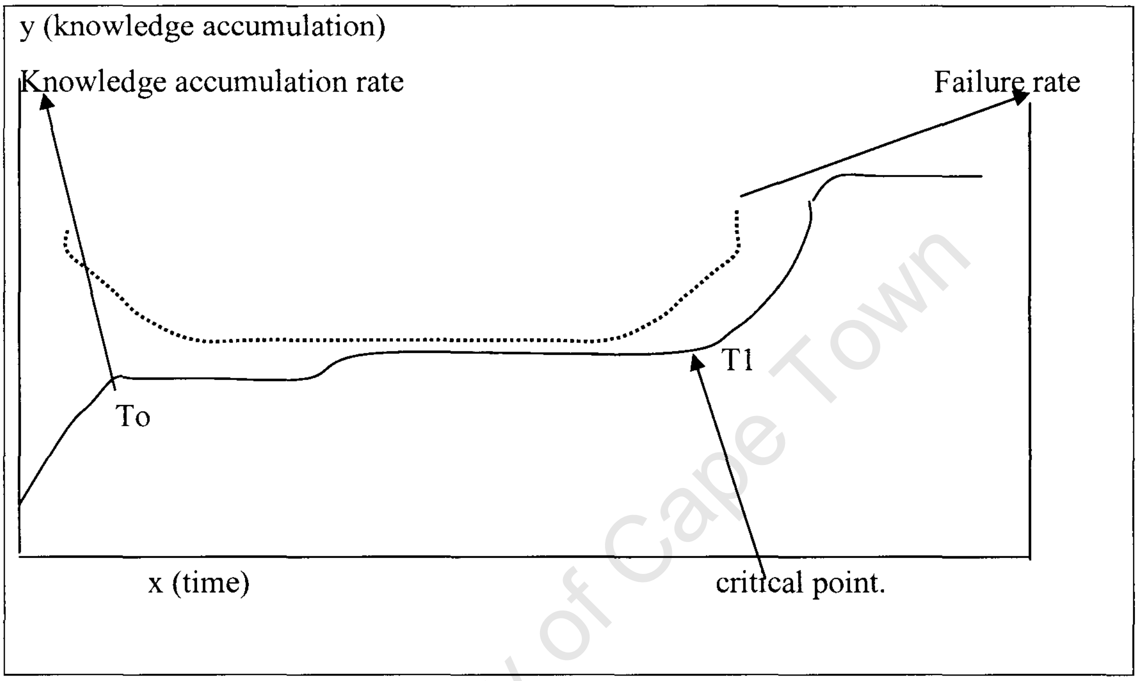


Figure 2: Knowledge accumulation, sharing and KM failure.

The continuous line represents knowledge accumulation rate and the dotted shows the failure rate.

Before the establishment of KMS, there is always some accumulated knowledge, hence the graph not starting from the origin. The first sloping part represents knowledge accumulation due to the introduction of better knowledge capturing, processing and sharing tools. As the organisation rapidly learns, knowledge sharing also increases. Eventually, the organisation reaches the peak where it has enough knowledge and relatively few problems managing and sharing it (T_o).

At time T_1, however, the organisation may start to lose knowledge due to employee moves, lack of system response to threats and challenges, or *knowledge gaps* (huge differences in knowledge levels between people in an organisation according to status [Hil02]), leading to a large difference between what an institution should know and what it does know [Zac99]. It is at this point (T_1) that better strategies should be formulated to boost knowledge accumulation, with corresponding re-focused ways of sharing. This calls for effective evaluation and restructuring of the current social constructs and corresponding technological changes. Such undertakings would require an environment that supports intelligible integration and coordination of complex social skills [Dig04].

2.8 When systems are likely to fail

From figure 2, it can also be noted that sometime before T_0 there is a high likelihood of system failure. This is because many of the implemented system components and concepts are still very new. Systems may also fail when users are not motivated to contribute. However, as time progresses, the failure possibilities diminish as shown. Eventually, other challenges arise, requiring new, better ways of knowledge management and sharing. Without this, at this critical point T_1, failure rate increases as the system responsively transforms.

In order to preserve continuity, knowledge should flow sufficiently, implying that knowledge sources must be readily available. Insufficient information and knowledge flow can be blamed for poor knowledge management qualities. Providing incentives for knowledge sharing can increase the flow but needs further research [Han00]. Monetary incentive alone is not a solution. In one case it resulted in a repository full of garbage [Han00]. It is also evident that people contribute quality items on Usenet and other forums, without monetary incentives. One potential driver may be social status or *self-esteem* [Con96, SK91].

2.9 Critical requirements of Knowledge Management Systems

From the knowledge cycle in figure 1, the following points were identified as critical to a KMS:

A) Information should be collected in a systematic way and integrated in the KM system. This includes the use of appropriate knowledge filters and acquisition methods

B) There should be flexible structures to organize knowledge in a meaningful way, capturing all the relevant attributes such as publication, origin and the creator's profile.

C) Information should be readily integrated with the existing knowledge body

D) Knowledge elements should be accessible easily. This requires common terminology (or an ontology) for the domain in consideration.

E) Knowledge elements and sources should be retrievable quickly.

F) Knowledge should be structured so that concepts can easily be interpreted

In addition to the above, we identified additional needs which we address in this thesis, namely the need for knowledge measures, for making inferences based on knowledge measures and metadata, and for automatic adaptation of the system based on this.

2.10 Summary

This chapter has discussed the foundation of this study. By outlining essential concepts such as epistemic theories, sharing, and the knowledge cycle, it has provided a basis upon which the framework and consequently the prototype have been built.

Chapter 3

Semantic Weblogs

This chapter explains why a Semantic Weblog is an appropriate tool for KM. It therefore discusses possible KMS platforms, introducing weblogs and their extended features such as RSS, aggregators and RDF. It further illustrates KM needs that weblogs do not meet.

3.1 Possible KMS Platforms

In order to build a system accommodating requirements stated in the previous chapter, three KMS platforms were considered. These were a Central Knowledge Base (CKB), peer to peer (P2P) and web based systems (WBS). The WBS is in fact seen as hybrid between the Central Knowledge Base and the P2P.

In the CKB arrangement, there is one powerful computer processing the jobs that would not be easily done on any other computer in the system. The CKB architecture is more widely used than its P2P counterpart, due to the superior file organization it exploits. It avoids the difficulties of finding exactly which computer contains particular knowledge sources. Having a common knowledge base helps to reduce redundancies and inconsistencies. Poor item ranking is also less likely to occur as it is easier to prioritize items contained in the same knowledge base.

Some of the areas the CKB gives problems with, are search criterion, stability, robustness and flexibility. The search is normally restricted to keywords, focussing on huge databases residing on the central server. In such an environment, people become skeptical to share knowledge because it decreases one's own value. The reliance on the central server for critical network management causes system failure in cases were the server malfunctions. Even in cases were an application is downloaded from the server, once that program changes, it would require adjusting all the operations that depend on it. Programs that are invoked on the server through the client normally take time to reach the point of application. Lastly, because of the many administrative procedures in CKB, members do not readily access knowledge and often would not readily contribute.

A P2P system is a computer network that consists of nodes acting as both clients and servers. In a general sense, a peer can even be a piece of software, client or any computer peripheral. As issues of sharing started becoming crucial to KM, the P2P approach started gaining ground. Peer-to-peer KM is relatively new. A P2P network is conducive for file sharing and e.g. Lotus, Napster and FastTrack have been successfully implemented [Tsu04]. The exponential increase in electronic knowledge and tools, the need for a distributed system, openness and impromptu access to knowledge material have all contributed to the popularity of P2P KM [ZR05]. Security concerns and the lack of supportive technologies are some of the major bottlenecks to P2P KM. Intellectual capital protection problems are already reported to be high [Tsu04].

3.2 Weblogs

It is evident the Web has become the supreme knowledge resource. This creates one space in which knowledge is inseparable from the community creating it, an essential quality of KM [VD04]. Cayzer and Roll suggested the use of weblogs as basis for KMS [CR05] – they avoid compatibility problems, entry barriers and unnecessary formalities. At the very basic level, a weblog typically comprises a list of snippets which can cross-reference each other.

Essentially, KM requires a non-disruptive and pervasive system; an adaptable, dynamic one endowed with rich semantics. This is where the Web-oriented system outweighs the P2P platform. Knowledge essentials, such as forums where practitioners and researchers can discuss ideas, share experiences and exchange knowledge can easily be achieved with Weblogs. A forum is also most likely to extract tacit knowledge from people's minds.

2) One feels forced to write if one has readers waiting for something new.
3) It is easier to write if one has blog-readers in mind, easier than having nobody to direct one's writing to.
4) Scholars write a lot that few people ever see.
5) It satisfies the need to give opinion and comment.
7) It is less intrusive than e-mail and easier because people don't need to manage all those addresses.
8) Weblogs are easy to read because they are dynamic, informal, short entries and in order of date.
9) They are better than discussion boards because people are not locked into a particular system as they contribute.

In comparison with the CKB, remote access to the system is easier with weblogs because of platform independence. Other advantages are usability, simplicity, flexibility and cost benefits. It is much easier to deal with complex interfaces in a web application such as html. They are easier to program, are flexible, and there is little or no cost at all in acquiring the software. Often, protocols are easier to implement and therefore, it becomes easy to attach links to sites. Weblog development tools can be very effective for developing KMS. They allow rapid system development and permit easy system customisation. They can send and receive information via email and users can easily switch between the work at hand and the KM system.

Another advantage is that, generally, the web has become very much integrated with education systems where online exercises, notes, and chatting are often conducted via this medium. Blogging can support all these and even more. Even departmental websites are in essence weblogs [EW01]. The only problem is that, they are not recognised as knowledge management systems – they are forums for sharing, but lack structured presentation of knowledge and associated KM tools. One of the risks with the web hosted KM are security and privacy, but these are also common with the other two platforms.

For the system to be easy to use and fit in with the basic working culture of researchers, a weblog promises appropriate underlying technologies [KNR04]. The main reasons include: high popularity [DE04], simplicity in appearance, and attractiveness as a result of the informal way in which contributions are organised with weblogs. Currently, there are many (see appendix D) popular weblog systems. In most of these, sharing occurs through exchanging of news summaries. The summaries are published and displayed through an export and interchange format called rich site summary (RSS) [Dow02, Col02]. Versions 0.9 and 2.0 of RSS are based on resource description framework (RDF) [RDF05]. Unlike RSS that is merely used for information transfer, RDF provides more structure with the triples (source, predicate and object) [RDF05]. With such structures, information/knowledge is well-defined, enabling computers and people to work in co-operation [LHL01]. One of the uses of RSS is the formation of RSS-feeds that outline the knowledge resources of a blog. Aggregators can then collect these feeds from different blogs and display them in one list.

3.3 RSS

RSS contains two parts: the static, enveloping information about the web page, and the dynamic, constituting the news stories or publications in the form of headlines and other textual and graphical contents. Figure 3 gives the overall structure of an RSS file. A fragment of RSS is shown in figure 4. RSS is shared through a process called syndication. Syndication is a way of posting news summaries on the web page in a chronological order of date of publication.


```
XML Declaration
Container
  Channel Description
  Image Description (optional)
  Item Description
  Item Description
  ...
  Item Description
```

Figure 3: RSS structure Sample

```
Sample code:
<?xml version="1.0" ?>
<channel>
<title>weblogs in KM education scenario</title>
<link>http://knowledge-centre.uct.cs.ac</link>
<description>This code gives an introductory example of syndication in KM education scenario.</description>
<language>XML::RSS</language>

<item>
<title>Readings: Application of RSS in KM</title>
<description>Exploring the web semantics By Chanda...</description>
<link>http:// knowledge-centre.uct.cs.ac,za/weblog/education /
archives/012.html</link> </item>
<item>
...
</channel>
</rss>
```

Figure 4: detailed RSS sample code

A brief explanation of the above code is as follows: The first part is an elaboration of the channel itself: The title of the entire weblog, a link [http://knowledge-centre.uct.ac.za] that points to the home page, and a description of the entire blog.
The points below, describe the items contained in the channel. Each item contains a title or name, a brief description explaining what the item is all about, and a link (URI) describing the location from where the full details may be extracted, i.e. a link to a specific page within the blog: [http://knowledge-centre.uct.ac.za/weblog/education]

A more elaborate example of the RSS/RDF code is shown below

```
// The name space is given as: xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#//
//The specification for the html used is given as: "xmlns=http://purl.org/rss/1.0/" and the definition of the RSS version as <?xml version="1.0" encoding="utf-8"?>//
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns=http://purl.org/rss/1.0/>
// Below is the channel component of the RSS. The title element should suggest to the user that it is a link from the web page to the feed. For example, the Knowledge-centre below should be the same title in the html appearing on the home page //
 <channel rdf:about="http:///knowledge-centre.uct.ac.za/rss.rdf">
  <title>Knowledge Centre</title>
  <description>
  Download Database notes from Knowledge-centre!
  </description>
  <link>http://knowledge-centre.uct.ac.za</link>
// The image part is option. This can be used for rendering purposes.//
  <image rdf:resource="http:///knowledge-centre.uct.ac.za/images/logo88x33.gif" />
  <textinput rdf:resource="http://knowledge-centre.uct.ac.za/search.pl" />
 </channel>

// The item part containing information from a specific source page follows.//
 <item rdf:about="http://knowledge-centre.uct.ac.za/databases/relational.html">
  <title>Relational Databases </title>
  <author> Francis Katoma</author>
  <date>2006-11-04</date>
  <publisher>Joseph</publisher>
  <quality>high</quality>
  <link>http://knowledge-centre.uct.ac.za/databases/relational.html</link>
 </item>
 <item rdf:about="http://knowledge-centre.uct.ac.za/databases/object.html">
  <title>Object Oriented Databases</title>
  <link>http://knowledge-centre.uct.ac.za/databases/object.html</link>
 </item>
</rdf:RDF>
```

Figure 5: RDF Sample Code

Characteristics of RSS/RDF appropriate for KM systems are:

- The ability to capture metadata (answering questions such as "Who wrote this", "When was this published" and "What are the topics of discussion") makes RSS very suitable, for example, in indicating the reliability of a source, the age of a publication and relevance of content.
- It can be repurposed, for example, could include capturing with an item its author's profile, comments and ratings given it by reviewers along with reviewer's names, etc.
- The common format of presenting material from different sources, both files and web pages, makes Weblogs very inclusive.
- Through RSS, an item can be posted on multiple sites making it more accessible. However, duplicates can be a menace in this case.
- Because XML is self contained, there is an opportunity to include a description of the methods required to use the data, along with the data itself. This can make the data re-usable many years from when it was deposited.

An example weblog at Harvard University appears in figure 6.

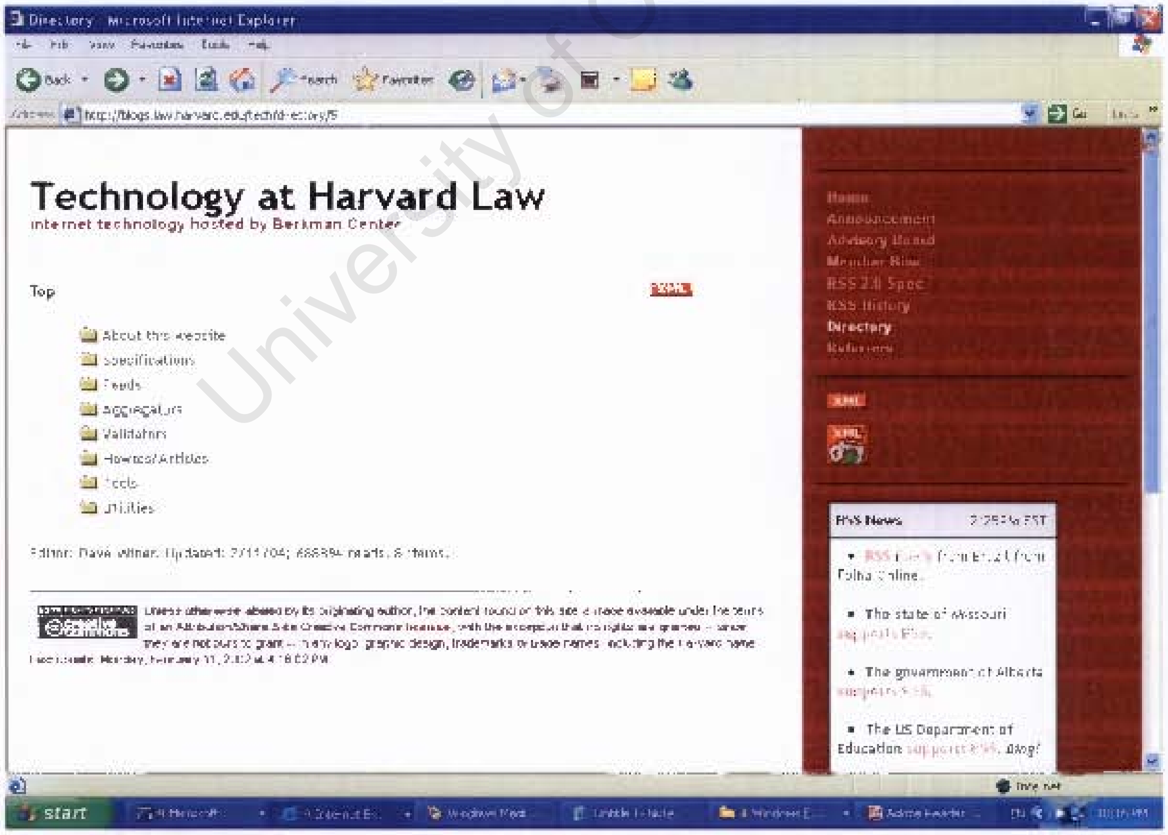


Figure 6: Sample Weblog [http://blogs.law.harvard.edu/tech]

3.4 *Aggregators*

The difference between the RSS feed and a news aggregator is that the RSS feed outlines the knowledge resource from one site whereas an aggregator collects items from different sites (weblogs) and displays them on one list for easy scan and referral.

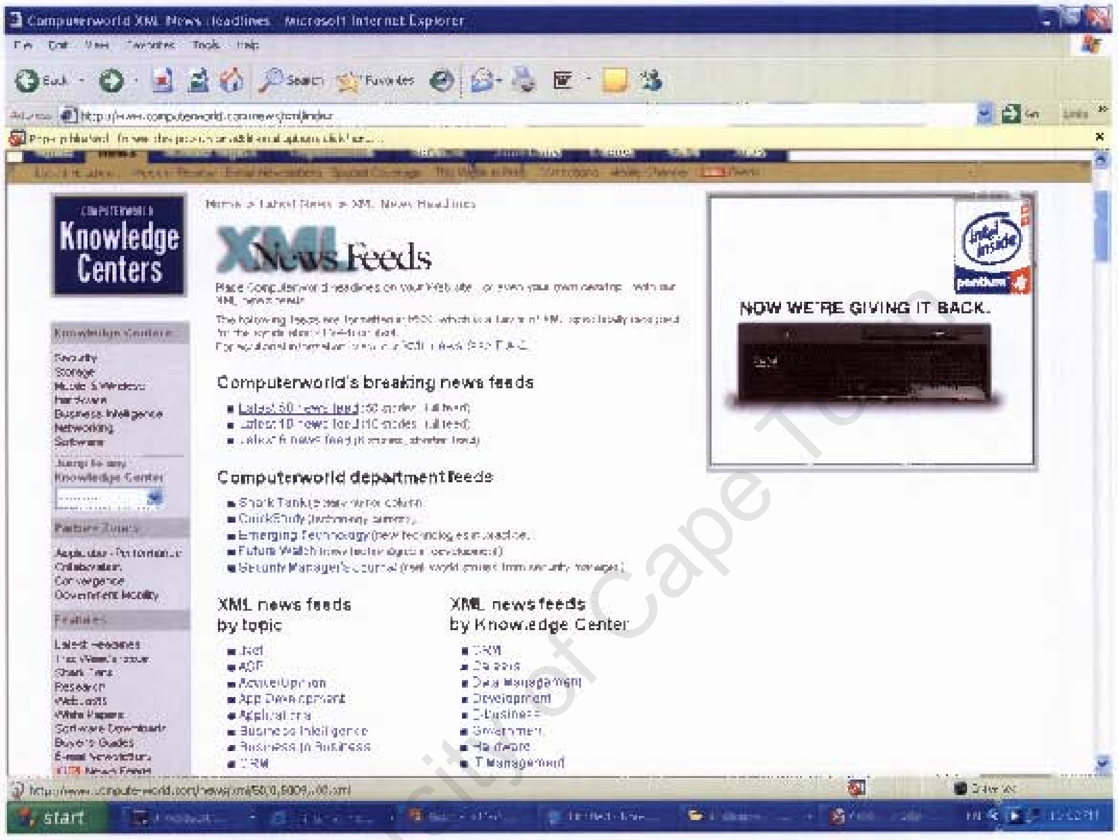


Figure 7: XML News Feed [http://www.computerworld.com]

The figure 7 above is an example of a news feed while figure 8 below is an example of a news aggregator.

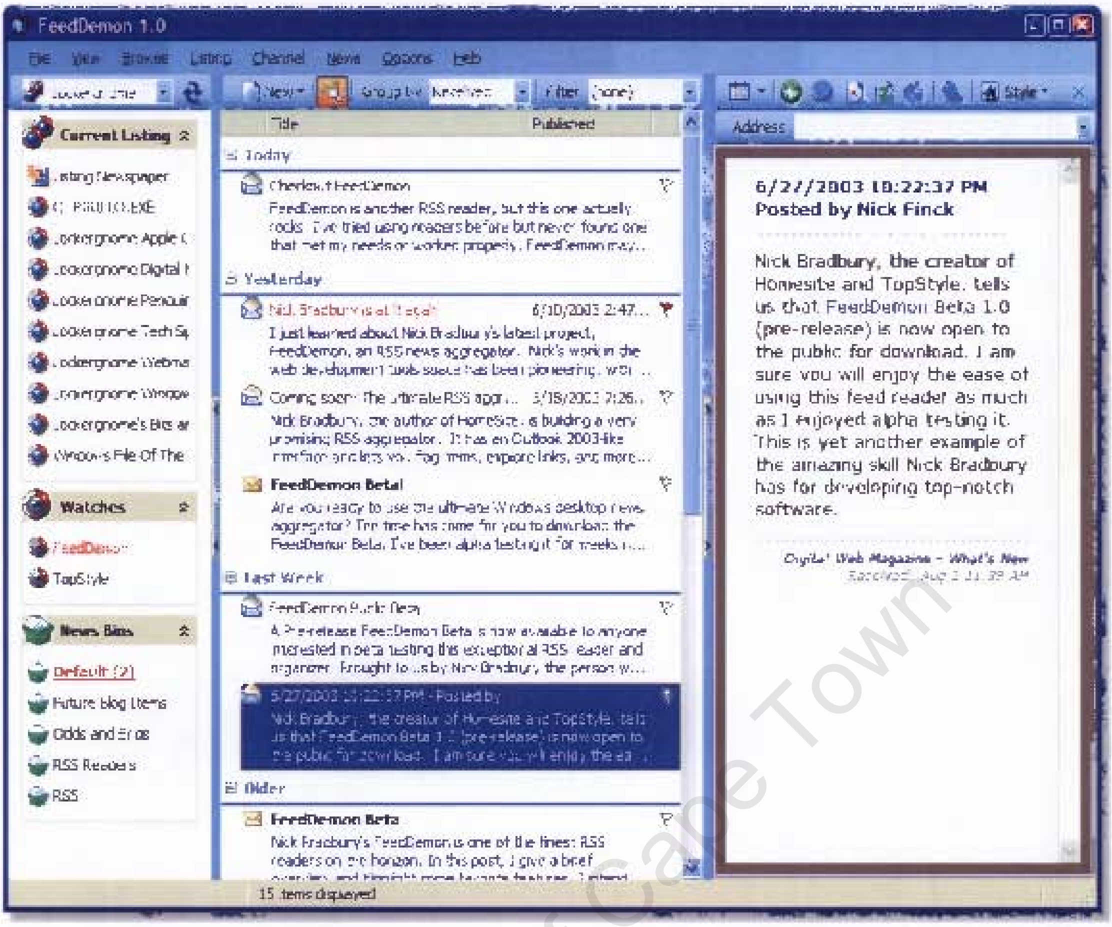


Figure 8: XML News Aggregator [http://www.computerworld.com]

Collected items are placed in folders. These items can be coming from different sources or from the same site but allocated to appropriate folders. On this page in figure 8, items can be browsed easily by following the explanations beside the folders.

One way of limiting aggregators to trusted knowledge providers is to employ epistemic theories. For example if weblog p is trusted by weblog q, then if the system trusts weblog q, it should trust p as well (*reliabilism*). Furthermore, if q gets publications from world renowned writers and society trusts that it is credible knowledge, then the system can include those publications (*consequentialism*).

3.5 RDF

Knowledge needs change according to people's interests, levels of understanding and perception. The way people look for knowledge is also dependant on where they search, how it is presented and whether it is available. According to [Lar96, Zac99, JD, PE01] information seeking is typically probabilistic and hence needs to be treated with appropriate tools. One useful way to find and assess knowledge items is by keeping metadata.

RDF introduces structures for capturing metadata. RDF consists of a *resource*, *predicate* and a *value*. A resource can be anything, for example a web page or book. A resource has some property and the property has a value that can sometimes be another resource. The predicate links the resource to the value. A complete statement is shown in figure 9 and represented as a triple (*S*, *P*, *V*). *S* represents the source, *P* the predicate (property) and *V* the value.

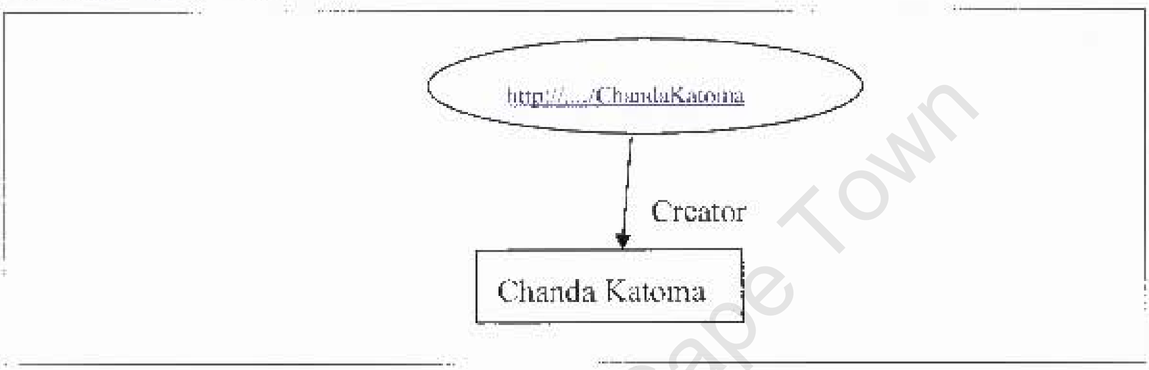


Figure 9: RDF graph

The ellipse contains the unique resource shown by the URI http://...../chandaKatoma. The property is shown by the arc, and the value Chanda Katoma in the rectangle. The value can also be a URI as shown below:

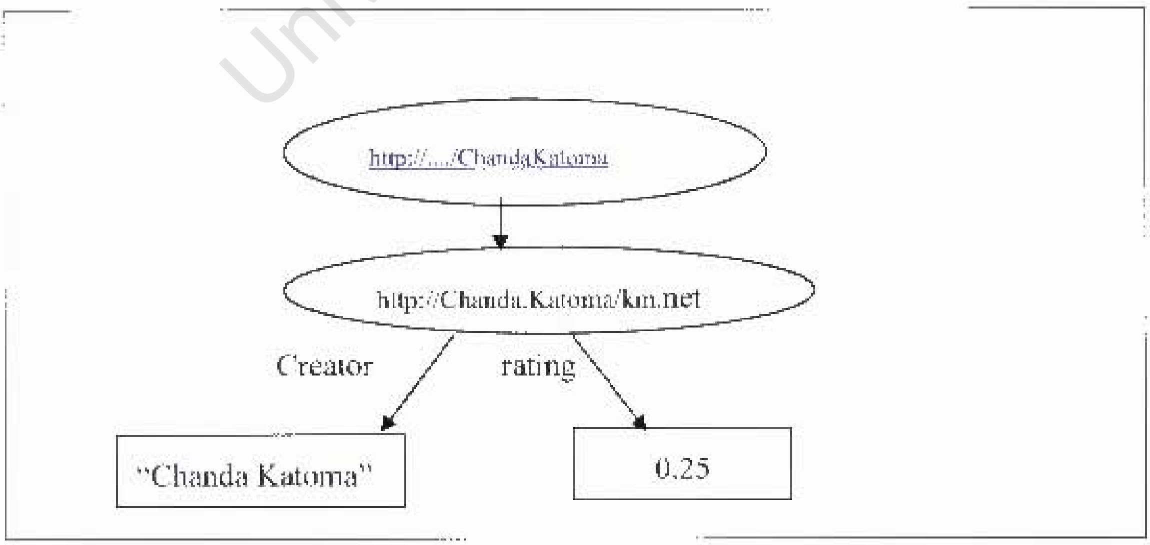


Figure 10: RDF extended graph

The graph produced through RDF contains more information than just how objects relate to each other, because of the implicit information in the attributes [W3C04]. RDF is also recommended for analysis of topic maps and for collaborative processes [BPM00].

3.6 Semantic Weblogs

Semantic blogging is a way of introducing structure and meaning to weblogs. Since most weblog construction tools support RDF generation, semantic blogging becomes a relatively easy extension to blogging. Steve Cayzer at HP has pointed out how semantic blogging has attracted extensive attention in KM related research [Cay04]. In particular, a lot of work is focused on the capturing of metadata along with snippets, and using this to enhance navigation, querying and display. Semantic Web for Advanced Development in Europe (SWAD-E) [Cay01] uses RDF [RC06] for knowledge management. Bettina and Roberto have looked at the use of the easily captured data and information from blogs to estimate domain relevance and domain consensus [BN06]. This is mostly done through the use of probability theories such as expectation. Flavious and Alexandru have looked at adapting graph visualisation techniques for the visualisation of RDF data, and show how inferences can be made from this [FTH03].

3.7 KM needs that weblogs do not meet

To improve the ability of a blog to meet the KM aims as described in chapter 2, the following properties of weblogs need to be addressed:.

1). Knowledge acquisition

Blogs easily lose focus especially if there is no strict control for example on who should contribute and what they should write.

2). Knowledge organisation

Blogs are not structured enough (typically lists sorted by date/sender/topic) and navigability can be difficult. They don't use domain ontology to structure content, they do not clearly relate items to each other and there are no icons representing summarised annotations next to entries (for example how much discussion an item has caused, how people rate it, etc).

3). Search / Knowledge retrieval

There is no search on metadata like rating by readers/ amount of discussion/etc.

4). Usability and Sharing

The system may be filled with irrelevant or poor items unless there is an incentive to only make valuable contributions (for example ratings from others, reward of marks, etc.) Group discussion is not used well enough. There should be someone to control the flow so that participants don't diverge from the topic of discussion.

3.8 Summary

This chapter has discussed different possible KM platforms and the reasons a Semantic weblog is proposed. It has also shown how KMS can be enhanced by the use of emerging technology such as RSS. These are essential as far as important components of KM such as sharing and knowledge transfer are concerned.

Chapter 4

Maximum Likelihood Estimation

This chapter focuses on maximum likelihood estimation (MLE) used for estimating parameters. It discusses sigma algebras, measure theory and their specific applications in KMS.

4.1 Sigma Algebras

RDF knowledge (in triples) does not convey much about the usefulness of the content. This is exactly when a *mathematical model* is needed. The following is a description of the mathematical concepts that the KMS framework builds on.

Informally, a set is defined as a collection of objects belonging to a particular class. The objects are called elements [Pat02]. This is very important because measures are rigorously established on this principle. Let X_1 be a knowledge item, with x_{11}, x_{12}, x_{13}....x_{1n} as its metadata, for example quality rating or number-of-hits. X_2 would be another item with similar metadata. Let X_1, X_2 ... belong to a sigma algebra X. A sigma algebra X can then be taken as a collection of sets, with x_{j1}, x_{j2} , x_{j3}......, x_{jn} as elements of set j.

So, let X be the collection of these metadata information sets (taken from RDF triples). Given a sample space $\{X, S\}$, a σ-algebra χ on a set S, is a family of subsets of *S*. By definition the empty set $\phi \in \chi$ and if $A \in \chi$, then $(S\text{-}A) \in \chi$ and if A_n is a sequence of subsets of *S*, then the union $(\cup^{n}_{i=1} A_i) \in \chi$. Thus, a sigma algebra is a Boolean algebra as these closure properties fulfil the OR, AND and NOT operations [Wik06]. Sets in the form of sigma-algebras have been used to model information [DE01] and can naturally be extended to model knowledge constructs. There are also other algebras, but σ-algebras are particularly appropriate because a probability measure is applicable to any σ-algebra, providing a computational basis for KM [Aum76].

4.2 *Measure Theory and MLE*

Measure theory is a branch of real analysis that investigates sigma algebras. Some information retrieval uses measure theories to determine the "importance weight" of a particular word in a document i.e. the ability of the associated word to single out the document from the others in the knowledge base. [PP03].

Formally, a countable additive measure (μ) is a function defined on a sigma algebra X in sample space *{ X, S }* with range $[0, \infty]$. A measure is a function that assigns a number to a set, with properties as below:

1. The empty set has measure zero
2. Countable additivity or σ-additivity: if *E1, E2, E3*….is a sequence of pairwise disjoint sets in *X*, then $\mu(\bigcup_i E_i) = \sum_i \mu(E_i)$.

The members of *X* are called measurable sets. If the range of μ is [0, 1], then μ forms a probability measure, applicable to random variables [AS06]. This is very important in KM because information seeking is generally probabilistic. A probability space (*X, S,* μ) incorporates a sample space *S*, defines a set of events of interest, the σ - algebra *X*, and its probability measure μ.

Sigma algebras are linked to Maximum Likelihood Estimation (MLE) [Tul03] as follows. Given an independent sample space $x_1, x_2 ....x_n$, in some $\{X, S\}$, with a σ -finite measure μ on X that has distribution $f(\chi, \theta)$ where θ is unknown; MLE can be used to find the most likely value of Θ. For *n* observations, $x_1, x_2 ....x_n$, the likelihood of any particular value for θ is given by

$L(\theta|\chi) \equiv p(\chi|\theta) \equiv \prod_{i=1}^{n} p(x_i|\theta)$. This holds because it is known from the multiplication rule in probability theory that the joint probability of these observations is $p(X|\theta) \equiv \prod_{i=1}^{n} p(x_i|\theta)$. The primary reason for constructing a likelihood function is that, given a sample data set, one would want to solve for the values of the parameters that make the joint probability of the data most likely or optimal. Therefore, the first step is to maximise $p\ (\chi\,|\,\theta)$ by differentiating the equation and solving for θ. A fundamental result is that, as the sample size increases, L divided by the sample size n tends to converge to a constant function. In the process, the distribution of θ becomes increasingly concentrated in the vicinity of the true population parameter θ_o.

4.3 *Example*

As a simple example, suppose a coin has probability p of yielding heads in a toss. A number of tossing or attempts N can be done and the outcomes recorded. Recordings can be made so that H represents head and T tail. Let *F(p)* represent the binomial probability function. By differentiating *F(p)* and solving for p, it estimates the most likely value of p that maximises the data. For a fair coin, heads and tails have equal (0.5) chances of occurrence. So, to check if the coin is biased, one simply compares the p with 0.5. If they are equal, then the coin is unbiased and biased otherwise. Moreover, the degree of bias can be determined.

Applying this to KM can be explained as: suppose a knowledge item is rated as helpful or unhelpful. Take N as the total number of ratings for a particular item from users where H is considered helpful and T unhelpful.

F(p) = 1 if helpful and *F(p)* =0 otherwise.

By taking the derivative of *F(p)* and solving for *p,* the most probable parameter is achieved. If the estimated value is not 0.5, the bias towards unhelpfulness or helpfulness can then be computed.

4.4 *MLE requirements*

As the example illustrated, maximum likelihood estimation or MLE is used to estimate the parameters of a probability distribution function. Let X be a sample data set drawn from some population P. In order to draw conclusions about the population P, the probability distribution function (PDF) that describes P has to be estimated (e.g. normal, binomial, gamma distributions) and the parameters of that PDF, e.g. its mean, variance, minimum, maximum, etc. The tentative model is normally proposed by simple graphical analysis of the data, some physical theory, and/or previous experience with similar data or other expert insights [CDK03]. Before using MLE for statistical inference from a given data sample, the following preparatory steps are thus required:

(I) Identify the model/probability distribution function (PDF) of P

(II) Estimate the value of the unknown parameter(s) i.e. provide initial parameter estimate(s) for them

When both the above are input to a statistical analysis package in order to apply MLE, the MLE estimation

(A) Evaluates the quality of fit (of the sample data to the PDF) and

(B) Accordingly estimates the parameter value(s) of the population.

In other words, given sample data along with (I) and (II) above, MLE evaluates how closely the data fits the given PDF, and computes better estimates of the PDF parameter(s) of the population from which the data sample was drawn.

Example 1

For a normal distribution, the probability distribution function (PDF) is:

$$f(x_1,..,x_n \mid \mu,\sigma^2)=\prod\frac{1}{\sigma\sqrt{2\pi}}[e^{-[x_i-\mu]^2 / 2\sigma^2}]\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.....(1).$$

where the $x_1, x_2.........x_n$ are the given sample values, the average is represented by μ and σ is the standard deviation. Applying the natural logarithm, gives

$$\ln f=-\frac{1}{2}n\ln(2\pi)-n\ln\sigma-[\sum(x_i-\mu)^2/2\sigma^2]\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

and differentiating with respect to μ and equating to 0, leads to: $\partial(\ln f) / \partial\mu = \sum(xi-\mu) / \sigma^2 = 0$. Solving for μ, the results are $\mu` = \sum x_i / n$ and plugging in the sample values, the estimated population parameter is computed.

For the standard deviation, we differentiate (1) with respect to σ and get $\sigma=\sqrt{\sum(xi-u`)^2 / n}$.

Example 2

The PDF for a binomial function is given by $f=p^k(1-p)^{n-k}$ …………………..(1)

were p is the probability of obtaining a certain event, n is the number of events considered and k the number of times p occurred. Taking the natural logarithm, we have $\ln f = k \ln p + (n-k)\ln(1-p)$ ....(2)

Differentiating (2), gives $\partial \ln p \big/ \partial p = k/p + (n-k)\, 1\big/1-p$ ....(3)

Equating (3) to 0, $\partial \ln p \big/ \partial p = 0$ and solving for p we get, $p = \dfrac{k}{n}$. In principle, the proportionality gives the population average.

4.5 Sigma algebras and MLE in a Knowledge Management System

Let X_1 be a knowledge object with associated metadata { x_{11}, x_{12}, x_{13}....x_{1n} } where an x_i is for example its quality rating, usefulness, status, popularity, etc. We can view a collection of such sets X_1, X_2 ... as a sigma algebra X if we map their range onto the same type e.g. Boolean or a rating scale of 1 to 5. We may have a single sigma-algebra for the whole KMS, or we may have a σ-algebra for all the contributions authored by a specific individual, etc. This is a collection of sets, with x_{j1}, x_{j2} , x_{j3}......, x_{jn} being elements of set j. These measure values (x_{ji}) can be drawn from the RDF metadata captured in the KMS. Using a Maximum Likelihood Estimator on such a σ-algebra allows a system to infer the true quality (usefulness, rating, status, popularity, etc.) of the associated items.

Selecting the metadata variables $x_1, x_2.........x_n$ appropriately is critical for accurate inference. To measure for instance usefulness of a group of items X, the program collects all the values from their usefulness metadata and forms a set. These would then be fed into the mathematical model and the MLE computed. If X comprises e.g. all items contributed by a particular person, then we can infer the true usefulness of that person. Such derived information can then be used to reorganise the site automatically, e.g. so that the most useful / popular / etc. objects appear above others.

A collection of data for investigation or σ-algebra could be expressed as a matrix of measurements **X** so that:

$\mathbf{X}=\begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Np} \end{pmatrix}$ where $\mathbf{x}_{ij}$ is the score on the j[th] measure value for the i[th] item. For example, on the x_1 entity, one could have x_{11} representing usefulness, x_{12} status, and so on. If the time element is considered, a Markov Chain estimator (stochastic process) can be used, providing different measures on the knowledge items. However, this is not done in this work and is left for future study.

4.6 Advantages of MLE

1. As an approach to parameter estimation, it achieves a better approximation than other estimators as the sample size grows (Cramer-Rao lower bound) [Wik06].
2. For independent observations, the maximum likelihood estimator follows an asymptotic normal distribution [Wilk06].
3. Because MLE results are single values (convergence points) they can be used with confidence bounds in hypothesis tests.
4. There is a lot of statistical software that provides algorithms for maximum likelihood estimation for most of the common distributions.
5. By establishing convergence values if and when they occur, *threshold* points can be set that people can over time use as *benchmarks*. If parameters are performance values MLE estimates of standard deviation can be used to set upper and lower bounds (for example, the number of contributions or the rating of items in a specified time can be expected to fall in a specified range).

4.7 Disadvantages of MLE

Maximum likelihood equations have to be specifically worked out for a given distribution and estimation problem and there is thus potential for errors. Starting values are very important and MLE can be sensitive to the choice of these, and omitting information (variables) can make the MLE very inconsistent. In addition, MLE can be heavily biased for small samples. In certain instances, even the optimality property may not apply. Nevertheless the principle of maximum likelihood still holds. For the vast majority of cases, MLE is a simple and useful estimator.

4.8 Summary

This chapter has looked at σ-algebras and described how measure theory can be applied to them. It discussed how the mathematical model can use values captured in knowledge attributes like usefulness to estimate the true quality of knowledge items and sources.

Chapter 5
A Knowledge Management Framework

Chapter five describes the KM framework. It first discusses research training and introduces the data model. The system overview is provided as well as the use cases for the framework.

5.1 Research training

Research training at postgraduate level must impart a variety of knowledge and skills. Some of these are domain-specific and others general. Examples of the latter are reading, writing, summarizing, arguing, criticizing, finding and organizing information. Students require tools and platforms to help them accomplish these requirements effectively. Didactic style, education strategy, technology development, appropriate selection of content, and the personal qualities of teachers and students should be considered when building these tools [Sla96, Hes96].

This framework for knowledge management tools is based on the ideas of knowledge management and the knowledge requirements presented earlier. The system framework seeks to integrate skills and address the aforementioned issues in learning to do research. Hence, it should be easy to use, and offer some added value to make its usage worthwhile. These include finding relevant information easily, organizing work effectively, *providing space* for reading, writing and critical analysis, and the ability to measure or evaluate knowledge items.

5.2 The data model

The framework was built upon a simple but powerful data model. "The main objects in the data model are the *source* (from which the knowledge emanated, typically a human or a published work), the *snippet* (for instance, a posting or message, usually one fact or idea) and the *file* (representing either the creator's own work or a reference they found)" [BK06]. Each of these can be specialized depending on how the framework is realized in a particular system. For example an implementation of the framework may distinguish snippets posted as knowledge items from snippets posted in discussion groups or on bulletin boards.

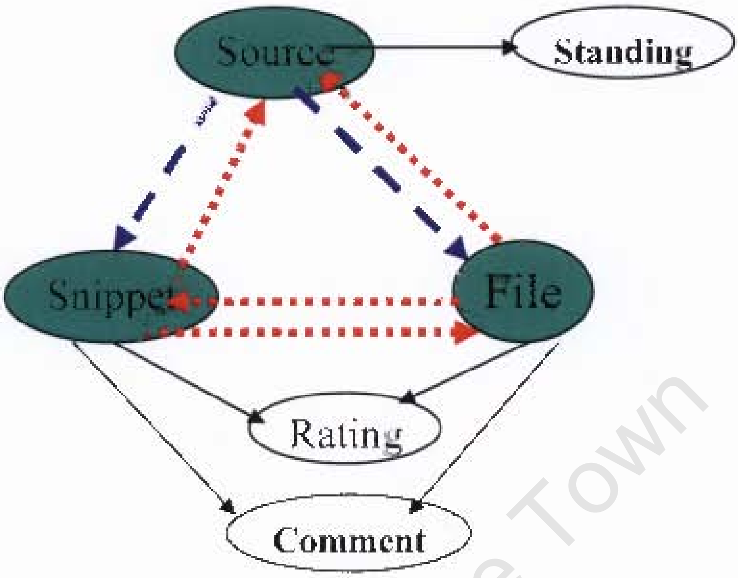


Figure 11: Data model

Objects in the model are associated with each other via directional links. Secondary objects such as comments and ratings can be related to any source, snippet or file. Snippets can reference any other object in the model - for example, when a message is posted about some paper (file) that was particularly useful, or about some other snippet that it is refuting or elaborating on [BK06].

Each of the three main types of object can be associated with secondary objects that add value to that object – *user feedback* and *system metadata*. User feedback comprises quantitative evaluations (ratings on a rating scale) and qualitative comments made by readers. System metadata includes the individual who entered the item into the system, the date on which this was done, the number of reads and links to that item, etc. In this thesis, primary objects are called items (i.e. snippets/postings, sources or files) and their secondary objects (i.e. user feedback and system metadata) are called knowledge properties. The model constrains the use of knowledge properties and the items they can be applied to according to the context. The properties are useful in computing knowledge quality measures. "In general, an item of the highest quality is one that is highly rated by users; is very often read, referenced and commented upon; and comes from a source of a high standing whose contributions are consistently highly rated" [BK06]. Knowledge properties can also be used to query the system in new ways, and for inferring how best to adapt the system, as described later.

5.3 The framework

The framework is an 8-step process, as outlined below:

1. Conduct an *audit* to evaluate the current system and needs of users.
2. Use a *weblog construction* toolkit to create a simple weblog based on the data model where snippets and files can be saved, shared and searched. Ensure that metadata including Dublin Core [KW01] is captured from the outset.
3. *Refine its interface* in an iterative manner, based on feedback from users
4. Extend the system to capture commentaries and *ratings* of snippets, files and any other knowledge sources
5. When necessary, introduce *incentives* to encourage extensive use of the system.
6. Compute *quality measures* and use simple inference to evaluate different parts of the system
7. Make the KMS an *adaptive system* that uses the results of the inference processes to extend and reorganize the knowledge structures appropriately
8. *Evaluate* the adaptation mechanisms along with the other features, and iterate again [BK06]

To fulfil the need for continued evaluation and hence system adaptation, quality measures based on knowledge properties is encouraged. Collecting user feedback such as item ratings can be a motivation especially to users who want to evaluate what they read or would like to feel valued according to their quality of contribution. Commentaries on publications give the perception that users add to what they read. Consequently, contributors are included as part of the quality control process. Apart from enriching the knowledge base, this improves accountability, security and transparency.

In academia, one aspect of easily *incentivising students* is to let students accumulate a few marks fairly easily from using a KMS. Some percentage of student's mark can assess how well they used the KMS for instance, how many snippets, comments and ratings they contributed, and the quality of these.

5.4 *System Overview*

An overview below illustrates how the system process can be realised.

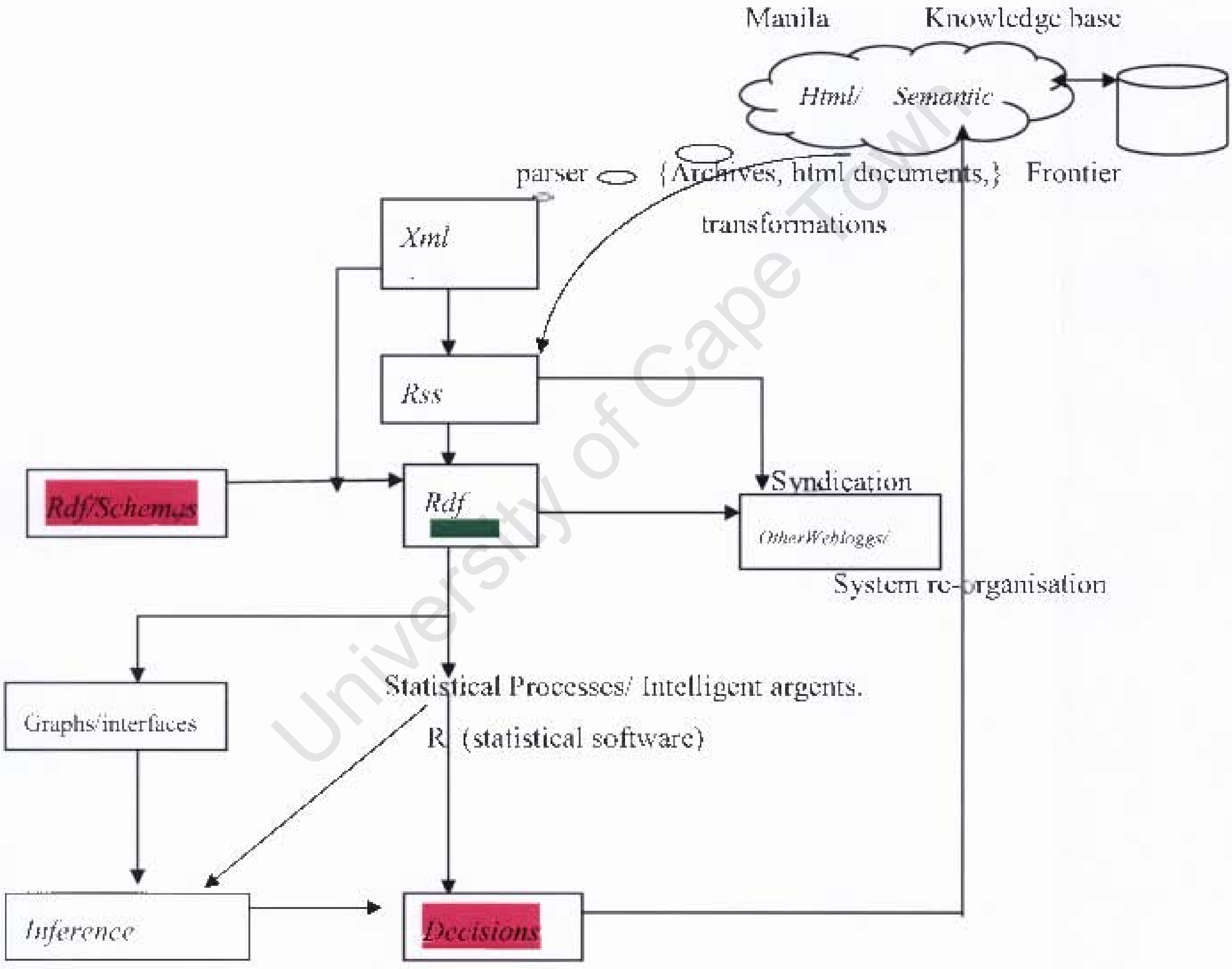


Figure 12: System process Overview

Firstly, the semantic weblog is constructed. By providing suitable RDF schemas, RSS can be extended to include more metadata viz. numerical values and Booleans that can then be used for quality measures and statistical computations. From this, inferences can be made and the system reorganised accordingly.

5.5 *System Use Cases*

Based on KMS requirements as outlined in chapter 2 the main use cases for the framework were formulated. The use cases below illustrate briefly the types of user and how they interact with the system.

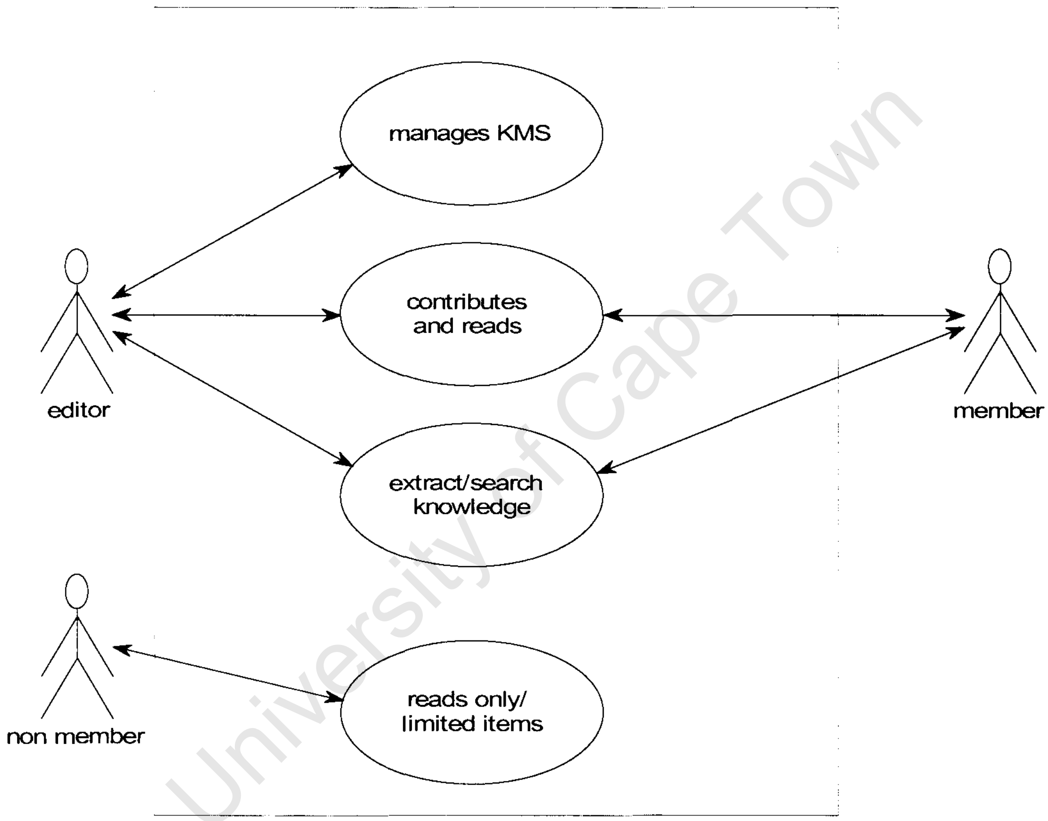


Figure 13: use case

Editor:
1. Manages KMS
2. Edits any postings on the system regardless of who posted it.
3. Selects quality published works and puts them in appropriate places on the system for easy access by readers.
4. Searches for knowledge material for users if asked to do so by members.
5. Regularly reads the incoming news from aggregators and subsequently saves important work in appropriate places.
Member
1. Contributes papers, postings, comments, discussion, topics and critiques. They can also edit their own work.
2. Does evaluations by rating knowledge material on the system.
3. Searches the system for needed knowledge.
Non Member
1. Reads only some public knowledge resources on the system.

Table 1: Use Cases

5.6 *Use Cases for the KMS Editor*

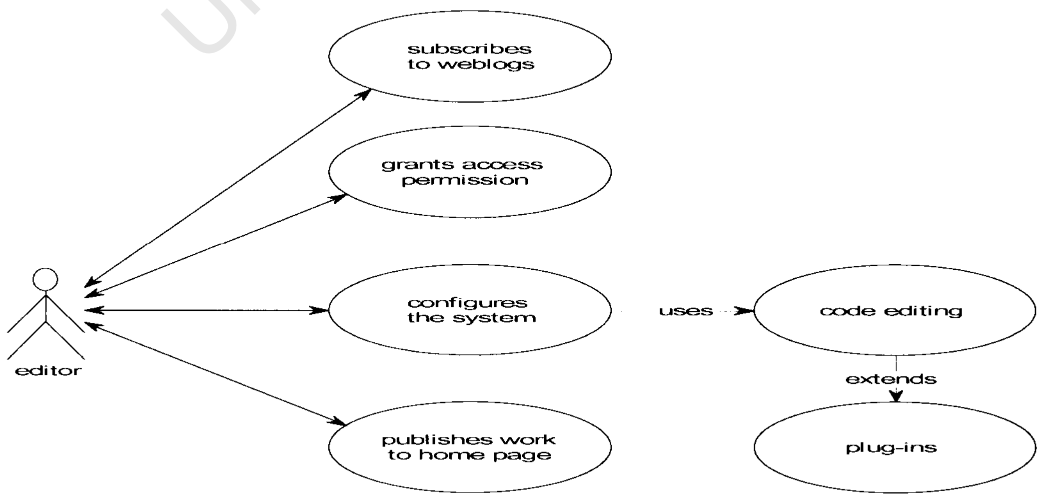


Figure 14: Manage KMS explored

Editor:
1. Subscribes to other weblogs by way of syndication so that external knowledge items can come in as syndicated news summaries.
2. Gives access privileges to members.
3. Sends items for syndication elsewhere
4. Configures the system for better system usage.
5. Extends the system as new functionalities becomes possible.
6. Forms groups and establishes links through discovered interests of the participants.
7. Sends news bulletins to members.

Table 2: Use Case 2 – the Editor

5.7 Summary

This chapter has discussed the composition of the envisaged framework and the details of how users should interact with the system once completed

Chapter 6
Prototype KMS Design

This chapter outlines the design of a Knowledge Management System called KC, the *Knowledge Centre*, a prototype KMS for research students based on the framework. It outlines an audit of post-graduate students to discover current practices and needs. This showed that, while several structures were in place as a result of research laboratory activities, participants desired a better knowledge management system, although several said they would only adopt this if it was sufficiently quick to use and provided noticeable benefits. This chapter then discusses the prototype system organisation, and the object lifecycle in KC, ending with a description of the knowledge properties that are captured.

6.1 Knowledge Audit

A survey was carried out to identify the difficulties postgraduate students were facing in seeking and managing knowledge. This was followed by interviews, focused on what students thought could be the remedy to the problems. An assessment of the existing knowledge strategies was also conducted and a simple feasibility study done to determine if a KMS would be used.

Through the feasibility study, it was realized that in most postgraduate education, there was a need for KM software especially since it is becoming increasingly difficult to do research without collaboration. In the computer science department, research groups were in place with basic websites often used for notices and advertising. However, certain individuals worked in isolation.

As with many other Universities, there was no specialized software that can be called a KMS. At undergraduate level, however, a lot of effort has been put into organizing repositories for course material, submission of assignments, commenting and displaying of results etc. on web based software which is most successful. Students were asked how they searched, stored, shared and disseminated knowledge. It was realized that, like industry, universities are also looking for KM software.

A questionnaire was formulated, based mostly on the KM auditing questions and ideas from epistemic theories (See Appendix A). The interviews engaged with the participants to obtain a feel for their working environment and ensure a thorough investigation had been done.

The results of the investigation are summarized as follows:

1. The web as the major source of knowledge			2. Problem in seeking knowledge				
Yes		Others	Yes	No			
90%		10%	92%	8%			
3. Second Sources of Knowledge Apart from the web			4. Willingness to share knowledge			5. In support of KM System	
e-mail	Books	Others (discussion)	Yes	No		Yes	No
18%	63%	19%	63%	33%		81%	18%
6. Could still do with the current system			7. Access to group knowledge			8. Any order in data storage	
Yes	No		Yes	No	Partially	Yes	No
9%	90%		63%	36%	1%	63%	36%
9. Already sharing his/her knowledge			10. Need software to structure knowledge				
Yes	No		Yes	No	Not sure		
27%	72%		90%	9%	1%		
11.What communities to share knowledge with			12. Usually got what they were looking for.				
Any	Some	None	Yes	Partially	Others		
63%	36%	1%	8%	92%	0%		

Figure 15: Analysis: shows the results of the survey:

6.2 Audit Analysis

According to the above results, there was a high dependence on the Internet for material. Still, paper based publications are used considerably. About 63% of the readers who use the web also indicated they get knowledge through journals and books. From interviews, this can be attributed to among other facts, order of presentation of content and source confidence. Another reason given was that some of the knowledge material uploaded on the web is not peer reviewed. About 27% indicated they used other sources, particularly lecture notes and e-mails.

6.2.1 Knowledge search

More than 92% indicated that they experienced problems searching for material. Some felt that it was even difficult to know where to start unless one asked people who know. Others raised the following dissatisfaction:

1. It takes too long to get to the required knowledge and sometimes one would have even forgotten what they were looking for.
2. Searching the web brings about too many results, and it takes time to sift through material for relevance and in some cases one is not even sure about its credibility.
3. Search results are not ordered by author, and regrettably, some of the materials don't even have dates.
4. Reading books is not only time consuming and expensive but it is also hard to find the better ones.
5. Some online research publications charge for access.

Over 95% of the respondents clearly indicated they only partially get what they needed through the Internet and paper-based publications even though they earlier indicated that these are their main sources of knowledge. Consequently, about 81% suggested that a better tool to manage knowledge was needed. This seems to confirm the earlier assertion that education systems, just like the corporate world, are in need of better KMS.

6.2.2 Knowledge Accessibility and sharing

63% acknowledged having access to shared group work. In further enquiries on how the current systems worked, some other knowledge sources were:
1) group meetings to discuss research problems. 2) publications accessible through shared files. 3) talks and colloquiums. 4) news groups and online forums. 5) discussions with supervisors.
Clearly, all these ways of knowledge sharing are traditional as they do not seem to meet the requirements of knowledge theories. They can in fact be viewed as prerequisite to better KM. As stated, earlier on, the vastness of knowledge today, coupled with the swift changes in KM strategies, much more advanced ways of sharing and managing knowledge are required.

6.2.3 Knowledge Restructuring

The survey showed also that about 63% of the participants have their work stored in unstructured and unorganized ways. Earlier findings in literature illustrated that about 80% of organizational knowledge is in unstructured form. It thus seems that universities contribute substantially to that problem.

The other problems the respondents faced with traditional arrangements were:
1) Knowledge items are not usually stored in a logical order within directories.
2) People do not know where to document what's relevant or are lazy to do so.
3) Much of the knowledge material both on servers and hard drives is not indexed, therefore, searching for specific papers or unpublished work is tedious.

Further enquiries through interviews on how people managed their work gave the following information:
1) A few stored knowledge items in hierarchical structures usually in files and directories but often not sharable. 2). some stored knowledge according to date and or topics.
3) Some stored knowledge in a manner that only their supervisors could share it but in most cases only the individual had access himself.

Only 27% indicated they were unreservedly sharing their knowledge.
Some of the answers given as to why people were not sharing are:
1) simply because no one was asking them; even though, that could mean people did not know they had the knowledge. 2). There was no proper forum to do so 3). absence of proper tools to structure knowledge so that it can become easily sharable. 4). some gave privacy and security reasons as the only impediment to sharing. 5). 63% were willing to share knowledge with any one, while 36% were eager but only with a class of people.

The three first reasons go to prove that, in spite of the research group arrangements and other forums established for knowledge sharing, there seems still to be a very big gap between what students want and what they actually get. There was also a lack of well-defined ways of describing what is in the knowledge item before one can read all of its content.

It was also interesting to know the kind of KMS students anticipated and the suggestions below were captured:
1) 50% of the respondents could only embrace the software if it was quick to use.
2) 8% were only interested if it was easy to use.
3) Some suggested however, that they could only make use of the system if it was able to fit in with the way they did their work and with the software tools they were using.
4) 33% would use the software any how as long as it would provide what they wanted.
Usability and security are some of the issues raised by the interviewees. Even though these do not appear as critical success factors and in the knowledge audit trail, they are very important and should be considered.

6.2.4 Audit Results Summary

It has been observed that there is not much difference between the general KM problems the world is experiencing and what is faced by institutions of learning. This to some extent shows that, KM systems generally require similar strategies. Despite the availability of powerful computers and networks, there is still a lack of effective KM tools to leverage knowledge. It was learnt also that effective knowledge reorganization and representation strategies are vital to sustainable education KM.

Some of the common problems research students experience are: difficulties to find relevant sources (such as books and articles), a feeling of isolation at times caused by lack of feedback, uncertainty about the direction of their efforts, and unwillingness to write up and or present their work without clear benefits. Time is wasted on poor ideas simply because these were not shared with those who know better. Even though research laboratories may host seminars and paper archives, students are often reluctant to utilise these.

Other hindrances could however, be attributed to cultural issues such as the will to share and contribute. It is also clear that islands of knowledge exist in research group databases and websites but the absence of a coordinating system makes it difficult to discover them. Hence there is need for a well defined knowledge management system.

6.3 KC organization

The KC system is made up of four basic types of page.

I. *Introduction* : pages that introduce the site and the institution
II. Search : pages used for querying and searching content
III. *Browsing*: pages used for navigating the content
IV. *Contribution*: pages for contributing one's own writing and/or files

The Introduction pages are: a "Getting Started" page (which briefly describes the site, its purpose and organization); a "Departments" page (which describes each of the main sections into which the institution is divided) and a "Subscriptions" page (which lists the sites from which RSS feeds are obtained). These pages are not intended for regular use, and are mainly targeted at newcomers.

The search page permits searching all or part of the site, and also offers the ability to search relevant sites elsewhere on the Web (these are maintained by the site's editor and should include the most useful online sites, bibliographies etc. for the institution).

The Browsing pages include a File Manager page (which presents the hierarchy in which files uploaded to KC have been stored), a Bulletins page (which comprises notices from the editor); a PublishedItem page (items that are in the public domain are shown in reverse-chronological order here) and an Aggregator page (items from other sites in reverse-chronological order, each comprising a summary and link).

There are two types of contribution page. The first is a Discussion page, with postings arranged by topic or by contributor, where members can add new items to any discussion they are reading. The other is the Unpublished Work page, where any posting is composed and kept until its author chooses to publish it - to Published Items and/or the File Manager.

Each page has the same format and layout which includes a navigation panel on the left. The Published Items and Discussion pages have a calendar in the top right corner. It is unnecessary to make Bulletins the home page since these are emailed to members anyway and should not change often. In fact, the home page is customizable, in that it can be the File manager, the Discussion or the Published Items page, depending on user preference.

Regular users will mostly look at their (personalized) homepage, i.e. they will start with the discussions (to check the "hot topics"), the published postings (to see the latest items posted anywhere relevant), or the file manager (to retrieve something they have read or stored previously on this system). They will typically switch between these three main pages, reading and adding new contributions as the need arises. Thus the site can be seen mainly as a knowledge repository, where users save files in their own personal Directory folders arranged in simple visual hierarchies and check files saved there by others, or in specific subject categories. At the same time, it can be seen as a forum for sharing of personal ideas (discussions and postings) and of files (in the directory). Screen shots of sample KC pages can be seen in Appendix B.

6.4 *KC object life cycle*

The object life cycle describes how creation and manipulation of the knowledge items proceeds.

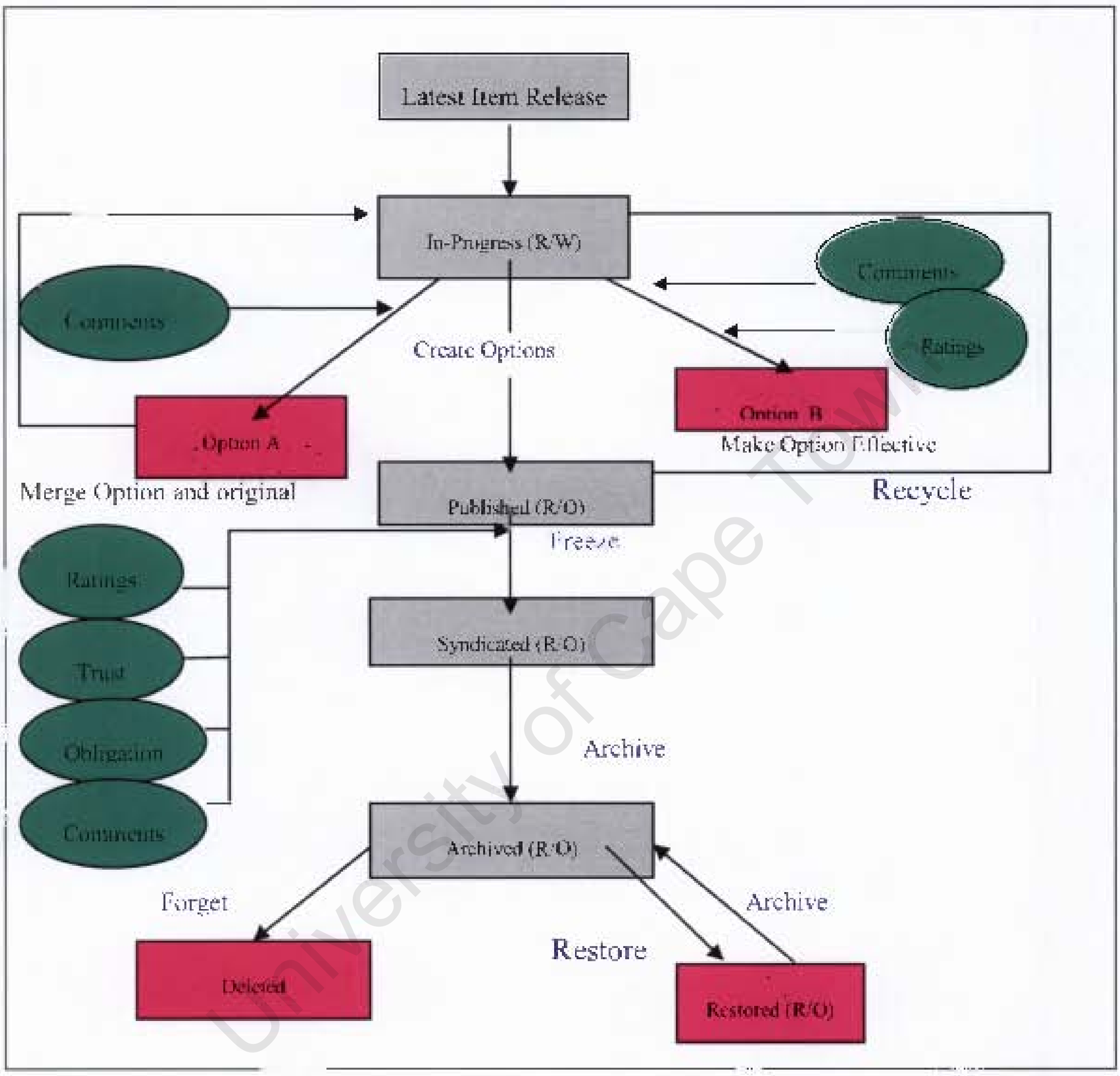


Figure 16: File (item) cycle

When an Item is created it becomes a work in progress. Work in-progress can be updated and alternative versions saved. A file remains a work-in progress until released by its author. Then, if accepted by the editor, it is added to Published Work. A published work is effectively a pre-release suitable for testing before external release (syndication). Before an item is syndicated, quality measures are collected for it. Eventually, released items are archived or deleted (archived items can be restored).

6.5 *Object progress through KC*

The figure 17 below illustrates the different sides of KC and how objects such as files can progress from when they are introduced into the system to their final destination.

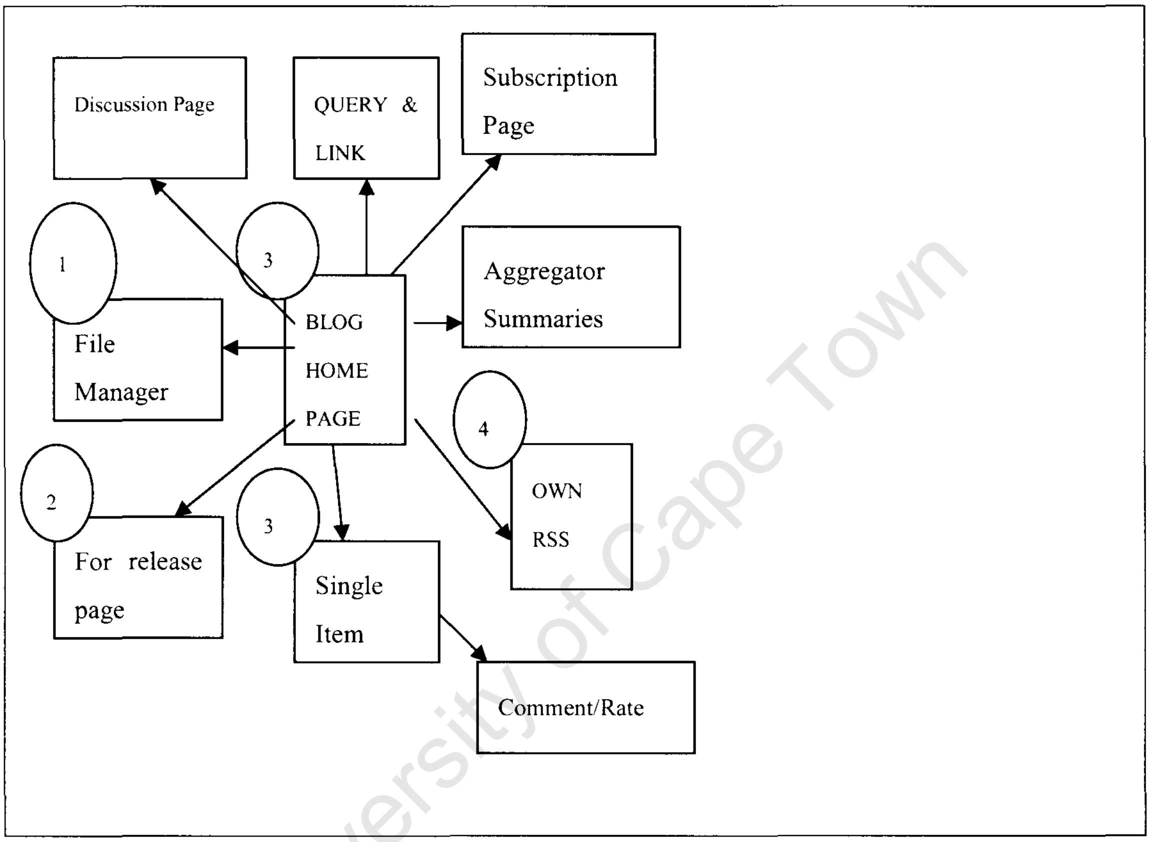


Figure 17: System Structure [BK07]

Numbers show how work items progress from private pages of the file manager to publication on the blog and externally.

Given that the KMS straddles technology, psychology, inference, etc. it is necessary to keep a thorough definition of all primary items. The same metadata must be kept for all items on the same kind of page.

6.6 *Knowledge properties*

The measure of quality in an adaptive system is a continuous process. Based on this, a system can be used to provide information about the current status and predict future trends. With such knowledge, the system can be reorganised appropriately. The knowledge properties implemented in KC comprise both system metadata and user feedback measures. The system automatically collects the following metadata for each KC item (source, file, snippet/posting): author, entry date, number of reads, number of links to it. More metadata can optionally be added by members: topic, certainty value, originator (from whom the file was obtained) and reference. The editor is also able to enter the status (role or job) of each contributor, such as lecturer or student. Three types of quality feedback measure can be supplied: The *usefulness* of a knowledge item (snippet or file) is a Boolean scale where readers rate the object as either useful or not useful. The *rating* of an item is selected from a scale of 1 to 5 (e.g. poor to excellent). Lastly, a comment is a textual note on the item, which is not interpreted by the system but which is used as a measure of interest in the item e.g. in identifying "hot topics". The system keeps a record of who gave the feedback as well.

From these knowledge properties a number of additional quality measures can be derived. For example, the popularity of an item can be computed from the number of times it has been read, commented upon, rated or linked to; the quality of an item can be derived from certainty values, usefulness and rating values given to it; similar-interest groups of people can be identified based on what they write, read, comment or rate; etc.

6.7 Summary

This chapter has discussed the main features of the prototype. These include KC organisation, the file object process as it moves in the system, and knowledge properties (metadata) that is kept. It has also given an overview of the KC pages.

Chapter 7
The Manila Content Management System

Chapter seven discusses the Frontier database system. It then outlines how web pages can be easily developed using this data base. Manila software is introduced as a basis for the KMS development. A brief account of Frontier tables in the development of KC is also discussed.

7.1 The Frontier Database System

The Manila Content Management system [Fro] was used to build the prototype KMS. Manila uses the Frontier database system [Fro] as its back-end. In Frontier, everything is kept in tables, even scripts, which are pieces of code that interact and manipulate the database. Tables can be nested and are arranged in hierarchies. A brief account of the top-level tables is as given below:

- Scratchpad: is a temporary storage area used for developmental work.
- Suites: are related scripts and data that have a common purpose, for example RSS, or for receiving e-mails and storing them in the database.
- System: keeps the agents that run in the background, system scripts, and template scripts that users can extend. Agents are useful for triggering important operations, for example (system.agents.schedulerMonitor) calls the scheduler, which manages scheduled tasks.
- User: is where development of a specific system occurs. This table stores the scripts and data. The data is in fact stored independently of the script and if there are any changes in the structure of the script, the data remains intact.
- The workspace: is designed for testing by developers but, unlike the scratchpad, can store work permanently.

Every object must have a name, value (or content) and data type (kind), so all tables have three columns: Name, Value, Kind, as shown in figure 18. Their rows can be any mix of item kinds. To name but a few kinds, there are Boolean, character, number, float, date, direction, string, enumerator, file specifier, alias, object specifier, address, table, WP-text, picture, outline, script, menu bar, list, record and binary.

Managing XML-structured information in Frontier is easy. A script that would read an XML file, compile it into a database structure and then open it in a new window is shown below:

```
Local (f = " C:\\xml item\\myKnowledge.xml" )
Local  (adrtable =
@xml.item.myKnowledge.tableStructure)
xml.compile (file.readWholeFile (f) , adrtable)
edit (adrtable)
```

If myKnowledge.xml contained:

```
<?XML VERSION= "1.0"?>
<myKnowledge username=" Victor Katoma">
        <research>KMS</research>
        <interest>quality measure</interest>
        <address>room 300</address>
        <university>UCT</university>
        </myKnowledge>
```

then its content would appear in the Frontier table as shown below:

Name	Value	Kind
>myKnowledge	5 items	Table
>/atts	1 item	Table
>username	Victor Katoma	String
>research	KMS	String[3]
>interests	Quality	String[7]
>address	Room 300	String[8]
>university	UCT	String[3]

Figure 18: XML in Frontier

The language used in Frontier scripts is called UserTalk [Fru]. Here, any object can have a value of any type regardless of what it initially had (not strongly typed language). Scripts can work together and interact in the database. This way, a complex but well structured program can be achieved. To access an object, one has to specify the full path, e.g.: ((@["C:\\Applications\\Frontier\\GuestDatabases\\apps\\nameOfitem.root"].nameOfitem.respond). In short one can write @[f].KC.msg. Such referencing of objects discourages changing of folder names or the database location on the disk.

7.2 *Developing web pages*

Manila is built on top of Frontier as an interactive tool for website management. For the user, it enables the use of WYSIWYG for simple tasks and directives such as sending e-mails and syndication instead of having to write HTML. On the developer's side, more complex HTML and Java scripts can then, via the #prefs menu that links Manila and Frontier, be added to these saved pages.

Web pages can be developed rapidly simply by using the index object. In the figure below, one simply creates the kmresearch table under the root table. The kmresearch table would contain the index object as show in figure 19.

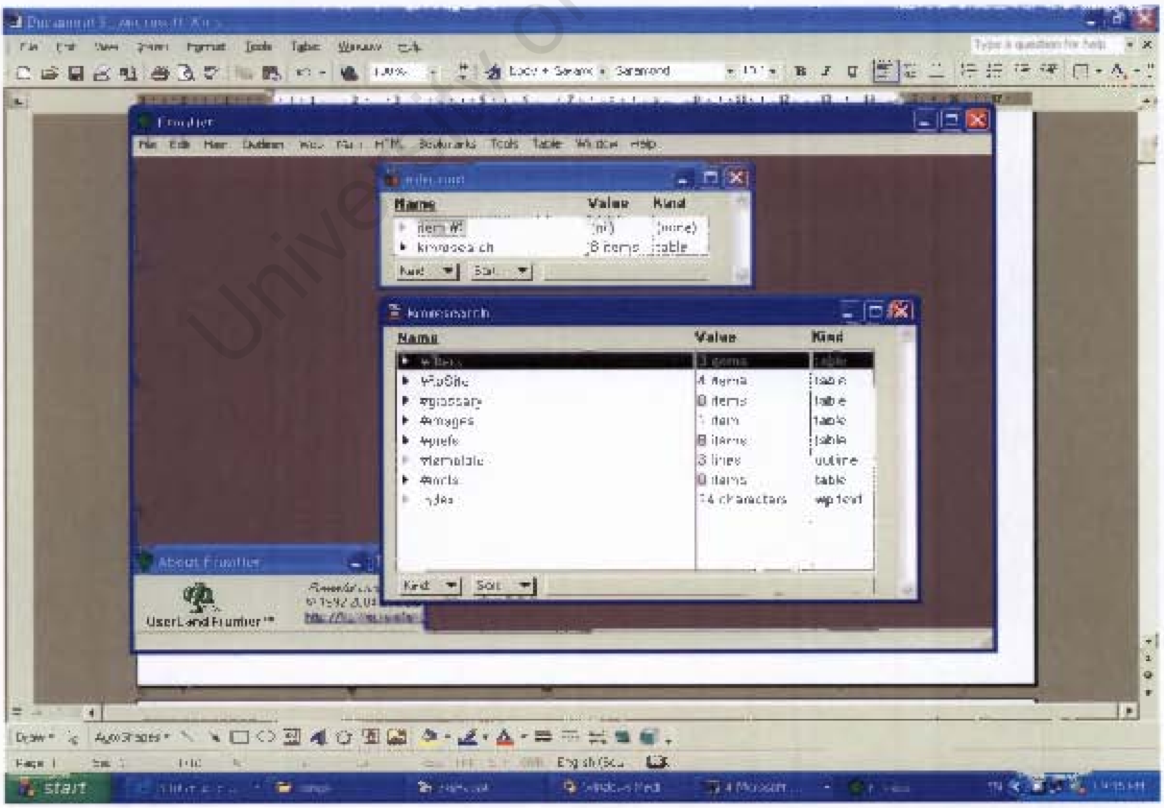


Figure 19: Development of the web pages

This gives the new page the index value and a plain window then comes up, as it would appear in the browser. Subsequently, the HTML code generated is shown below:

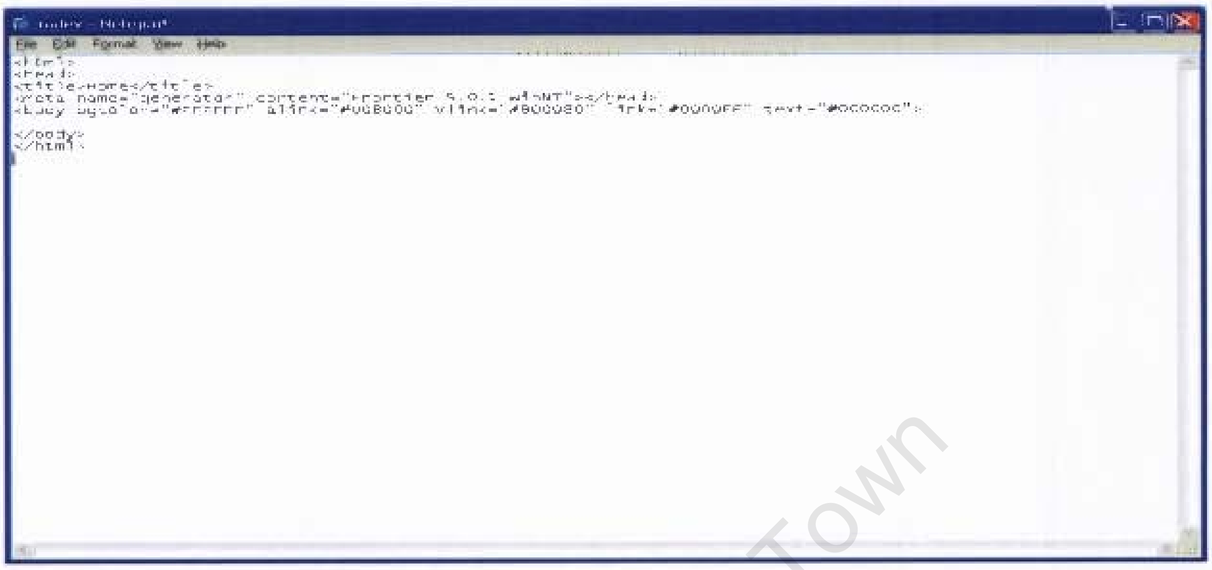


Figure 20: html generation

To write something in the page, the index has to be opened for the page that has been created and the text included as shown below:

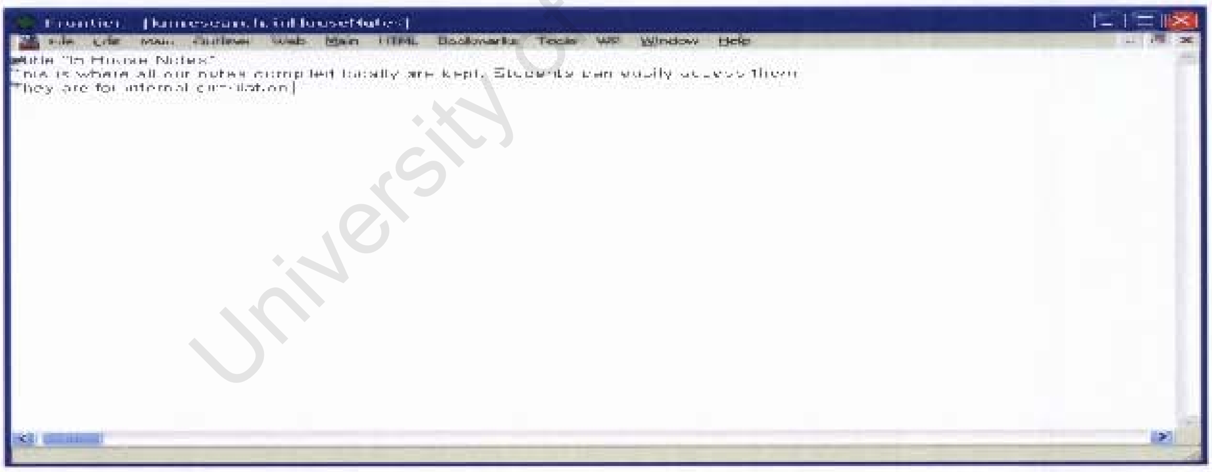


Figure 21: Indexing pages

By rendering that to the browser the out put comes out as follows:

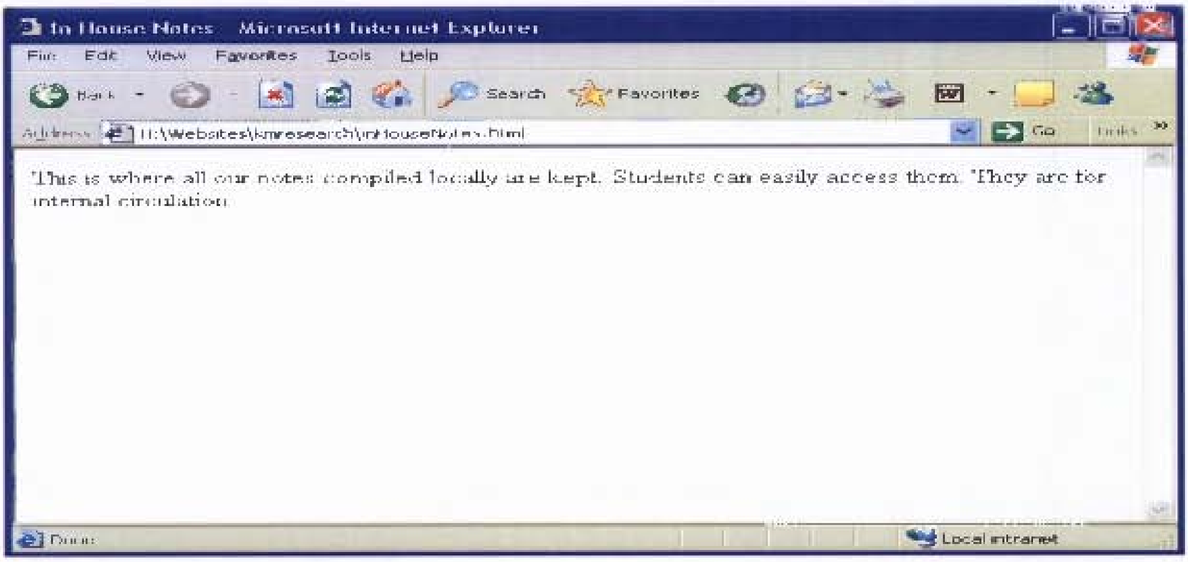


Figure 22: rendered page

To customize pages, one uses the collection of features as shown in figure 23. Some of these are very simple, for example, to alter the background colour, the #prefs feature is selected and then in the slot "bgcolor" one simply types in for example "white" (Frontier translates to hex for you). Others involve using macros, e.g {imageRef ("somepic")} or {imageRef ("somepic", hspace:2, border:2, width:100)} invokes the Frontier macro that generates a relative link to one's image in the "somepic" file.

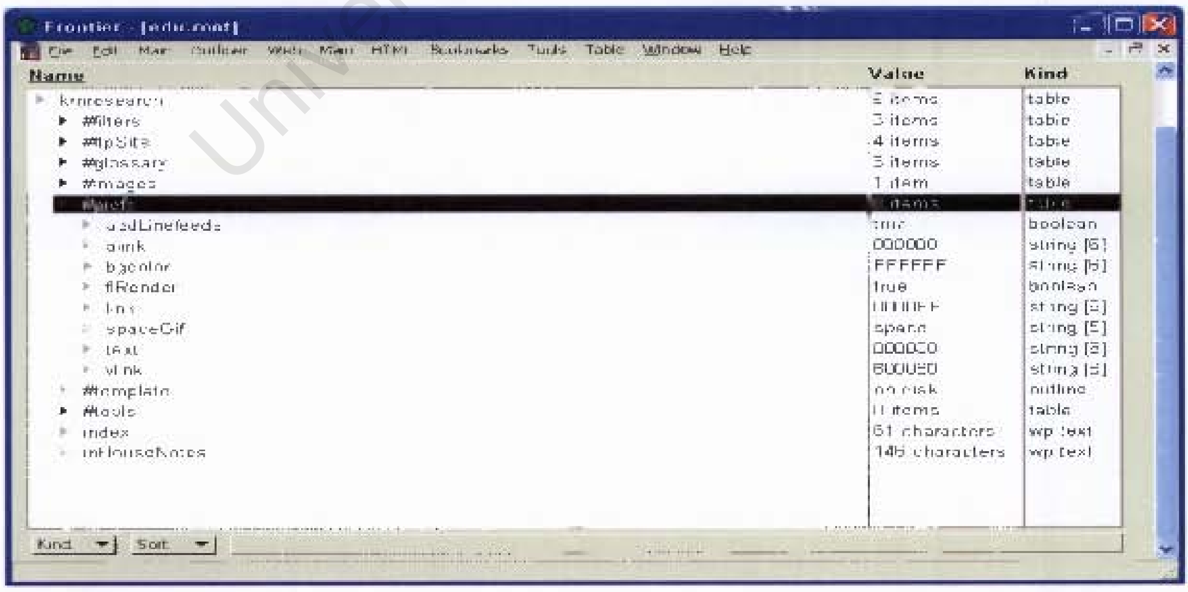


Figure 23: Adjusting Web Pages Properties

Everything that is needed to configure the page is in the list above.

"Template" is used to provide a default look to all the pages (for an example, giving borders, or referencing the {clock.now ()} macro to show the date and time).

The #renderOutlineWith feature for example is used to create structured text, as distinct from one long piece of writing: one selects from a number of predefined text-structuring styles using the #renderOutlineWith directive. One can choose a predefined style that employs white space, font size, bolding, indentation and so on, making headings and subheadings more readily apparent.

To create links to other pages, one simply gives the page title in quotes within the text for an example, "Bulletins". External pages one would like to link to frequently can be entered in the Glossary and then subsequently referred to in the same way. The nextPrevs facility allows one to drag pages (by their arrowhead icon) into the correct order and then Frontier's linkPrev and linkNext macros automatically direct the browser accordingly.

System development normally occurs on local drives. There are many ways of publishing to the server but the easiest is to publish to files and then manually publish these files to the server.

7.3 Manila's built-in macros

While scripts are used in Frontier, macros are used in Manila. The main work of macros is inserting text in the web page and returning a value. A macro of the form "{metaDataMacros.query(metaname,metavalue)}" would for example let one get a "knowledge" article by specifying {metaDataMacros.query("Topic",knowledge)}.

Table 3 shows Manila's built-in macros.

app	base64	basic	betty	bit	clipboard	clock
coercion	com	core	date	db	**dialog**	dll
editMenu	export	fatPages	file	fileMenu	Finder	Frontier
html	inetd	kb	keywords	launch	log	macros
mainResponder	menu	misc	mouse	objectModel	op	operators
osa	people	pict	point	quickTime	re	rectangle
required	rez	rootUpdates	scheduler	script	search	searchEngine
semaphore	soap	speaker	stack	string	sys	table
target	tcp	thread	ubase	webBrowser	webServer	window
		winRegistry	winShell	wp	xml	

Table 3: Manila built in macros

The list of verbs for just the dialog macro is as shown below:

- dialog.alert - Display an alert dialog.
- dialog.ask - Get input from the user via dialog box.
- dialog.confirm - Display a confirmation dialog.
- dialog.fileInfo - Display a file information dialog.
- dialog.getInt - Get an integer from the user via dialog.
- dialog.setItemEnable - Enable or disable a dialog item.
- dialog.showItem - Show an item in a dialog.
- dialog.threeWay - Display a dialog with three buttons.
- dialog.twoWay - Display a dialog with two buttons.
- dialog.yesNo - Display a Yes/No dialog.

• dialog.yesNoCancel - Display a Yes/No/Cancel dialog.

7.4 Manila as a basis for KMS development

Frontier communicates over TCP/IP. For example, it can act as a client and can communicate with the FTP server to upload a file and with the HTTP server to check whether links are binding. Unlike some content management systems, Manila uses remote procedure calls (RPC) and simple object access protocol (SOAP), and is therefore compatible with most of the internet protocols. This makes it an effective distributed management system, appropriate for knowledge management. This database is however unreservedly open giving access and configuration rights to almost anything including system objects. The idea is to give the developer freedom to exploit the full potentials of its functionalities. This means, all the variables in scripts must be declared as local otherwise system variables may be inadvertently changed. As an object database, Frontier has high performance and uses less memory. This is because objects only occupy memory when they are needed.

When KC was extended to capture *Dublin Core* [KW01] metadata, and the query facility was extended to incorporate this, this was easily achieved because Manila provides for RDF generation automatically. While the semantic weblog can be a good product of the Frontier/Manila software because of easy web publishing, the Frontier database system goes way beyond this. It is a powerful and independent database system that can be used for storage of notes, electronic mail massages, organised outlines of electronic mail accounts and records of user on a local area network, outlines of projects and reminders about appointments, all important to knowledge management.

7.5 Frontier tables used in the prototype KMS

The main Frontier tables that are used in the KC prototype are:

Navigator: Stores the site structure, having (URI, author, type)

Bulletin: Stores all the bulletins sent out and it has (bulletin text, date, author)

Filer: Stores all the files and has (author, date, type of file, size, title)

Membership: Stores particulars of members and has (name, email address, country, profession).

Aggregator: Stores all collected knowledge items from different sources and has (URI or source, date, title, topic, author)

Stats: Stores the statistics that can be used for inference and has (numberOfHits, topic, author, date, URI or source)

Metadata: These tables store all metadata in the system: name, status (e.g. lecturer) and standing (expertise level) of sources; author, URI, topic, sub-topic and certainty value of snippets and files; comments, usefulness values and ratings along with the reviewer name; originator (person from whom a file was obtained); etc.

MetaSearch: Indicates which metadata can be used in the Search page specifically; its entries are author, topic, sub-topic, body, publisher, date, title, originator, etc.

Published Page: Stores information about the Published Items page and has (author, URI, quality, topic, title, status, standing)

7.6 Summary

This chapter has described the system used to build the prototype. Clearly, this makes system development easy and quick yet achieves complicated functionality, an added advantage for KM systems. Some screen shots have been provided in appendix E.

Chapter 8
Statistical Analysis in KC

This chapter discusses statistical analysis and its use to improve KC. It demonstrates how R statistical software is used in the application of MLE in KC (inference). A multivariate analysis is also provided to compare statistical values and some visual outputs described and analysed.

8.1 Improving KC using knowledge properties

The knowledge properties of primary objects can be used to evaluate and improve the KMS. Recall that these comprise metadata (author, date, number of reads, number of links to the item, certainty values, topic, originator and status such as "lecturer"); user feedback (usefulness indicators, ratings, comments); and derived estimates of quality (based on status, usefulness, ratings, certainty), of popularity (based on number of reads, links, comments, ratings) and of similar-interest groups.

To optimise the use of KC, simple inference is used. The idea is to have a system that is flexible and adaptable. Knowledge properties extracted from KC are input to a Maximum Likelihood Estimator [R FSC05] in order to estimate the population mean, median, minimum, maximum and quartile value. The results can then be used in KC as follows:

1. displaying icons alongside an item to represent any of its knowledge properties, or its derived quality or popularity
2. re-ordering lists based on quality or popularity
3. explicitly showing derived similar-interest groups and their members
4. explicitly showing the most highly-rated items and the most popular items
5. showing the overall quality/popularity of the entire KMS, or of any of its pages or discussion topics, so as to monitor how this changes over time.

Most of the KC pages can benefit from these automated improvements. Items on any of the contribution pages can be re-ordered by quality or popularity. This can also be applied to files and directories, so that the most popular or best quality appears on top.

The next section first shows how MLE is used to derive statistics for the population based on sample data.

8.2 *Using R for MLE*

Quality measures and quality control are centred on the use of maximum likelihood estimation over a data set. The significance of using MLE in KM is to have:

I) the ability to identify knowledge points of convergence/concentration. These are data points with maximum expectation of occurrence.

II) ability to find and to show information clusters and trends.

III) ability to learn and adapt by, for example placing the data with high likelihood of being good quality on top of a page.

KC is integrated with the R statistical analysis package [FSC05] so that MLE can be applied to the knowledge property values, as show in figure 24 below.

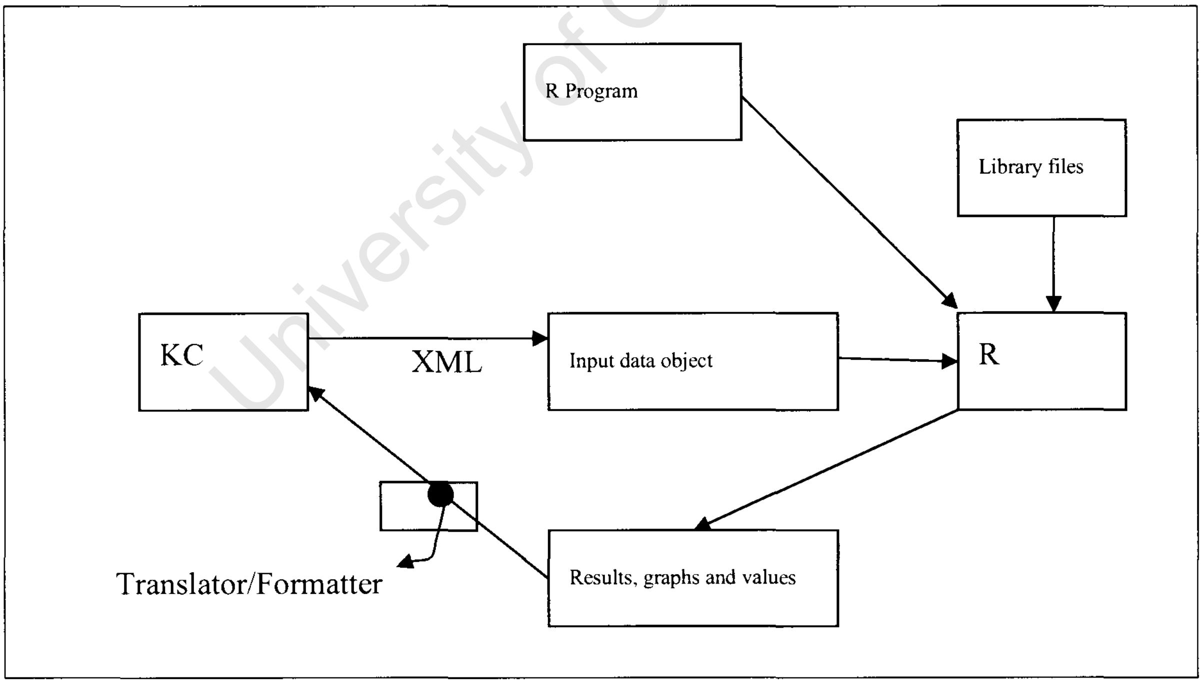


Figure 24: R software

A simple synopsis of the MLE program structure in R is as follows:

These are the specifications such as data file, output file, data specification, the model and run specifications (e.g initial mean)

```
MLE
    DATAFILE("mykcdatafile.dat")
    OUTFILE ("mykcoutfile.out")
    TITLE = "...."
    MAXITER = 100
   DATA
        <Data specification>
   END
    MODEL
        <Expression>
  RUN
        <run specification>
 END
```

Figure 25: R MLE Model

8.3 Using MLE in KC

MLE can be applied in different ways according to the complexity of the desired results. Here, three examples are given, two of them based on normal distributions, one on binomial. The names of contributors (participants) in the tables are abbreviated for convenience.

Example 1 – normal distribution

Students	1	2	3	4	5	6	7	8
Contribution Probabilities	0.11	0.14	0.10	0.22	0.08	0.25	0.07	0.03
Names	V	B	S	M	K	D	N	W
Status	Student	Student	Student	Student	Student	Student	Student	Student

Table 4: contribution frequency

The data in table 4 was captured from one of the contributions pages. This was to determine the contribution frequency over a given topic. The number of contributions by each individual on that topic was first converted to a percentage (of items on that topic contributed by him/her).

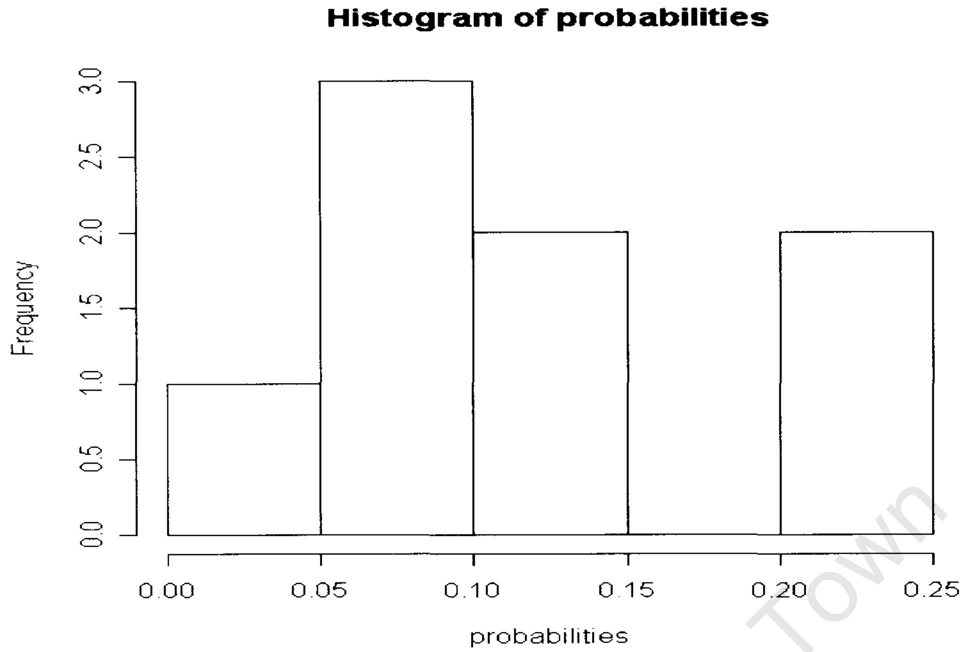


Figure 26: contribution bar graph produced by R

Figure 26 is the graphical representation of the data in R, suggesting a normal distribution. To check the distribution, we use the input below:

```
> h<-hist(x.norm,breaks=15)
> xhist<-c(min(h$breaks),h$breaks)
> yhist<-c(0,h$density,0)
> xfit<-seq(min(x.norm),max(x.norm),length=40)
> yfit<-dnorm(xfit,mean=mean(x.norm),sd=sd(x.norm))
>plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)),main="normal pdf and histogram")
> lines(xfit,yfit,col="red")
```

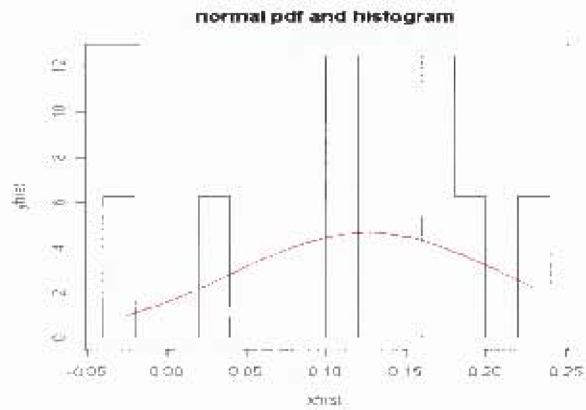


Figure 27: Fitting a normal distribution graph to sample data in R.

Figure 27 shows the output of how a normal distribution graph is fitted to the data in R. R gave the following results:

summary(probabilities)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0300	0.0775	0.1050	0.1250	0.1600	0.2500

```
> fitdistr(probabilities,"normal")
     mean          sd
 0.12500000   0.07053368
(0.02493742) (0.01763342)
```

From the above values, it can be deduced that, the average or most likely possibility of contributing is 0.125 and the lowest 0.03. The high value for the mean implies that the 8 participants were contributing equally to this topic, indicating that the topic is of wide, general interest. Such topics can automatically be placed on top in the list of topics on KC.

Example 2-Normal distribution

Figure 28 displays the number of reads an item gets, normalized to values between 0 and 1. These values are of interest in evaluating a topic or a primary object. In this example the items were all contributed by one individual.

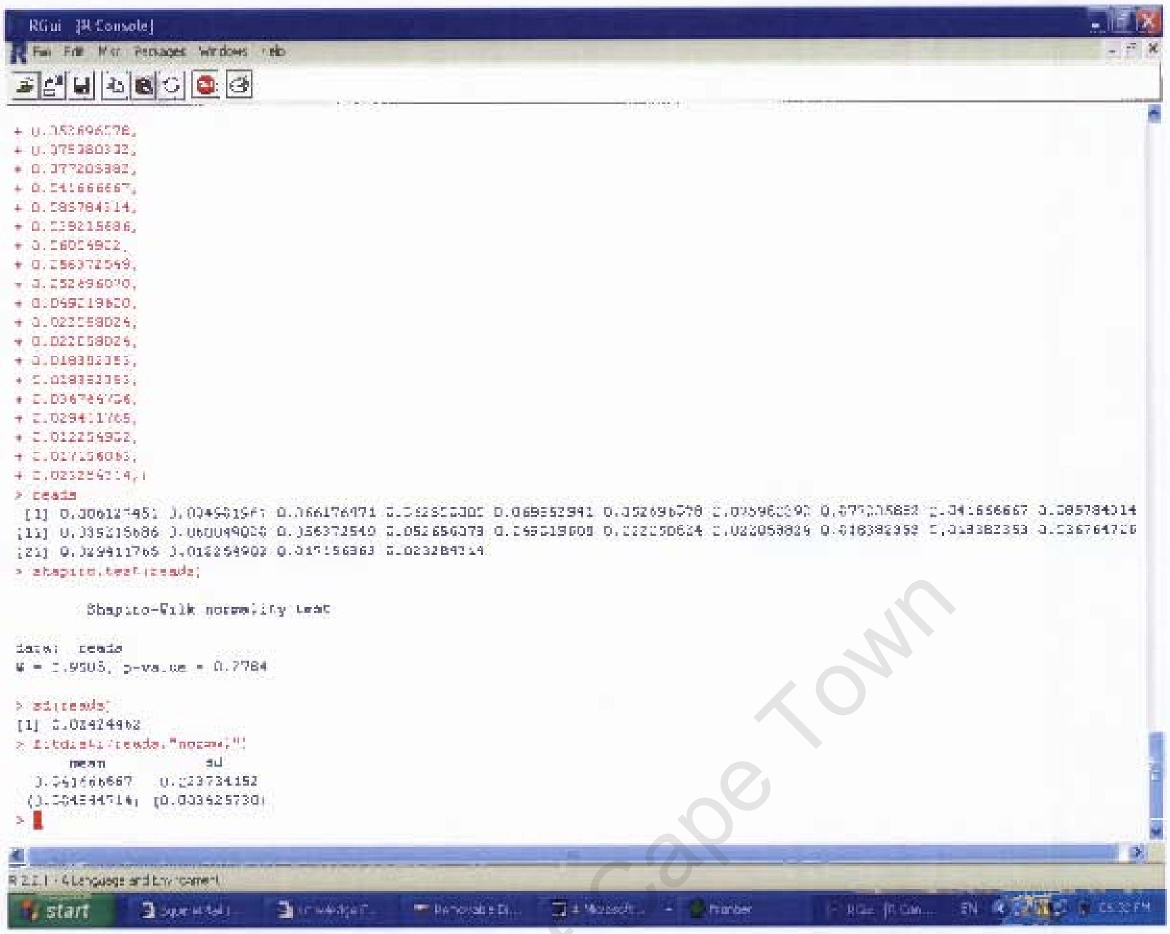


Figure 28: MLE for item reads

> summary(reads)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.004902	0.021140	0.040440	0.041670	0.060660	0.085780

From the results above, the average proportion of reads is 4.2%, the minimum 0.5% and highest around 8.6%. If there are many items, this indicates a high likelihood that people will read a contribution from this person, and the site can be adjusted to take advantage of this. A similar computation for a specific discussion topic can indicate its popularity; and if the mean is low, it can be deduced that the topic was not very relevant, so the KMS can move this to the end of the topic list.

Example 3 – binomial distribution

Research papers	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Usefulness	1	0	1	1	1	0	0	1	1	1	0	1	0	1

Table 5: Data about item usefulness

Example 3 contains data obtained from assessments of students' research proposals, which were judged as useful or not useful. An analysis on the usefulness of content was done and the following were the results

```
summary(usefulness)

 Min.   1st Qu.   Median   Mean   3rd Qu.   Max.
 0.0000  0.0000   1.0000  0.6667   1.0000  1.0000
> sd(usefulness)
[1] 0.4815434
> fitdistr(usefulness,"normal")
     mean          sd
  0.66666667   0.47140452
 (0.09622504) (0.06804138)
```

Since the value of content is either useful (1) or not useful (0), the data follows a binomial PDF naturally. MLE for the mean was 0.66666667. If a threshold point was 0.5% which is normally the case, then it can be noted that 67% would be a good result, so in this example it indicates that those evaluating the proposals found them a useful collection, which can be indicated on the KMS with a usefulness icon showing about 2/3 i.e. well above average. The median being 1 also shows that most proposals are useful. When similar analysis is done for a topic or for the set of contributions from a specific source, results can be used to highlight particularly useful ones or to re-order material, etc.

One advantage of MLE is that, apart from seeking to optimise the quality measure of knowledge which is the ultimate goal, the intermediate steps and results (through the use of libraries and run specifications) can be used for inference to reorganise the system. These include the mean, standard deviations, maximum and minimum values and inter-quartile ranges. In this way, MLE can in fact be seen to deal mostly with value and state of knowledge and value and state of KM as a whole.

Apart from the mean value that is generated by MLE, the other values can also be useful for KMS. For example if the 3rd quartile value for data on reads is high, it is clear that items (from this source or on this topic / page / KMS) will be read by most people; if the first quartile value for usefulness is 1 it is clear that most items (of this source/topic/page) are useful; if the median over all ratings (of a source/topic/page) is high then the majority of those items are good quality, etc. When such values are higher or lower than a threshold value, the KMS can automatically annotate the corresponding object (source/topic/etc.) with an appropriate icon, and/or move it to a more (or less) prominent position on the site.

8.4 Multivariate analysis

Table 6 below shows sample values extracted from KC pages. An analysis was carried out showing how knowledge properties could be used to deduce meaningful results for KM.

Snippet represents the item considered, *date* is the day the item was posted, *reads* is the number of times the item was opened, *author* is the initial of the contributor and *status* distinguishes lecturers (1) and students (2). *Usefulness* is the Boolean value of 1 if the item is useful or 0 otherwise. *Rating(ranking)* is a measure chosen from a scale of 1 to 5, with 1 being worst and 5 being best.

Data Input:

```
> measure<-data.frame(,date, snippet, reads, author, usefulness, ranking,)
> measure
```

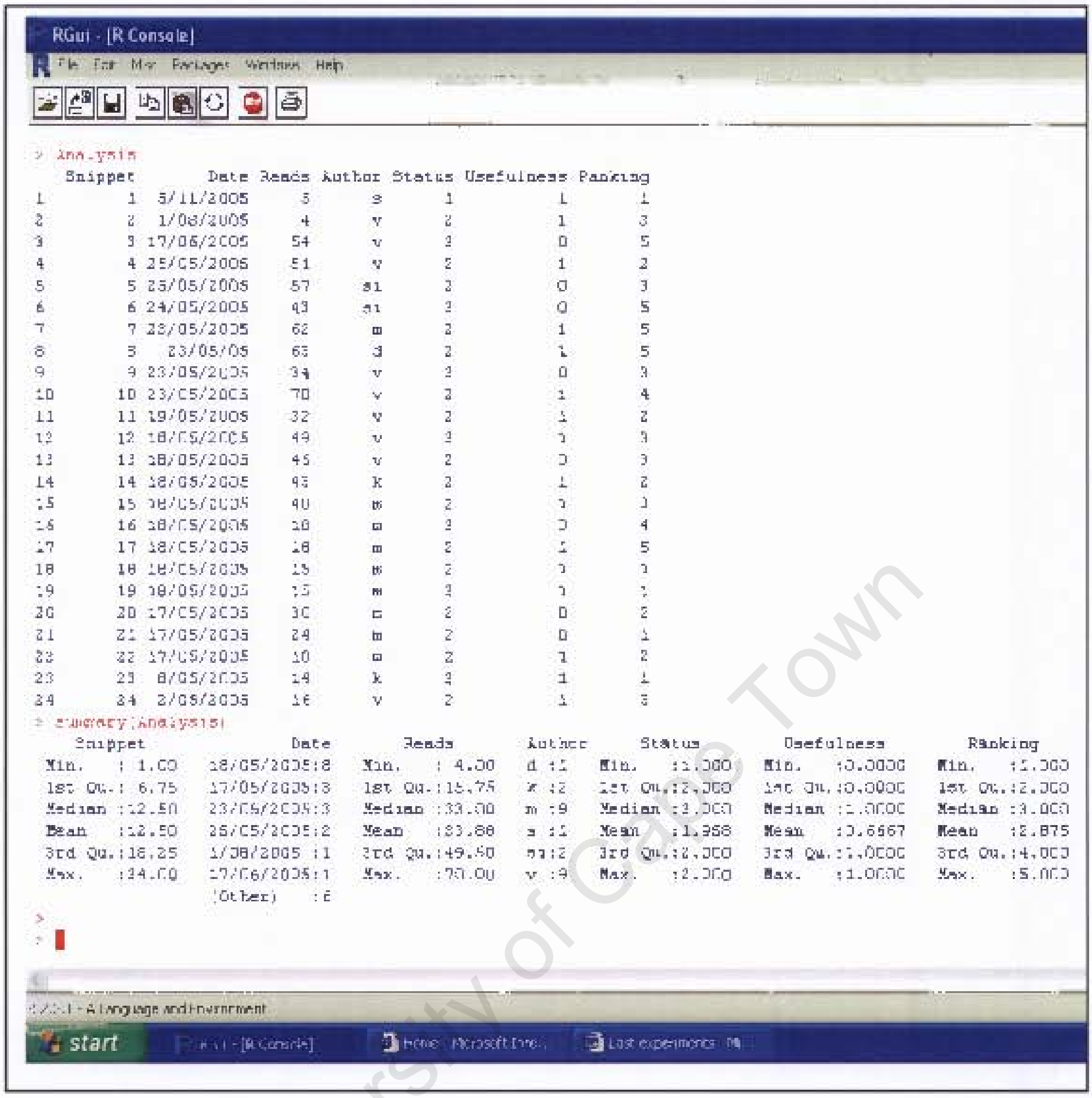


Table 6: data object for multivariate analysis

For each of the distributions from the sample data, MLE can be computed by taking some of the values such as the mean as initial estimates.

Summaries

	Min	1st Qu	Median	Mean	3rd Qu	Max
>summary(Usefulness)	0.0000	0.0000	1.0000	0.6667	1.0000	1.0000
>summary(Rating)	1.0000	2.0000	3.0000	2.875	4.000	5.000
>summary(Reads)	4.00	15.75	33.00	33.88	49.50	70.00
>summary(Status)	1.000	2.000	2.000	1.958	2.000	2.000

Table 7: Analysis summaries

From table 12, the usefulness of the items was above average with mean value of 0.6667. This means that, the contributions were very relevant to the point of discussion. Even if the content was relevant, the rating in certain cases was very poor. The mean was about 2.875, indicating that most of the items were ranked average. This suggests that participants have useful ideas to contribute but that these ideas could have been presented better. This shows that usefulness and rating are different indicators and it is good to keep both knowledge properties.

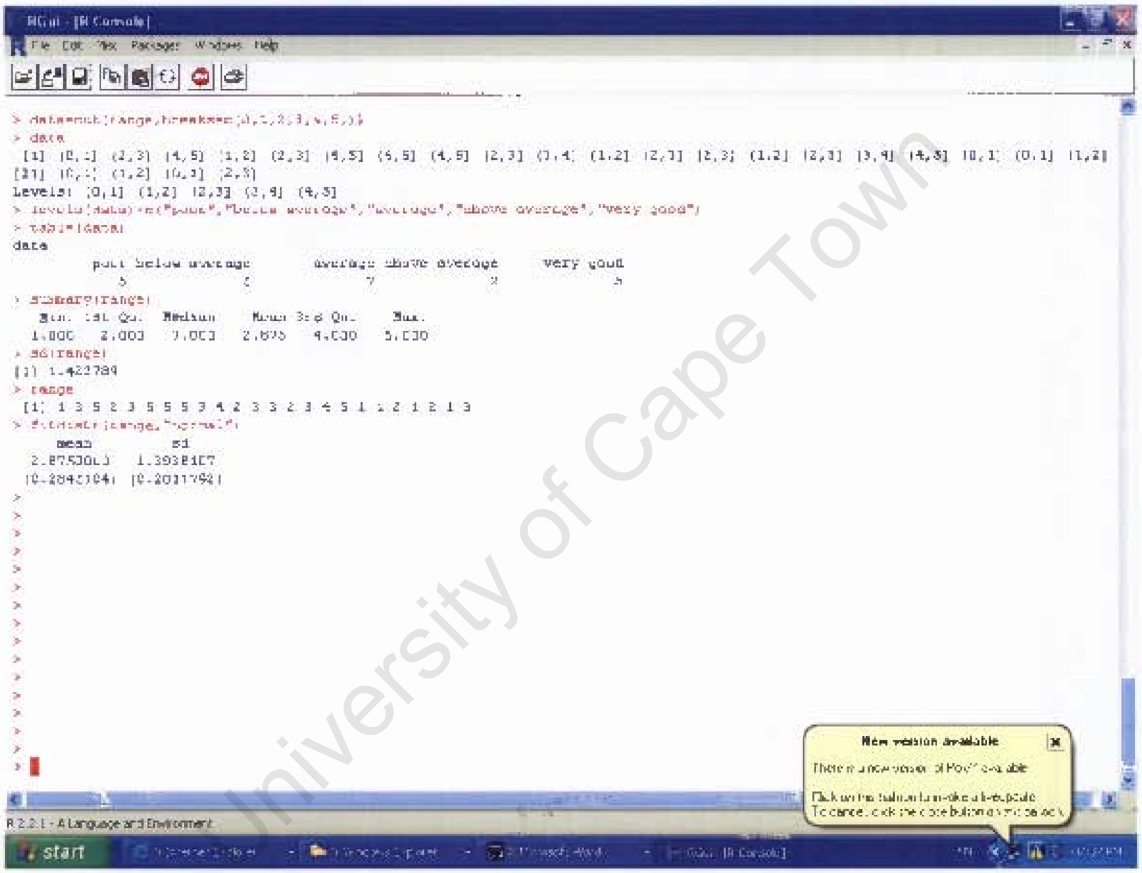


Table 8: Placing of item on rating scale

In the above example ratings were done by the lecturer, but in the future it is hoped that participants would have to evaluate items and rate them. This can easily be done by the use of the voting page as shown in figure 45 in the appendix.

On number of reads, the minimum was 4 and the highest 70. This shows that there is a considerable range, and so a list of most-read items is valuable. The status average of 1.958 indicates that most participants are students (value 2). In this case one would look at bringing more lecturers on the system.

8.5 *Visual knowledge measures for KC*

In addition to the statistical results obtained from MLE, R is also used to generate visuals that can show how successful the different primary objects in the KMS are. Tables, boxplots and cluster diagrams are generated for KC, some examples of which are shown here.

Table 9 shows the number of useful snippets and the number that are not, in this case 16 were and 8 were not. The data came from one topic on the discussion page, showing this discussion was fairly useful but not very.

- table(analysis$usefulness)

0 (not useful)	1(useful items)
8	16

Table 9: Usefulness of items on one discussion topic

Table 10 below shows the initials of participants, the number of items they contributed that were useful and the number that were not. This shows almost everyone made some useful contribution(s), but those who contributed many items were not always useful.

Name	0 " not useful"	1 " useful"
D	0	1
K	0	2
M	3	6
S	0	1
Si	2	0
V	3	6

Table 10: Participants (Initials) and number of useful items

The boxplot below (figure 29) shows the spread of ratings against number of reads.
Boxplot(reads(y-axis), ~rating(x-axis))

This shows clearly in a visual way that highly rated items are read more often (minimum just below 20 reads) and that the average number of reads increases as the rating increases.

> dotchart(reads(x-axis),author(y-axis))

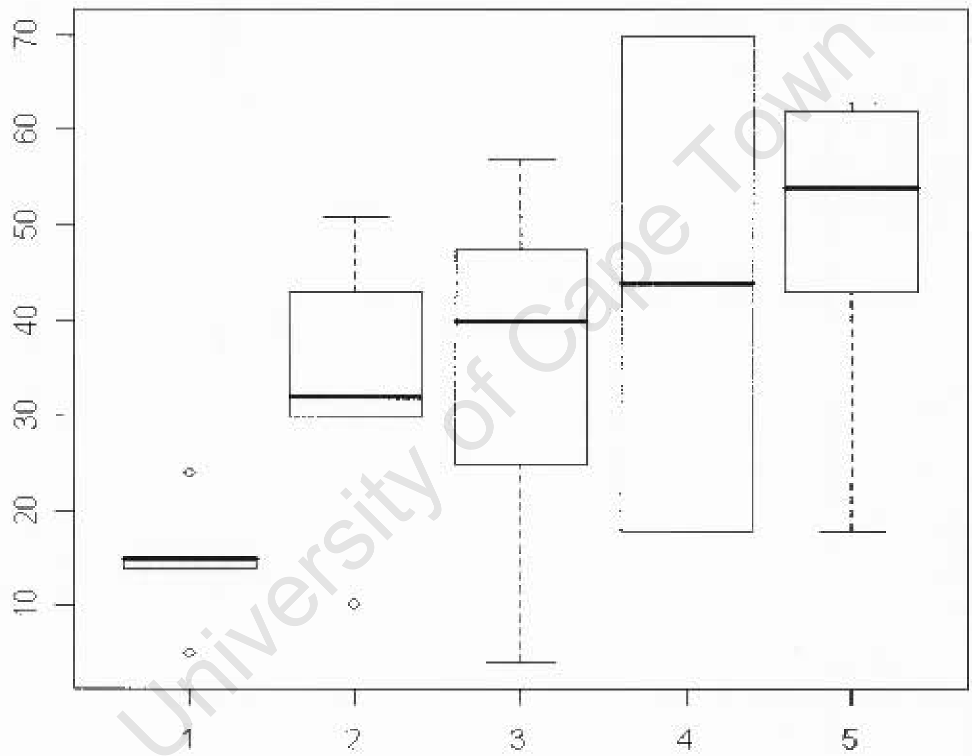


Figure 29: box plots: spread of the ratings(x~axis) over number of reads(y~axis)

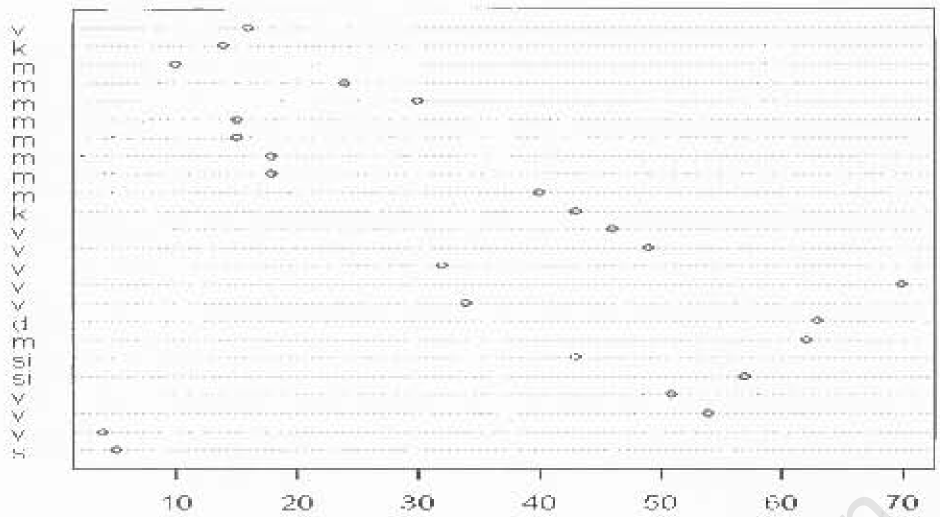


Figure 30: Dot chart: names (y~axis) and number of reads (x~axis)

Figure 30 shows the dot chart generated for number of reads per snippet, and figure 31 shows the rating obtained by these same snippets. The two diagrams are very different, showing that number of reads and ratings are two very different knowledge properties. Figure 31 also shows that more recent items (smaller snippet no.s) have been read more than older items, indicating that usage of the KMS has grown over time.

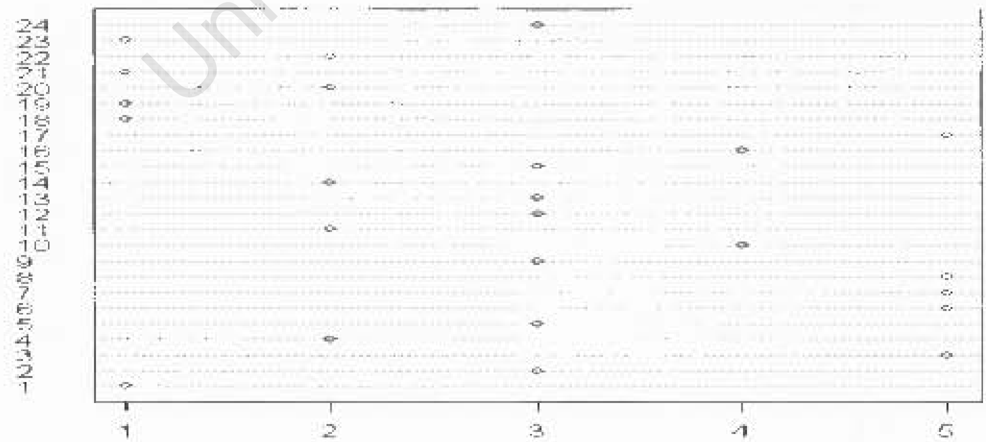


Figure 31: Snippet number (y~axis) and ranking (x~axis)

- ➢ dotchart(range(x-axis),snippet(yaxis))

Figures 32 and 33 below show some of the pages generated from statistics of KC usage.

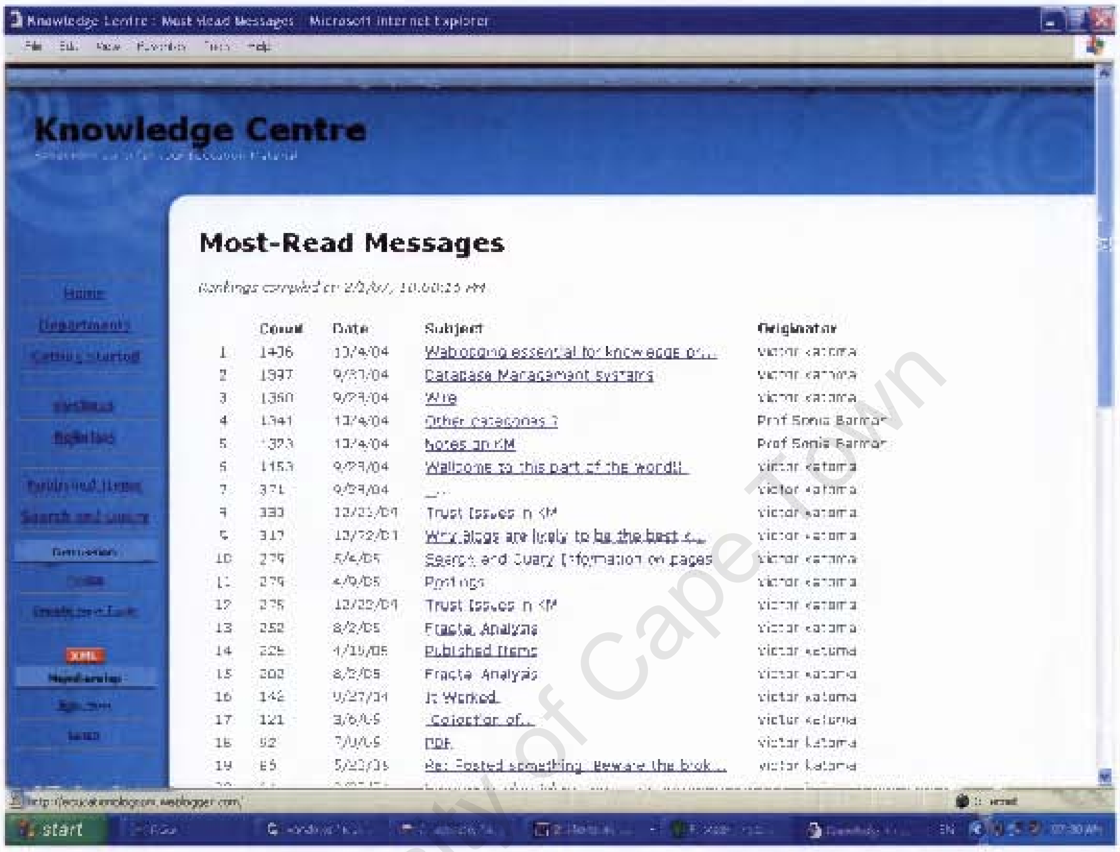


Figure 32: syndicated items and stats

Figure 33: hourly hits

Figure 33 shows the number of hits on snippets on the discussion page. These can be used to check when people use the system so that announcements (or Bulletins) can for example be made at that time.

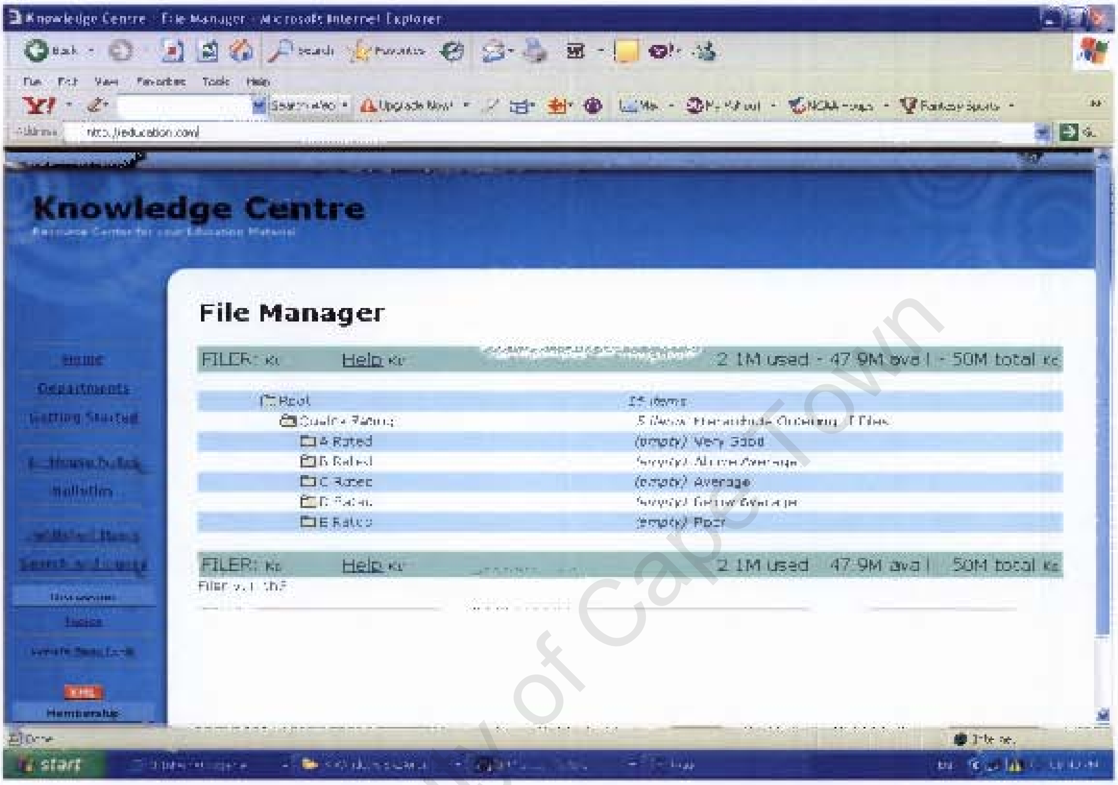


Figure 34: Partitioning of items according to quality rating (or ranking)

8.6 Summary

Chapter 8 has discussed how MLE can be used to make inferences based on knowledge values. Through this, it is possible to evaluate knowledge objects and facilitate re-organisation of KC.

Chapter 9
KC Evaluation

This chapter discusses the evaluations of KC. The first evaluation examined the usefulness of the system to MSc students and the second was concerned with usability. Both evaluations were accompanied with comments and suggestions from participants.

9.1 Initial Evaluation

After the initial audit, a second questionnaire aimed at getting feedback from the usage of KC and to follow up on some of the issues that arose in the audit. This questionnaire (See Appendix C) was completed by MSc students who used the system for their research module component.

Respondent No	Usefulness of search criteria	Usefulness of search links	Clarity and simplicity	Style of file storage	Encourages Sharing	Navigability	Use KC in future
1	Not useful	Very useful	Very clear	Better	Yes	Better	Maybe
2	Not useful	Very useful	Very clear	The same	Yes	Good	If faster
3	Very useful	Very useful	Very clear	Better	Yes	Good	Yes definitely
4	Not useful	Very useful	Very clear	Better	Yes	Better	Yes definitely
5	Quite useful	Very useful	Clear	Not used	Yes	Good	Yes definitely
6	Useful	Very useful	Very clear	Not used	Yes	Good	Yes definitely

Table 11: Results from the survey of KC usage

Based on the results in table 11, the evaluation was satisfactory. It became clear that most of the issues raised in the audit were appropriately addressed. In addition some usability issues were raised.

The following comments were made by the participants:

1. Metadata query and search are good but only when applicable; for example search by author requires knowledge of the author's name.
2. Deployment strategies were a major concern. The system would work well if hosted on powerful servers and with high bandwidth. This is a big problem especially in Africa.
3. The search techniques were useful and helped reduced the usually voluminous information quantities that traditional search methods would produce. Precision was also enhanced especially when the searches are based on metadata.
4. Knowing exactly what users need takes time, and a clear understanding of how they adapt to new systems comes with experience. Therefore, if trends of needs are not established, even a system that seems to deal with current problems effectively can become useless over time because knowledge needs change in subtle ways.

Drawbacks:

1. Users would like to precisely get what they look for from the search and therefore enough information should be captured by the system [Respondents (1), (4), (2)]. It however takes time to capture quality material. This could only be fulfilled over time as knowledge material accumulates.
2. The system could not do well on some browsers such as Mozilla Firefox. [Respondent (1)]. RSS/RDF is not compatible with some browsers (especially the old versions of internet explorer)
3. Users would prefer some graphics to make the site a bit more attractive and navigable. [Respondent (2) and (5)]

9.2 Follow-Up Evaluation

A second evaluation (by use of a third questionnaire) was conducted focussing on usability. Participants were drawn from honours and third year students from Computer Science at the University of Cape Town. The students were directed to use the system, comment, critique and provide any other suggestions on the usability.

The questionnaire is attached in appendix D (96).

Section 1	Question	Respondent 1	Respondent 2	Respondent 3	Respondent 4	Respondent 5
Was storage using directories easy	1	1	2	1	2	2
was searching items in the directories easy	2	2	1	2	1	1
Was browsing of the system easy with folders	3	2	1	2	1	1
Do you think storage was Robustness	4	2	1	3	1	2
Do you think it takes less time to get items	5	1	4	3	2	2
Was the system structured good enough for KM	6	1	2	1	3	2
Section 2	**Question**					
Are links easy to follow	1	1	1	1	1	1
Is ranking of Items Helpful	2	2	4	2	1	3
Section 3 Evaluation of the system	**Question**					
Did you find the system as a whole easy to use	1	1	2	1	1	2
Was it easy to get expected items	2	2	2	2	2	2
Was it easy to add items	3	1	3	1	2	1
Was the system helpful for KM	4	1	1	1	1	1
Would you use the system in future	5	1	1	2	2	1

Key

1. Strongly Agree
2. Agree
3. Not sure
4. Disagree
5. Strongly disagree.

Table 12: Second evaluation

Respondents' comments:

- Storing of files in hierarchical directories made it easy to browse and share knowledge resources. Knowledge items should however, be classified according to the users needs. This may need time and further experimentation to understand how users may want their items stored. Some users for example, indicated that items should be classified according to concepts. For an example a paper discussing measure in KM can be stored under measure "as a concept" within KM, instead of directly in the KM "as a topic" directory (*Partitioning refinement*).
- Grouping of items in different categories (*segmentation*) works very well for bulky storage as it focuses users to the targeted knowledge items. Description of knowledge items must also be more explicit.
- Storage strategies are the very core elements of systems that are intended for sharing and managing information and knowledge. Therefore, rights of access to files should be clearly noted.
- It was very easy to query items and metadata was very helpful especially that the system is meant for postgraduate students who would take less time to understand it.
- Rating/ranking of items can be very useful and can be a good motivator for contribution.
- By providing different search options on the search page, it was easy to find items.
- The system was very easy to navigate and simple to follow.

Suggestions from questionnaires are quoted below:

- I did not conduct extensive or complicated queries, so I am not sure whether huge volumes would be easily handled or if spelling errors are handled well.
- Rating was helpful but this is assuming all members of the group are rating honestly and that no repetitive rating happens and that system failures don't adversely affect the rating.
- It could be more useful if one can query over a specific group.
- It could be helpful to display file types (for example PDF) on the directories.
- If there is a spelling mistake in the search term, the system should make some

suggestions of the correct word or expression.

- The introductory page is very clear and simple but it needs a well structured layout to be more presentable.
- The number of times an item was read or opened may not reflect the true quality of the item as people may open an item only because of the attractiveness of the title and not content. Voting over an item is therefore a more concrete way to measure quality.
- Some tasks such as query may need a bit of tutorials especially for new users.
- Directory names should be much more meaningful
- Spelling mistakes can make search hard on the search and query page but this is not a problem as many search methods are available on the system.
- Rating would be better if measured by relevance to one's interests. Lots of things should be taken into consideration when rating. For an example rating by concept or idea.
- Search is the heart of the system and should be improved repeatedly
- Even though searching by metadata such as the author's name can be accurate, one has to understand metadata vocabulary.
- Even though the rating of items can help value knowledge items, it does not necessarily reflect the true value of the content.

9.3 Summary

The KC prototype was evaluated in two separate experiments involving different groups of users. The first experiment focussed on functionality and the second on usability. In the first evaluation, students from a Masters course that had used KC when writing their research proposals, were questioned to see whether it had addressed issues raised earlier by postgraduates regarding their knowledge management needs. The second evaluation assessed the design of KC and particularly its use of knowledge properties to obtain feedback and to measure quality. Results indicate that the system is useful and meets the needs identified in the original knowledge audit.

The following points were also noted from the second evaluation:

- It can be deduced that an effective storage strategy is of prime importance to KM. In order to attract people to share/contribute knowledge, the directory structure should over time be configured according to users needs. For example, items within the folders can be arranged alphabetically by topic.
- Rating/Ranking of items may not always determine the true value of an item. It should therefore be important to vote or else have the editor/s review the quality of the item content and rank them accordingly. During this evaluation, there was no voting because participants were not part of the Research Methods class.
- Because of the different views and needs, usability evaluation should be done on a regular basis and the system improved. From this evaluation, it was clear that the interface needed some improvement.
- Partitioning of knowledge items is a vast task and of practical importance to KM. Hierarchical directory structures provide a good basis for classification of knowledge items as observed in the first evaluation as well. Refining partitioning makes it even more accurate. However, studies have shown that more refinement can lead to loss of information [DC01].

Chapter 10
Conclusion

10.1 Thesis Summary

This study was aimed at providing a framework for developing a knowledge management system and to develop a prototype based on this. To achieve that, a knowledge audit was done and from the requirements raised, a semantic weblog was proposed to be the underlying technology. A prototype was built through Frontier and the Manila weblog publishing tool kit, focusing specifically on research training. The prototype was mostly based on metadata harnessing (RDF) with an extended notion of quality measure. The system was easy to access and use and it offered value through quality measures and inference. From the inference and measures obtained, the site could easily be evaluated and subsequently improved. The prototype was evaluated twice. Firstly "for relevance and effectiveness" by Masters students and secondly, "for usability" by honours students. Both of these evaluations were successful and the system improved accordingly.

10.2 Conclusion

The study has demonstrated that a knowledge management system (KMS) based on good KM structures and concepts (knowledge auditing, epistemology and need for knowledge measures) is plausible. Firstly, it has been observed that knowledge artefacts should be selected carefully and linked to grounded philosophical and mathematical theories for effective interpretation. To the theoretical basis, practical knowledge dimensions viz. storage, usage, sharing and dispersal should be gradually introduced. This study has proposed the use of sigma algebras to rigorously establish the process-oriented (usage) side of KM. Subsequently, the models of MLE were applied, allowing knowledge to be interpreted and presented in an effective manner.
This research has illustrated how the use of semantic weblogs can provide adequate structures and extensions for capturing metadata used in search and inference. In line with the earlier suggestion of encapsulating knowledge artefacts through object

concepts, the Frontier object database proved to be very effective. From observation it was becoming apparent that since knowledge concepts are highly divergent, the most appropriate properties need to be chosen which can also be used in computations. In the prototype, the R statistical software also fitted well with the rest of the system.

A KM that is coupled with appropriate incentives to use it wisely should be part of research students training. Letting MSc students who were in actual fact doing a research module use the system, provided the empirical experience and data in the study. To make it quick and easy to use, a semantic weblog is the best basis for such a system. To deliver added value that makes it worthwhile, knowledge quality measures, inference and self-adaptation are advocated. This research has presented a framework for constructing such a knowledge management system. It also described KC, a prototype implementation of this framework for research training of post-graduate students.

A KMS in which students contribute papers they have read or written, post snippets discussing their problems or ideas, and comment critically on items, ensures that important research skills are learnt in a non-threatening environment. The users found KC easy to use and all agreed that it encouraged knowledge sharing; while all but one were certain that they would continue to use the system in future. Some user comments noted that there was a lack of suitable sites to syndicate to, and that response time for remote users indicate that more powerful servers and greater bandwidth are needed. These aspects require more attention, and future work should also evaluate the inference and adaptive system more thoroughly once the system has been in use long enough. This study has provided a framework that is generic in nature. The idea here has been that, since sigma algebras are derived from information sets and provide entry to probability computations, they are worthwhile considering in KM. They are simple but sufficiently concrete to provide knowledge measuring.

10.3 Future work

Not all browsers support RSS/RDF and low bandwidths can be a problem especially for participants in remote places. In such places, the system can be very slow to respond which is often a cause for low user motivation. Since the system was not used extensively, some participants suggested inference on KC pages could not be done because there was too little data. Once the system has been in use long enough more meaningful evaluations could be done. As an example, with time the estimates obtained from initial usage could be used as starting values in more complex MLE. Many ideas for measure and adaptation can be explored, for instance the use of estimators other than MLE, such as Markov Models (model with time series). Such models could be helpful in monitoring knowledge needs and changes against continuous or discreet time. Another prototype can be developed to evaluate the framework in a different context such as business, medical or military. Future work can also consider a rigorous foundation for KMS based on knowledge measures and knowledge properties. Using metadata and user feedback as additional knowledge properties can be investigated further to find new ways of making inferences and adapting a KMS to better fulfil knowledge management systems requirements.

References

[Ack98] Ackerman, M.S. (1998), "Augmenting Organisation Memory." ACM Transaction of Information Systems (TOIS), 16(3), pp.203-224.

[AC69] Allen, T.J., Cohen, S.I. (1969). "Information flow in research and development laboratories", Administrative Science Quarterly, 14(1), 12-19.

[Ari45] Aristotle, (1945), "Nicomaclean ethics". Cambridge, MA: Harvard University press.

[BD98] Brown, Duquid. (1998) "Organizing knowledge", California Management Review,Vol.40 No.3, pp. 90-111.

[BK06] Berman S, Katoma, V. (2006), "A computer based framework for post-graduate education" 36 Annual Conference of the Southern African Computer Lecturers Association. pp.180-188.

[BK07] Berman S., katoma, V., (2007), "An e-learning framework based on semantic weblogs" to be presented at the 3rd Advanced International Conference in Telecommunications, Mauritius.

[BN06] Berendt B., Navigli R., (2006) "Finding your way through blogspace: Using semantics for cross-domain blog analysis" Humboldt University, department of computer science Berlin, Germany.

[Bla95] Blacker F., (1995), "Knowledge work and organizations: an overview and interpretation, Organization studies" The management school, Lancaster University, UK 1021-1046.

[Blo64] Bloom, (1964), "How people learn", British Computer Society.

[BM86] Michael L.B., Mylopoulos J., (1986), "Knowledge base Management Systems" 1986 .p 396. ISBN 0-387-96382-0, Springer-Verlag. New York.

[BPM00] Baskenville, Pawlowski, McLean, (2000), "Enterprise resource planning and organisational knowledge".Department of Information Systems, Georgia State University

USA.

[BP03] Berman S., Potgieter A., (2003), "Knowledge Management: What education can learn from Corporate Memory Systems", Technical Report, Computer Science Department, University of Cape Town.

[CDK03] Cook D., Duckworth W.M., Kaiser M.S, January 22, (2003), "Principles of maximum likelihood estimation and the Analysis of Censored Data". In Beyond Traditional Statistical Methods. NSF/ ILI grant Due9751644. pp 1-29.

[CFT99] Crooke P., Froeb L., Tschantz S., (1999), "Maximum Likelihood Estimation", Mathematica in Education and Research, 8(1), pp. 17-23.

[Car01] Carillo F.J, (2001), "Knowledge Management Consortium" Journal of the KMCI, Meta-KM: A program and A Plea. Inc Vol. 1, No.2, pg. 109.

[Cay01] Cayzer S., (2001), "Semantic Web Advanced Development for Europe" HP Laboratories, Bristol. (http://.w3.org.org/2001/SW/E/Europe/report/demo_1_report).

[Cay04] Cayzer S., (2004), "Semantic blogging and decentralized knowledge management", Communication of the ACM Vol.47.No. 12,pp. 47-52, 2004.

[Cay04] Cayzer S., (2004), " The blogosphere" Communication of the ACM, Vol.47.No. 12.

[Cay06] Cayzer S., (2006), "What next for Semantic Blogging". HP Laboratories Bristol. HPL-2006-149.

[Chu72] Churchman C.W., (1972),"The Design of Inquiring Systems, Basic concepts of systems and organisation", New York, Bensic Books.

[Coh03] Cohen S., (2003), "Keeping Current: dvanced Internet Strategies to Meet Librarian and Patron Needs", American Library Association. Chicago, IL.

[Col06] Cold J.S., (2006), "Using Really Simple Syndication to Enhance Student Research". ACM SIGITE, Newsletter, Vol.3, No. 1.

[CLW99] Crossan M.M., Lane H., White R.E., (1999), "An organisational learning framework. From intuition to institution", Academy of management review, 24, pp. 522-537. Richard Ivey school of business.

[DDB98] Devenport T.H., De Long D.W., Beers, M.C., (1998) "Successful Knowledge Management Projects," Sloan Management Review, pp.43-57.

[DE01] Dubra J., Echenique F., (2001) "Measurability is not about Information" March 2001. Cowles foundation discussion paper no. 1296, Yale University.

[DE04] Desoza K.C., Evaristo J.R., (2004), "Managing knowledge in distributed projects", Communication of the ACM 47(4), pp. 87-91, 2004.

[Dig04] Dignum V., (2004), "Personalised support for knowledge sharing". 11th ACM International conference proceedings Vol. 65 page 1, June 2004 ACM, Cambridge University Press.

[DP97] Davenport T.H., Prusak L., (1997), "Working knowledge: How organizations manage what they know", Harvard Business School Press, 1997.

[DP97] Davenport T.H., Prusak L., (1997), "A framework for management systems: Issues and Challenges for theory and practice", ACM conference 21 international conference, p.302-312.

[DP98] Davenport T.H., Prusak L., (1998), "Working knowledge: How Organizations Manage What they know", Boston, MA: Harvard Business School press.

[Dow02] Downes S., November 2, (2002) "An Introduction to RSS for Education Designers" Merlot Centre for learning and instructions. Vancouver, British Columbia.

[Eve79] Everith B., (1979), "Cluster analysis", 26 June 1979 Handbook of statistics, p155.

[FD02] Farhoomand, Dury (2002), "Management information overload". Communications of the ACM, 45(10), 127-131

[Fro] Copyright Userland Software, 1996-2007.

[FSC05] R : Copyright (2005), The R Foundation for Statistical Computing, Version 2.2.1

(2005-12-20 r36812), ISBN 3-900051-07-0

[FTH03] Frasincar F., Telea A., Houben G.J., (2003),"Visualisation of RDF-based Information", Department of Information Technology, Eindhoven University of Technology.

[GM96] Gains B.R., Mildred L.G., (1996), "Knowledge modeling and inference through the World Wide Web". In Proceedings of KAW96, the 10th Knowledge Acquisition Conference, Vol 46, Issue 6, pp.729-759. Academic Press, Inc, USA.

[Gol96] Goleman D.P (1996), "Vital lies simple truth: the psychology of self-description". New York: Touchstone Books.

[Gol99] Goldman, A.I., (1999), "Knowledge in a social world", Oxford: Clarendon Press.

[Hah00] Hahn J., (2000), "A framework for knowledge management systems: Issues and challenges for theory and practice" Proceeding of the 20th international conference ICIS, pp302-312. University of Minnesota, USA.

[HH03] Herrmann T., Hoffmann M., (2003), "Concepts for usable patterns of groupware Applications", ACM Information and Society, University of Dortmund, Germany.

[HK01] Hampel, Keli-Slawik R., (2001), "Structuring Information in Team-Distributed Knowledge Management in Cooperative Learning Environments". Journal on educational Resource in computing, Volume 1, Issue 2es (summer 2001). Article 3. ACM Press NY, USA

[Hof03] Hofr-Alfers J., April (2003), "Measuring knowledge and KM in an organisation with an explicit top-down knowledge strategy" Siemens AG, Corporate Information and Operations Enabling Process and Knowledge Management. Vol 9, issue 7 Suedalle 1 85356 Munich, Germany.

[Hof03] J. Hofer-Alfeis, (2003) "Effective integration of knowledge management. into the business starts with a top-down knowledge strategy", Journal of Universal Computer Science, 9(7), pg. 719-, 2003.

[Hof03] Hofer-Alfeis J., (2003), Siemens AG, "Corporate Information and Operations. Enabling processes and KM", pp2-4. Journal of Universal Computer Science.

[HPR01] Harper R., Procter R., Randall D., (2001), 'Safety in Numbers' : "Calculation and document re-use in knowledge work", Colorado, USA. 2001 ACM 1-58113-294-8/01/0009.

[Hyl02] Hylton A., (2002), "Measuring and valuing knowledge. The role of the knowledge auditing", pp 4-5.

[Hyl02] Hylton A., "A knowledge management initiative is unlikely to succeed without a knowledge audit", pp. 1 – 13, www.annHylton.com, 2002.

[Hyl04] Hylton A., (2004) "A KM initiative is unlikely to succeed without knowledge audit" (www.annHylton.com)

[JK96] Johnasson J.K, Nonaka I., (1996): "The Japanese way of Marketing", London, UK: Butterworth-Heinemann, 1996.

[ISO03] ISO/IEC TR 11179: "Metadata registry" (all parts). Geneva: International Organisation for standardization and International Electrotechnical Commission.

[ISO04] ISO. (2004). ISO/IEC 11179: TR 20943-3: Proceedings for achieving metadata registry content consistency- part 3: "Value domains", Geneva: International Organisation for Standardisation and International Electro-technical Commission.

[ISO95] ISO/IEC 11404: Language independent datatypes. Geneva: International Organisation for standardization and International Electro-technical Commission.

[Kal99] Kalman, M.E., (1999), "The effects of organisational commitment and expected outcomes on the motivation to share discretionary information in a collaborative database" University of South California, department of computer Science.

[KC04] RDF/XML Syntax specification (Revised) 10 February 2004. HP labs, University of Bristol. http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/
Editors, Graham Klyne and Jeremy J. Carroll, Hewlett Packard Labs.

[Kel96] Keli-Slawik R., (1996) "Information and communication technologies in education and training" (Part A). Working document for the project steering group, PE 1665.710 Directorate General for Research Luxembourg.

[KMM02] William R. King W.R., Marks P.V., McCoy M., (2002), "The most important issues in knowledge management", Communication of the ACM September 2002/vol.45,No. 9.p93

[KNRT04] Kumar R., Novak J., Raghavan P., Tomkins A., (2004), "Structure and evolution of blogspace":, Communication of the ACM 47(12), pp. 35-39.

[KW01] Koch T., Weibel S., (2001), Editorial "Introduction to a Special Issue on Metadata: Selected papers from the Dublin Core 2001 Conference", J. Digit. Inf. 2(2), 2001.

[Lou01] Loungino H., (2001) "The fate of knowledge", Princeton, NJ: Princeton University Press

[LT04] Lucy T., Terry (2004)". "Information Management Systems", New York press P. 35

[Men05] Mendelson E., (2005), "Syndicate Yourself with RSS Feed Creator". PC http://www.pcmag.com/article2/0,1895,1922822,00.asp

[MG03] Mutton P., Golbeck J., (2003), "Visualization of Semantic Metadata and Ontologies". In proceedings of the 7th international conference on information visualisation 2003 Ieeexplore.org pp 300-312, University of Kent, Canterbury.

[Min95] Minsky M., (1995), "A framework for representing knowledge", in the Psychology of Computer vision, Winston, Patrick H.(e.d), McGraw – Hill New York, 1995.

[MM00] Morey, Maybuy, (2000) "Can knowledge management succeed where other efforts have failed?" Classic and contemporary works, MIT press, 2000, 139-182.

[MQ96] Magill M., Quinzii M., (1996), "Theory of Incomplete Markets", Cambridge, London: MIT Press.

[NT95] Nonaka I., Takeuchi H., (1995). "The knowledge creating company: How Japanese companies create the dynamic of innovation". New York: Oxford University Press.

[Och00, HP04] IBM Exploring the standards of web-based metadata Ochi Ogbuji 01/December/2000. http://www.ibm.com/us/.

[Pat02] XML Syntax and RDF Semantics (Peter PatelSchneider, Bell Laboratories 600 Mountain Avenue Murray Hill, NJ, USA pfps@research.belllabs.com. J`er ome Simeon Bell Laboratories 600 Mountain avenue Murray Hill, NJ, USA simeon@research.belllabs.com. 2002 p.443 (ACM).

[Per03] Perisic I., (2003), Finding information today. "Combining advanced dynamic semantic profiling with data mining technologies to create unified enterprise information management solutions", Entopia, white paper, March 2003 .p6.

[PLG03] Perez-Contino M., Lopez-Lopez A., Gomez M.Y., (2003), 12-15 September. "A multi-agent system for web document authoring" First Atlantic web intelligence conference, Madrid Spain.

[Pol67] Polanyi M., (1967), "The Tacit Dimension", Double day Garden City 1967 New York: anchor Books.

[Pol71] Polya (1971), "Estimations in schools", p.135, Conference, British Society.

[Pol71a] Polya, G. (1971", "How to solve it." Princeton University Press, Princeton, NY.

[Pol94] Polanyi L., (1994), "A syntactic approach to discourse semantics. University of Amsterdam" Proceedings of the 10th international conference on computational linguistics. P 413-419. Association for computational linguistics.

[Vyg78] Vygotsky P., (1978), "How people Learn", National academics press.

[PW03] Pitman PW., Weare J., "Probability theory" (Spring 2003).

[Ran] http://en.wikipedia.org/wiki/Function (Mathematics).

[Rec00] Recanat I.S., (2000), "Knowledge management: A new idea or a recycled concept" Graduate school of Business. Tel Aviv University Vol3, Article 14 June 2000.

[RC06] Reynolds D., Cayzer S., (2006), SWAD-E Demonstrators- Lessons Learnt available at: (http://www.w3.org/2001/sw/Europe/reports/demo-lessons-report/accessed) 13 October 2006.

[RDF] RDF/XML Syntax Specification (Revised) 10 February 2004. Hewlett Packard Labs, University of Bristol. http://www.w3.org/TR/2004/REC-rdf-syntax-gramar-20040210/.

[RDF] "Resource Description Framework (RDF)", www.w3.org/RDF/.

[Ric05] Richardson W., (2005). "The ABC of RSS. Technology and Learning", 25 may 2005 issue 10, 20-24. Communications of the ACM July 2003/ vol. 46. No 7.

[Ric05a] Richardson W., (2005), "RSS: A quick start guide for Educators", Version. 1.5, Updated March 29, 2005. Retrieved November 5, 2005 from http://www.weblogged/rss_for_ed.

[Rey05] Reynold D., (2005), " Using Networked Technologies to support Conferences". University of Brostol (HP laboratories) New York, USA. ACM press.

[RM00] Rittenbrush M., Mansfield T., (2000), "Making sense of Syndicate Collaboration" Department of computer science, University of Queens Land 00, 1-2, 2000, city, state Australia.

[RT99] Ruth S., Theobald J., (1999), "A university-based approach to the diffusion of knowledge management concepts and practices" Proceedings of the ACM SIGCPR Conference on Computer Personal research p. 283-290, 1999. ACM press, USA.

[Rut99] Ruth S., (1999), "The university-based Approach to the diffusion of KM concepts". Communications of the ACM, 1999, 1-58113-063, 063-5-99-04, .p2836.

[Sch98] Schmitt F., (1998), "Social epistemology". In E.Craig(ED), Routledge encyclopaedia of philosophy. London: Routledge, article /po46.

[Shu98] Shun S., (1998) "Negotiation the construction of organisational memories" May 2001, http://d3e.open.ac.uk/general/negotiating-org-mem/negotiating-org-mem-01.html.

[SK91] Sproll L., Kiesler S., (1991), "New ways of working in a Networked Organisation", the MIT press, Cambridge: MA, 1991.

[Sky03] Skyrme D.J., (2003), "Effective integration of Knowlegd management. Models for metrics Issue" No. 69, January 2003. 13 update/Enovation international news.

http://www.skyrme.com.

[Sol01] Solomon M., (2001), "Social empriricism, Cambridge", MA: MIT press.

[SK00]. Srikantaiah K., Koenig M.D.E., (2000), Eds "Why knowledge Management Systems Fail?" November, 2000. Knowledge management for the information professional. Medard, NJ.

[SWADE] (2001), "Semantic Web Advanced Development for Europe. Semantic Blogging", http://www.w3c.org/2001/SW/Europe/report/demo_1_report/.

[Tho05] Thomas C., (2005), "Order From Chaos Via RSS". Information Week, 1050 (August 1, 2005), 41-45.

[Tin05] Tinkster, (2005, October 26). "Is XML universal?" [Msg 3]. Message retrieved November 5, 2005 from http://www.linuxquestions.org/questions/history/377082.

[Tlu03] Tulcea A.I., (2003), "Markov chains monte carlo methods", Christian P. Robert University Paris Douphe.

[Tsu04] Tsui E., (2004), "Technologies for personal and peer-to-peer (p2p) Knowledge Management". (http://www.2csc.com/lef/programs/complete_02.html).

[Tus77] Tushman M.L., (1977), "Special boundary roles in the innovation process". Administrative Science Quarterly, 22(4), 587-605.

[Tur02] Turner S., 2002, "Brains/ Practices/Relativism", Chicago: University of Chicago press.

[VK05] Van Deussen A., Knipers T., (2005), ACM 99/05. "Identifying objects using cluster and concept analysis", In the proceedings of the International conference on software engineering, ICSE-99. pp 246-255, January 1999. CiteSeer.

[Vro64] Vroom, (1964), "Work and motivation", New York: Willey.

[Wal90] Walker A., (1990), "Knowledge Systems principles and practices", IBM Journal of research and development 30(1).

[WH01] Weal M.J., Hughes G.V., (2001), "Open Hypermedia as a Navigational Interface to Ontological Information Space". Department of Electronics and Computer Science, University of Southampton .p227, 2001 ACM (wjw@ecs.soton.ac.uk).

[W3C] RDF/XML Syntax specification W3C recommendation 10, February 2004 (http://www.w3c.org/TR/2004/REC-rdf-syntax-grammer-20040210/

[Wik06] "Maximum likelihood", From Wikipedia, the free encyclopaedia, en.widipedia.org/wiki/Maximum_likelihood.

[YBCFW01] Ziv Bar-Yossef ZB., Berg A., Chien S., Fakcharoenphhol T., Weitz D., (2001) "Uniform sampling from the web via random walks", University of California at Berkeley, department of computer science, USA.

[Zac99] Zack M, (1999), "Developing a knowledge strategy", California Management Review, 41,125-145.

[ZR05] Zaihrayeu I., Robertson D., (Eds), (2005) " Peer-to-Peer Knowledge Management", proceedings of the second workshop on peer-to-peer knowledge management, July 17 2005, La Jolla, San Diego, California, USA.

[ZR99] B. Zhu, Ramsey M., (1999) Management Information Systems department "Support Concept-based Multimedia information retrieval", A KM approach, Proceedings of the 20th International Conference December 13-15 .p 1, University of Arizona.

Appendices

Appendix A (96): Audit Questionnaire

We are conducting a survey on Knowledge Management and we would definitely appreciate your contribution towards this exercise by answering the following few questions. Please feel free to discuss any relevant information pertaining to Knowledge Management.

1. How do you reach the knowledge material in your course/research or reading?

   a) By e-mail
   b) Searching the Web
   c) Books and /or other paper based Publication
   d) Other ways, please explain.

2. For each of the above sources you picked in question one (Q 1), is there any problem or difficulties you encounter in the process of seeking the Knowledge material?

   a) yes
   b) No

   If yes, please explain

....

....

3. For each of the above sources in question one (Q 1) do you feel there should be a better way of organizing it for easy and accurate access?

   a) yes
   b) No

   If yes:

   Any ideas of how these could be addressed?

....

....

4. Are you satisfied that you are reading what you need/want

   a) yes no problem
   b) partially
   c) just a bit
   d) Not at all

5. Do you have access to knowledge around your work/research group, either electronically or verbal?

   a) Yes
   b) No

6. If the answer to 5 is yes, explain how this works?

....

....

7. If the answer to question 5 is No, then what do you think are some of the reasons?

....

....

8. Knowledge Management encourages sharing of information/Knowledge just like on the Web and as such there is a need to contribute.

   How do you store your data/information and knowledge? Is it in a structured way or not?

   a) Yes
   b) No

   If Yes then how?

   If Not, then why Not?

....

....

9. Do you feel your knowledge is shared with others who are in need?

   a) Yes
   b) No

   If No, please explain why, if yes explain how?

....

....

10. Who would you be happy to share your knowledge with?

    a) Anyone
    b) Only people you trust
    c) Only some items
    d) No-one. Explain why

....

....

11. If you had a software tool to structure and manage your knowledge items (doc, pdf, own notes, summaries etc), would you use it?

    a) Yes definitely
    b) Only if it is quick to enter and retrieve knowledge
    c) Only if it would be easier to find relevant knowledge items
    d) Definitely not

    If d) is your choice then what is the reason?

....

....

Other criteria. Please explain

…………………………………………………………………………………………………

………

12. In a knowledge base, which of the following do you think would be practical and useful communities for sharing knowledge (Tick any number of answers)

   a. Anyone with interest
   b. People in the same department/research group
   c. Students doing research in same field
   d. Academics/professors in the same field
   e. Research groups in the same field
   f. None of the above. Explain why?

………

………

Appendix B - Sample KC pages

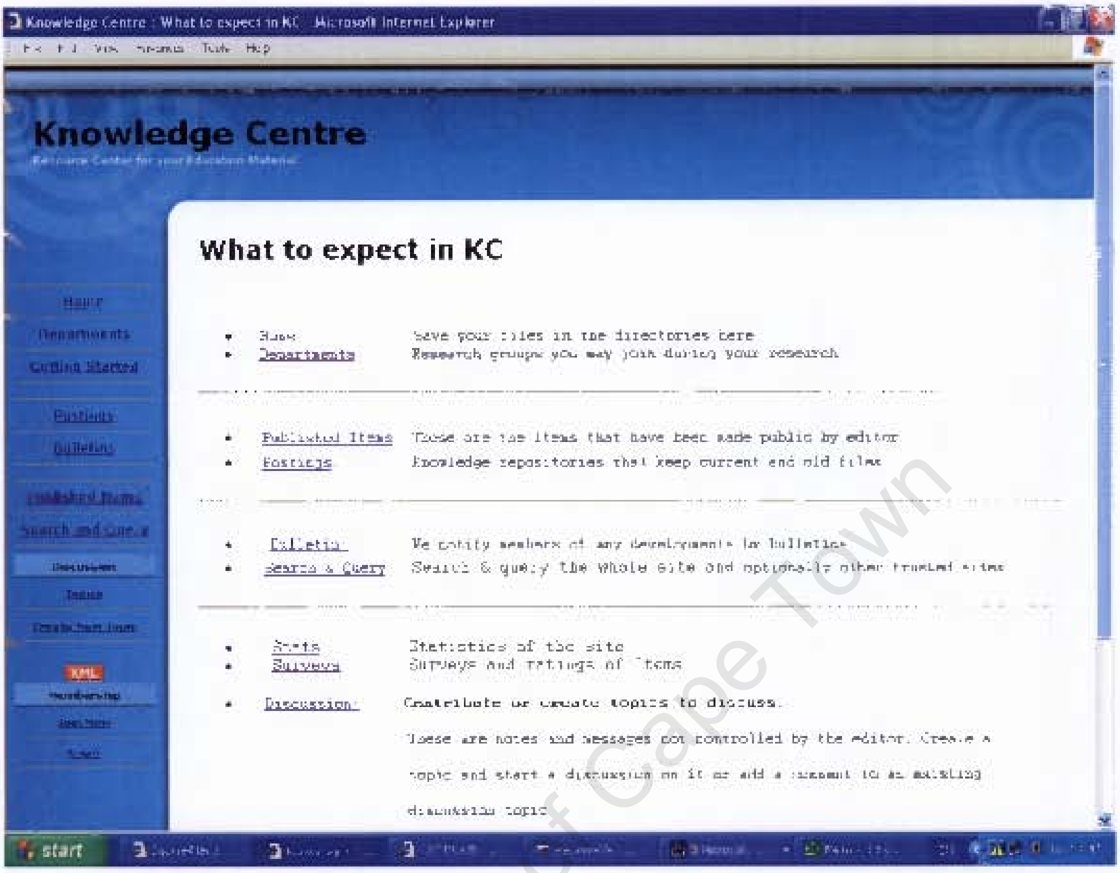


Figure 35: Introductory page

Getting started: An introductory page with explanations of what to expect in KC. The XML button on the navigation panel leads to all the syndicated items on KC.

Classified files can easily be managed with hierarchical directories as show below. Manila is compatible with file manager version 1.1b3 through which dynamic taxonomy was implemented.

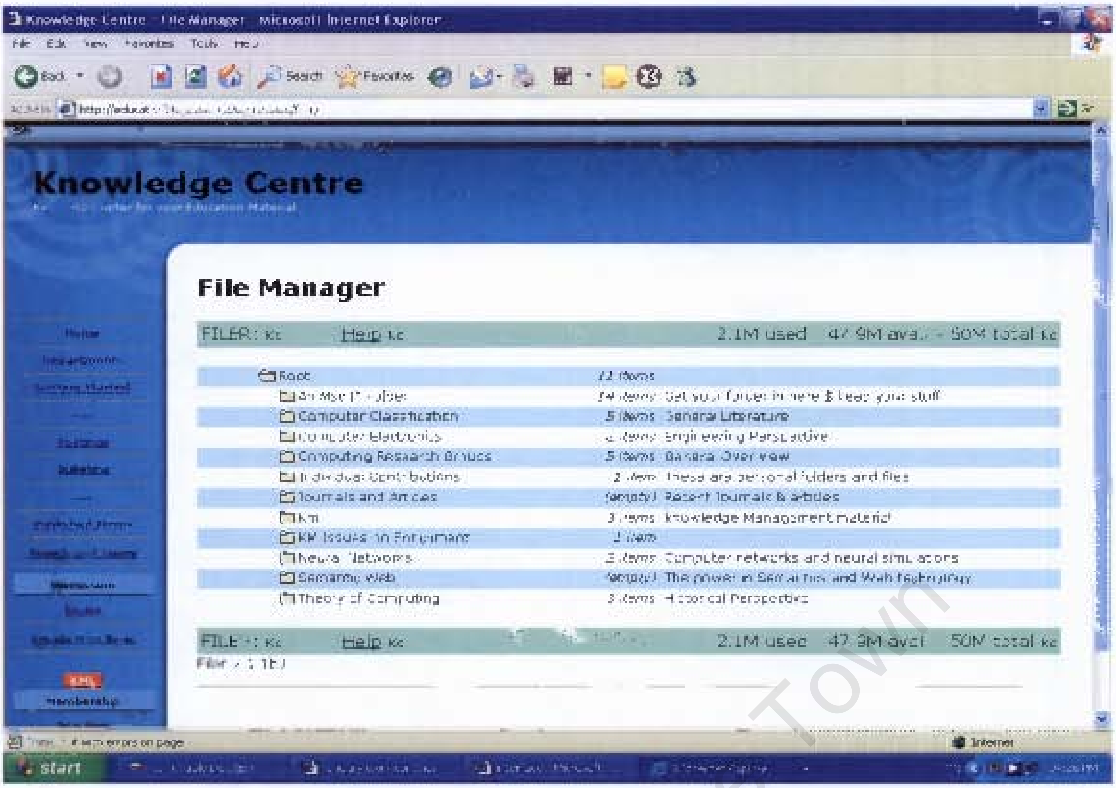


Figure 36: Directory management

File Manager (Simple visual hierarchies of directories)

Directory hierarchies simplify search tasks because the emphasis here is where a file fits in the directory structure. This offers a different search approach from search by keyword were one only relies on the meaning and structure of the specified words. Keyword searches usually have common spelling problems, inconsistent compounding of words and redundancy.

Figure 37 shows folders with names of Masters students who were users of the system.

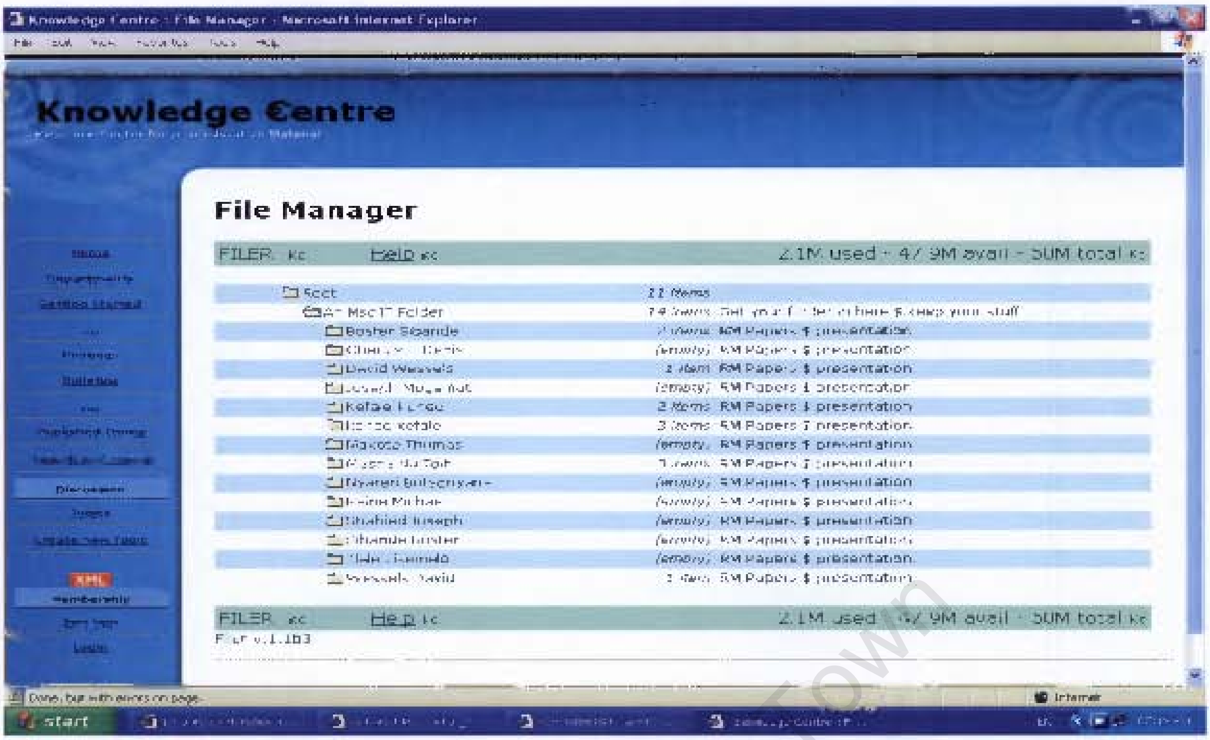


Figure 37: Private Directories of participants

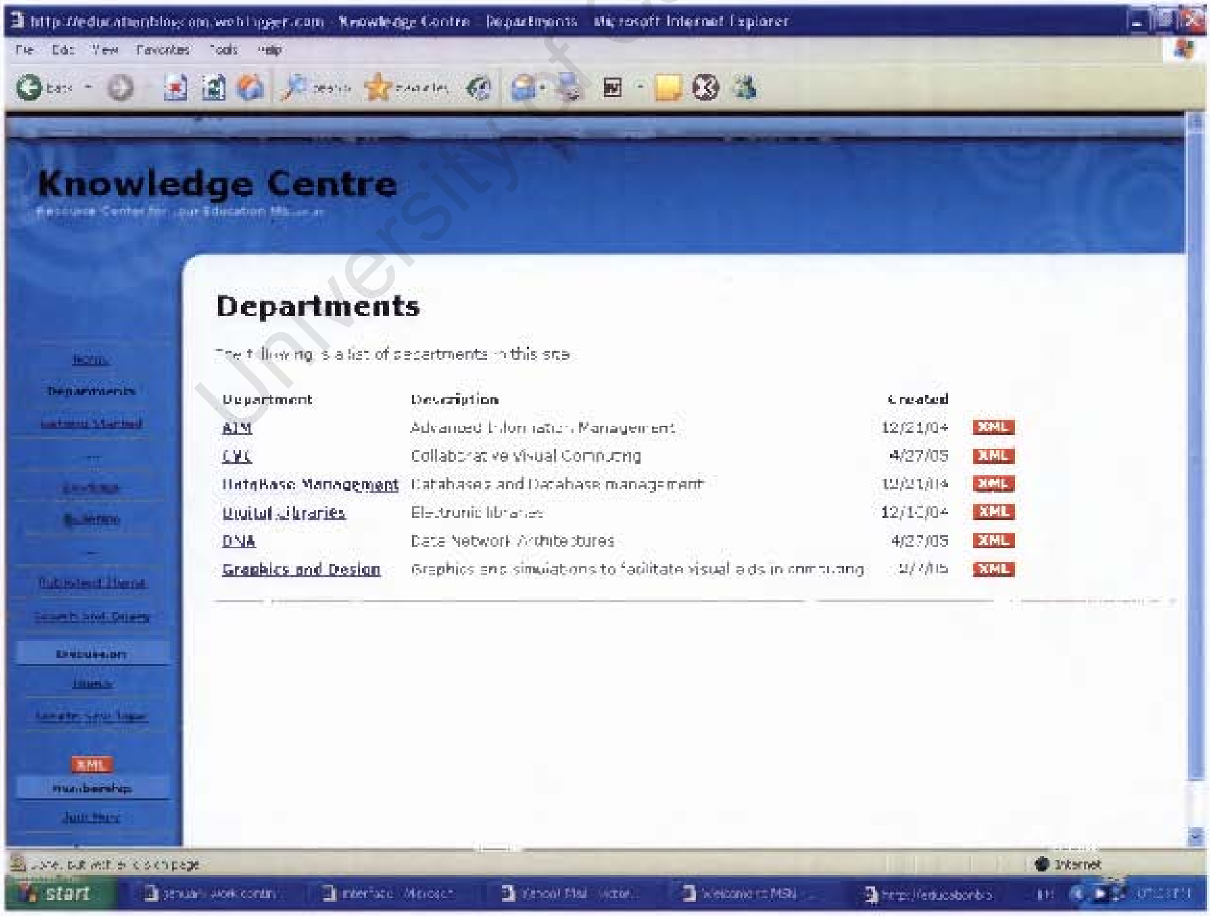


Figure 38 : Departmental groupings at UCT

Department page contains research groups. The content of each group is syndicated as shown by the xml buttons. These can be customised according to department's requirements.

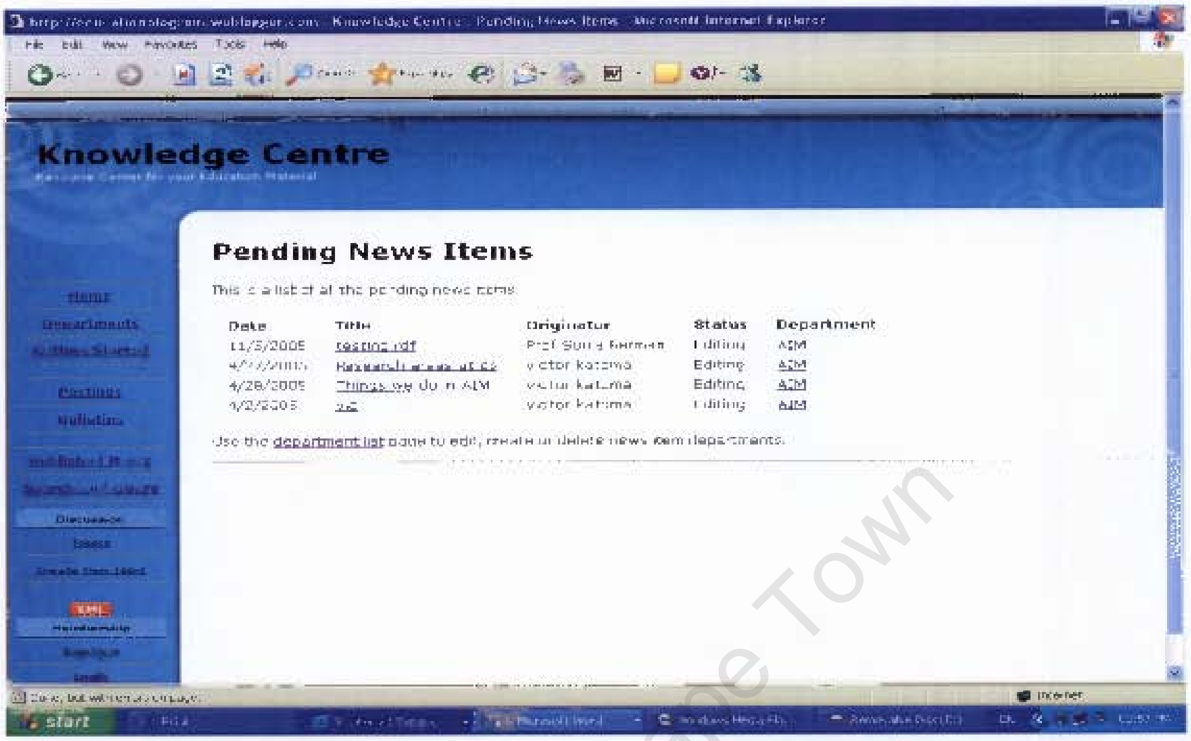


Figure 39: Items awaiting release by the editor

Figure 39 shows items that are not yet public. These could be coming from individuals from different groups.

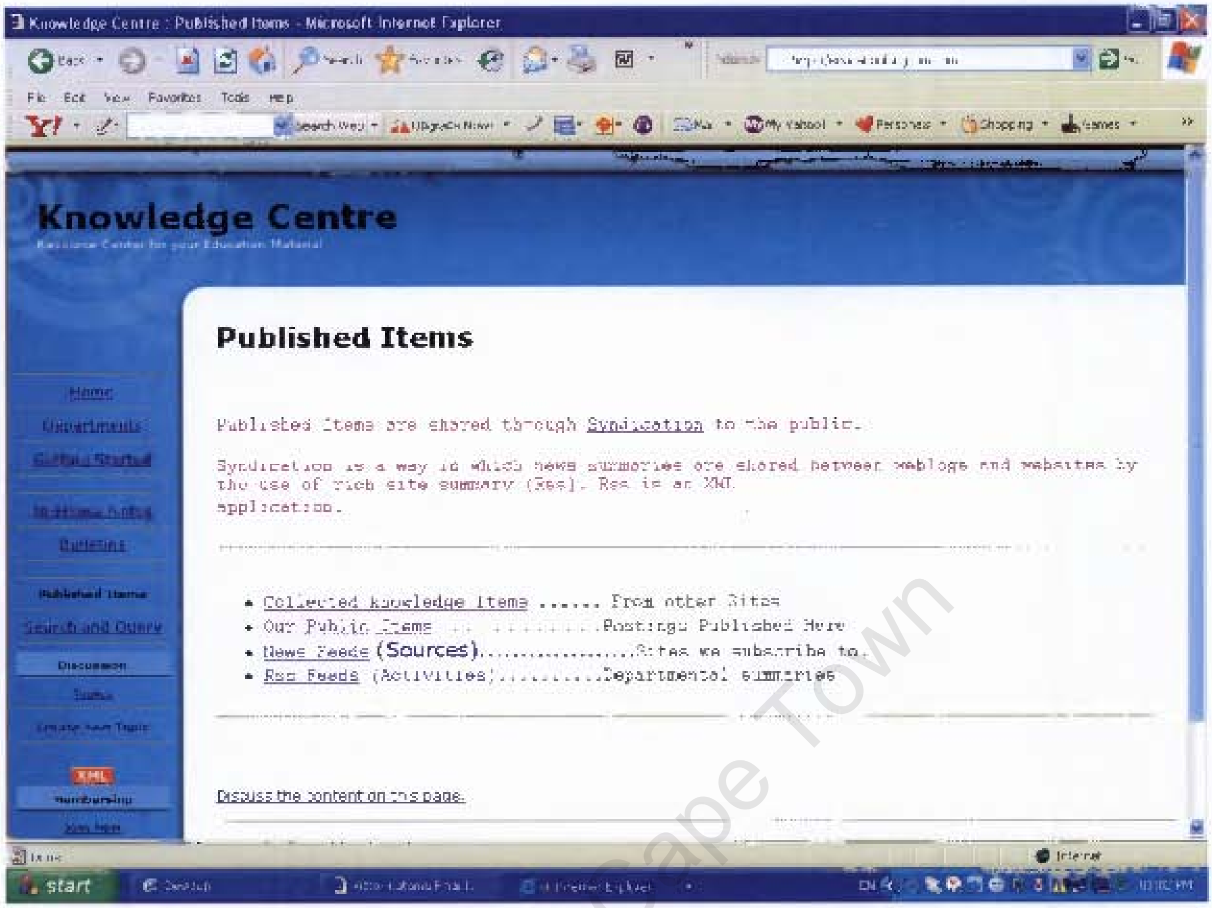


Figure 40 Published items can be accessed through this page

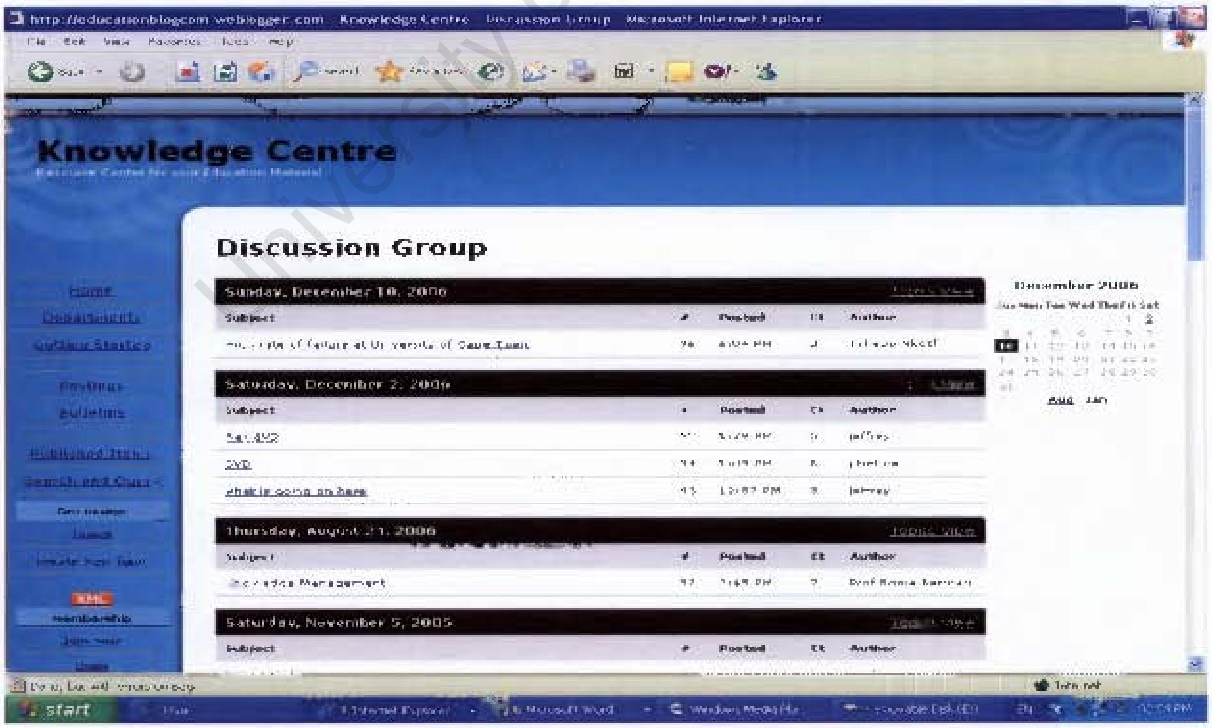


Figure 41: Discussion forum

Discussion page shows subjects and groups according to days of contribution. The items on the discussion page can also be arranged according to topic.

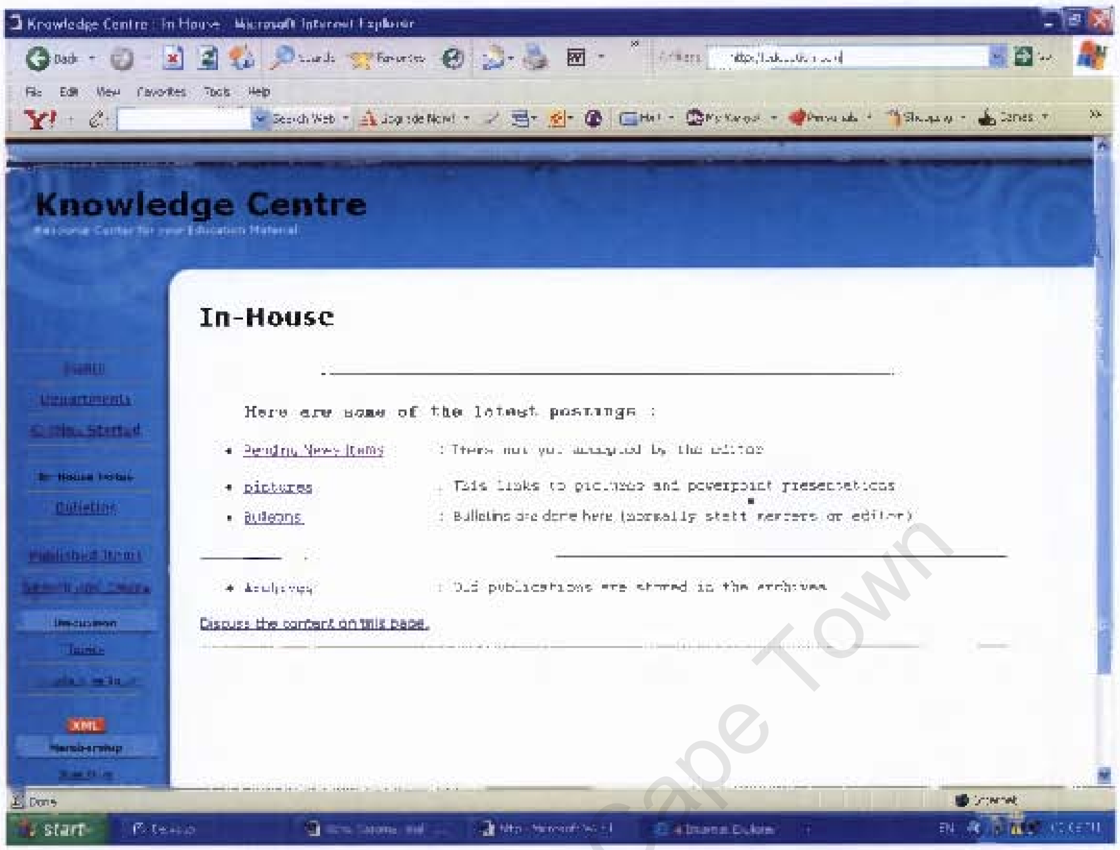


Figure 42: Posted items to be ranked and or commented on.

The In-House page displays the different links to postings made on the system.

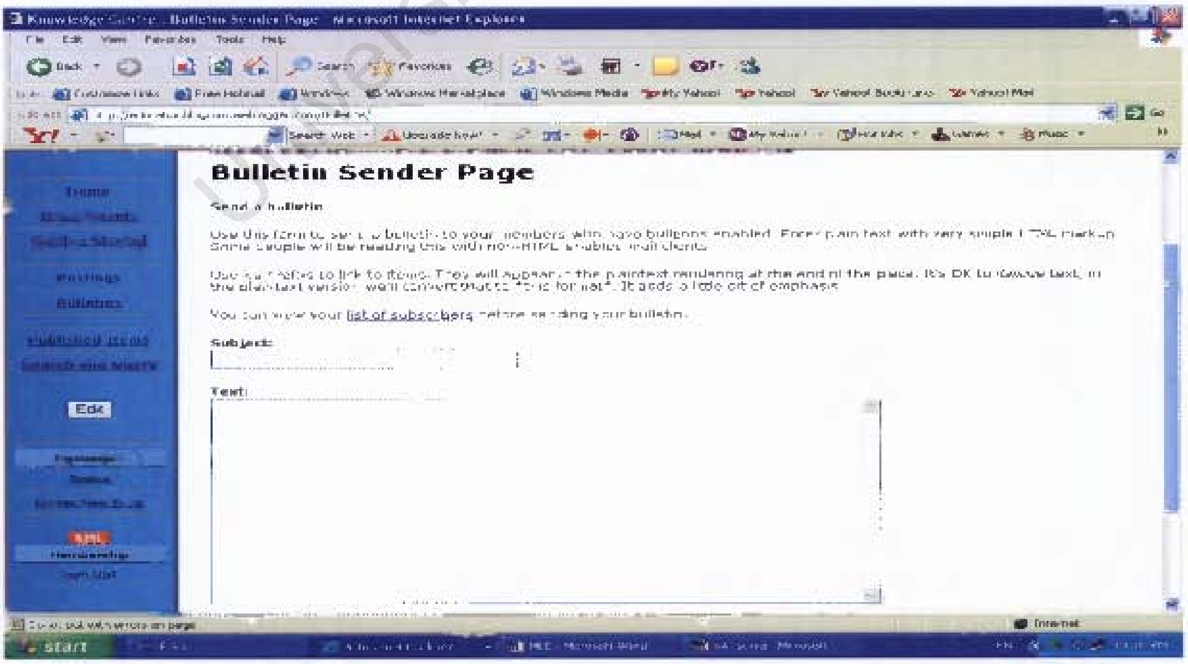


Figure 43: Bulletin page

The edit button on the navigation panel gives the editor access to the listed bulletins. The bulletin page is for the editor.

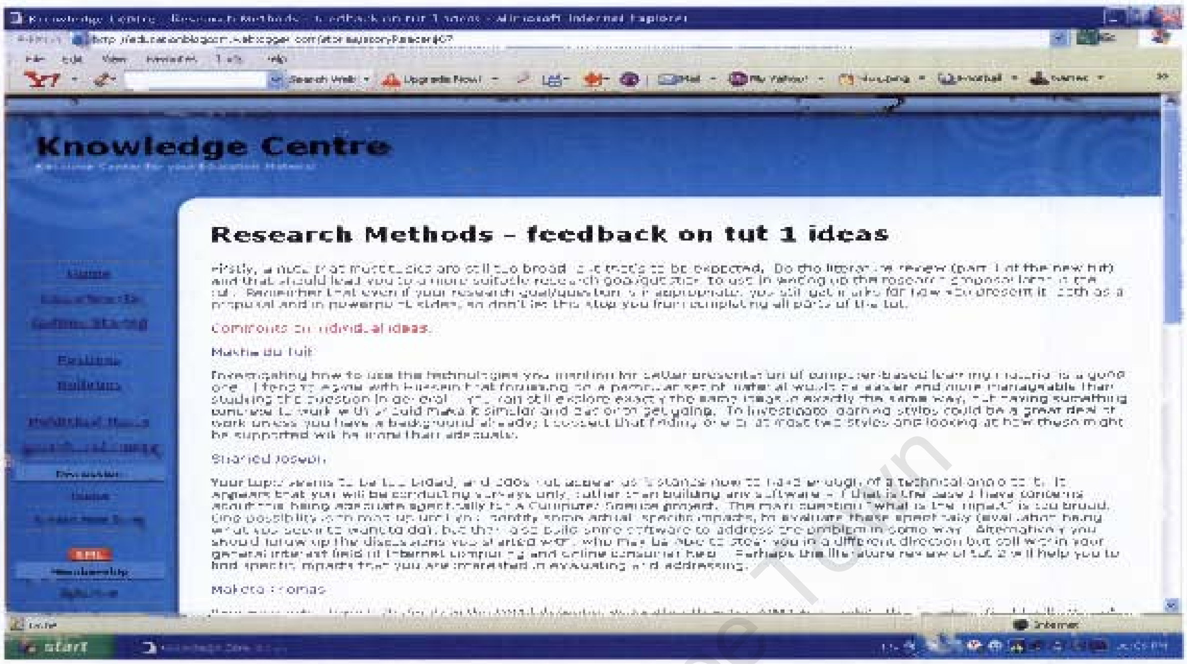


Figure 44 : In-house Notes feedback from lecturers or editor

Feedback page for Msc students who were in the process of coming up with research Topics.

To evaluate contributed items students would be required to use the voting mechanism as show in the figure below:

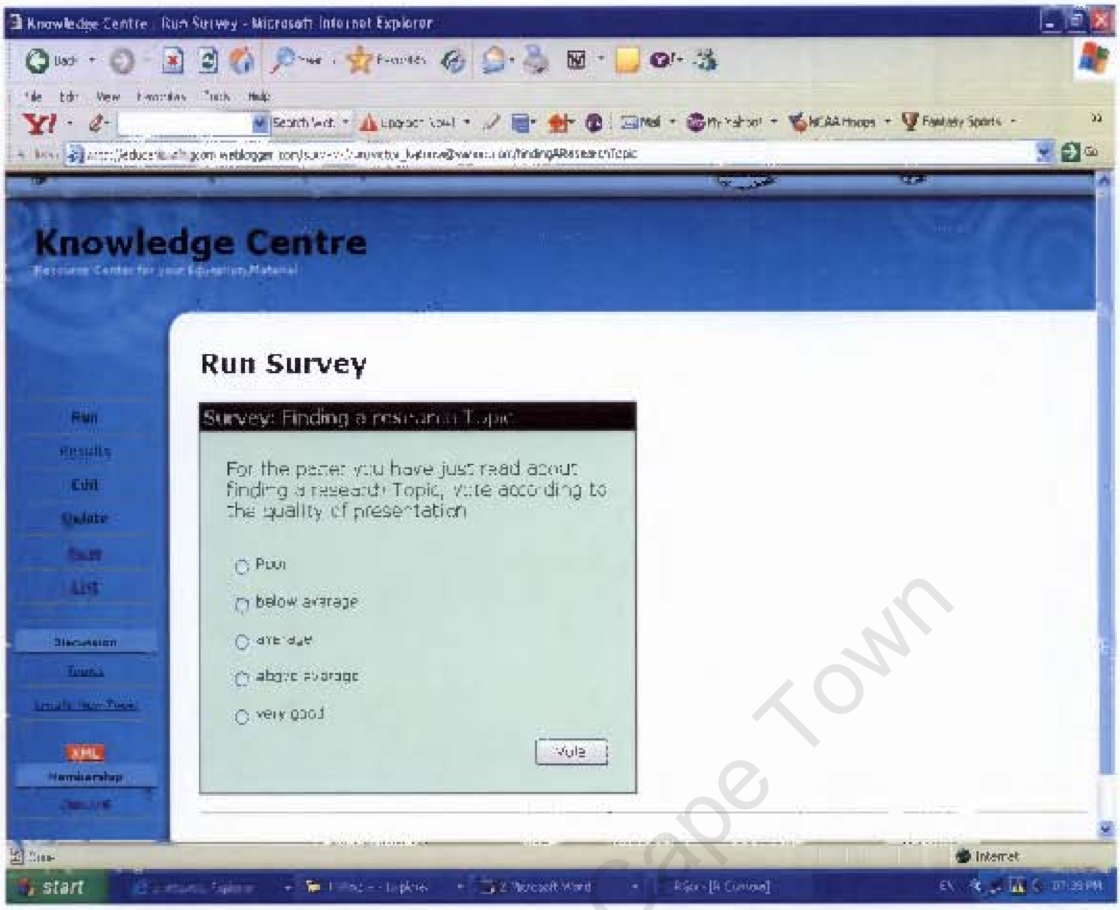


Figure 45: voting page.

Figure 46 : News aggregators' page

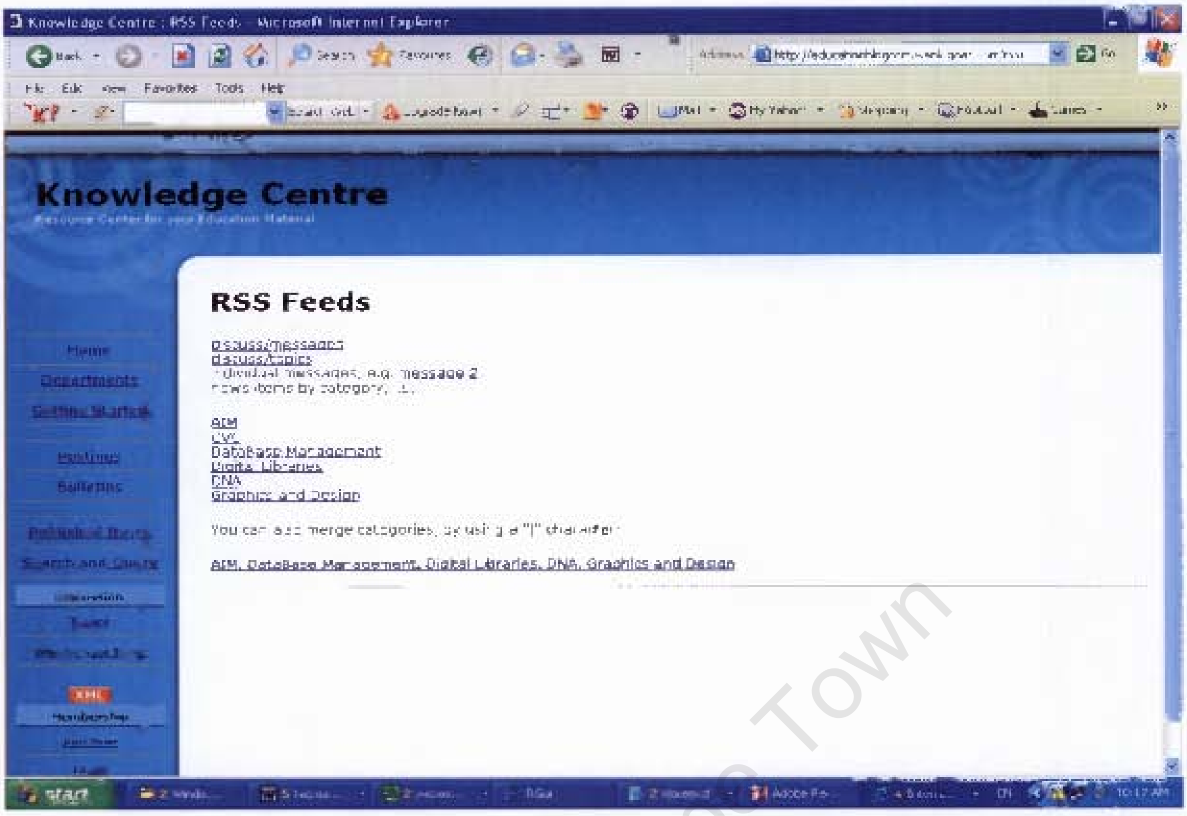


Figure 47: A collection of RSS feeds from departments

Search based on metadata: It facilitates querying for blog entries by author, subject or topic. In KC the search page (figure 51) additionally consists of the following search strategies:

1). Local Search which is a search for the in house items. 2). Global search: This also searches through the items that are received from trusted weblogs and sites.

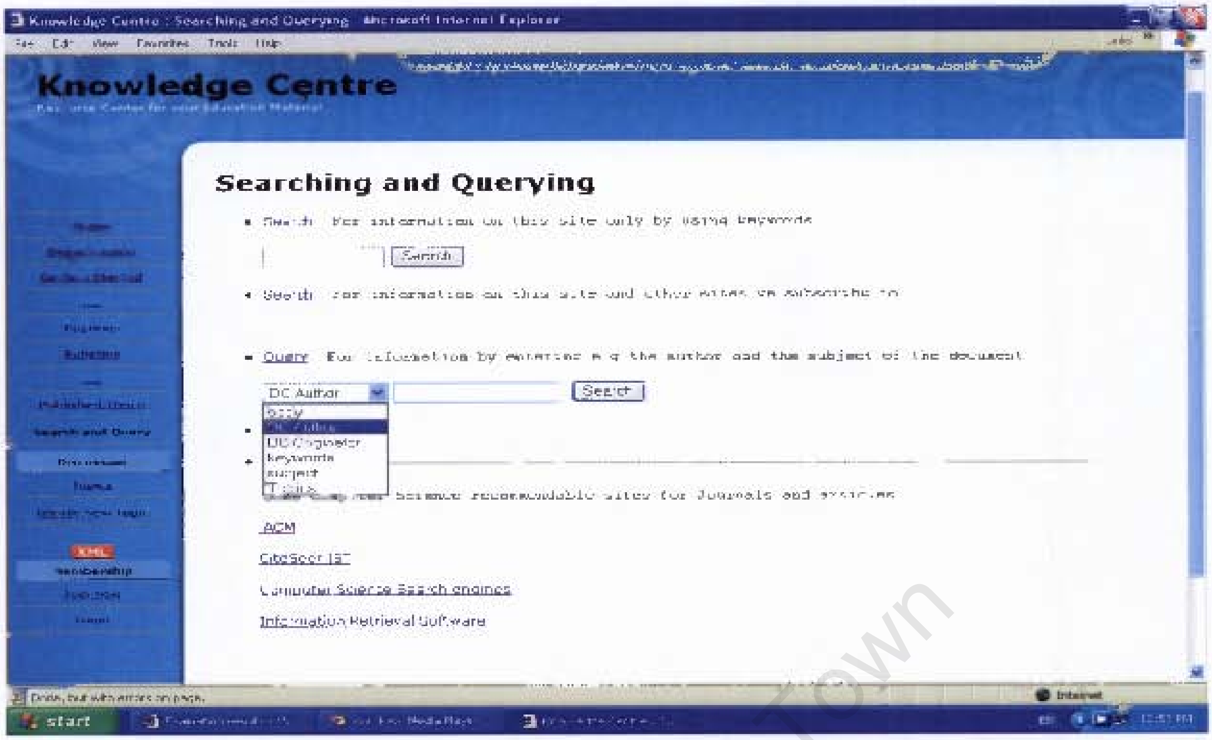


Figure 48: Query and search page

Appendix C (96): Evaluation questionnaire for MSc. students

Please answer the following questions by marking an X in the square bracket (e.g.[X]) besides the answer of your choice and where necessary giving a brief explanation:

1. In our earlier survey, we found out that, about 100% of the MSc students doing research or course work have problems in finding relevant knowledge material.
   How useful is our Search and Query page for finding relevant information?
   (a) Very useful [ ]
   (b) Quite useful [ ]
   (c) Not useful at all [ ]
   If you have any suggestions of what can be done to improve on the Search and Query page, please explain

2. At the bottom of the search query page, there are links to databases and computer science search engines. More links would be added as long as we establish they are trusted sources. Do you think, this would be useful?

   a) Yes [ ]
   b) No [ ]

3. Can you please comment on the usability (Clarity, simplicity and navigability) of the search query page as compared to others you have used? ....

4. Does the File Manager on the home page make storage and search for knowledge material easier for you compared to other ways of filing information?
   a) Yes, this File Manager is better [ ]
   b) About the same [ ]
   c) No, this File Manager is not as useful [ ]

   Explain..
   ....
   ....

We want to promote data and knowledge sharing through the open folders in our file system. You can make contributions in any of the specified fields on the folders where your friends can have access but can not interfere with your knowledge.

Do you think this would encourage people to share knowledge?

a) Yes [ ]
b) No [ ]

Explain.........

5. Does KC navigation links guide you well to knowledge material as compared to some of the knowledge systems you could have used on the web? Please explain some of the things that were interesting and those that were not

....

6. Which part of our system did you like best, and why?

7. Which part of our system did you like the least, and why?

8 Would you use this system when doing research in future?

(a) Yes, definitely [ ]
(b) Yes, but only if it was faster [ ]
(c) Maybe [ ]
(d) No, because it is too slow [ ]
(e) No, because I do not like using it [ ]

9 What things can be changed to improve this system?

....
....
....

Appendix D (96): Usability Questionnaire

Circle the statement that best fits your view of the system.

TASK 1.

a) Create a file named "storage" and save it on the desktop. Go to the home page and save this file in the folder that bears your name in the honours folder.
b) Go to the root folder. From there, find the folder called Ontologies. Open the file in there called semantics and save it in the folder named "AI".
c) Browse through the folders and complete the following:

I. Does this storage strategy make file managing easy?

Why?
....
....
Folders are very easy for storing files?

1. very easy
2. easy
3. fairly easy
4. very hard
5. not useful

Folders are very easy for finding files?

1. very easy
2. easy
3. fairly easy
4. very hard
5. not useful

Folders are very helpful for browsing the system

1. very helpful
2. helpful
3. fairly helpful
4. not sure
5. not helpful

II. Huge volumes of files can easily be handled by this strategy.

1. Strongly agree
2. Agree
3. not sure
4. disagree
5. strongly disagree

It takes less time to get the files and removes errors in searches that are based on keywords, such as spelling mistakes and poor combination of search sentences.

1. Strongly agree
2. Agree
3. not sure
4. disagree
5. strongly disagree

Please comment.

....

....

III. Does this provide a good structured way of storing and accessing files?

1. Strongly agree
2. Agree
3. not sure
4. disagree
5. strongly disagree

Any other comments about this page?

....

....

TASK 2.

Go to the Getting Started page. Familiarise yourself with this page.

I. Does this page guide you well to reaching the respective knowledge resources? Why?

....

....

The links are easy to follow?

1. Strongly agree
2. Agree
3. not sure
4. disagree
5. strongly disagree

II. On this page, open the Statistics page and open any items that were read yesterday. Then open the item that has been read the most.

Does rating help in assessing the quality of knowledge items? Why?

....

....

Rating of items helps value knowledge

1. Strongly agree
2. Agree
3. not sure
4. disagree
5. strongly disagree

Any further suggestion and /or comment on this page?

....

....

TASK 3.

Create a new discussion topic and enter any sentence on this topic, then Go to the "Search and Query page".

I. Search for the item you just created by using metadata such as topic or your name.

Do you find the query and search page helpful? Why?

....

....

TASK 4.

Open the Department page and check through the different research groups. Browse through the XML files to see the content of the research groups.

I. Does this help to guide you when looking for research groups?

Why?

....

....

Evaluation of the system as a whole.

Would you find this knowledge centre helpful if you were doing research? Please explain giving comments on the things you like and those you don't like about the system.

....

....

....

....

Did you find the system as a whole easy to use?

1. Very easy
2. easy
3. not sure
4. hard
5. Very hard

Was it very easy to find material?

1. Very easy
2. easy
3. not sure
4. hard
5. Very hard

Was it very easy to add items?

1. Very easy
2. easy
3. not sure
4. hard
5. Very hard

Is the system helpful for Knowledge Management (KM)?

1. Very helpful
2. helpful
3. not sure
4. not helpful
5. I don't know

Would you use the system for KM in the future?

1. I would definitely use the system
2. I would use the system
3. Maybe
4. I would not use the system
5. I don't know

Any other comments

....

....

Appendix E (96): Blog software features analysis chart

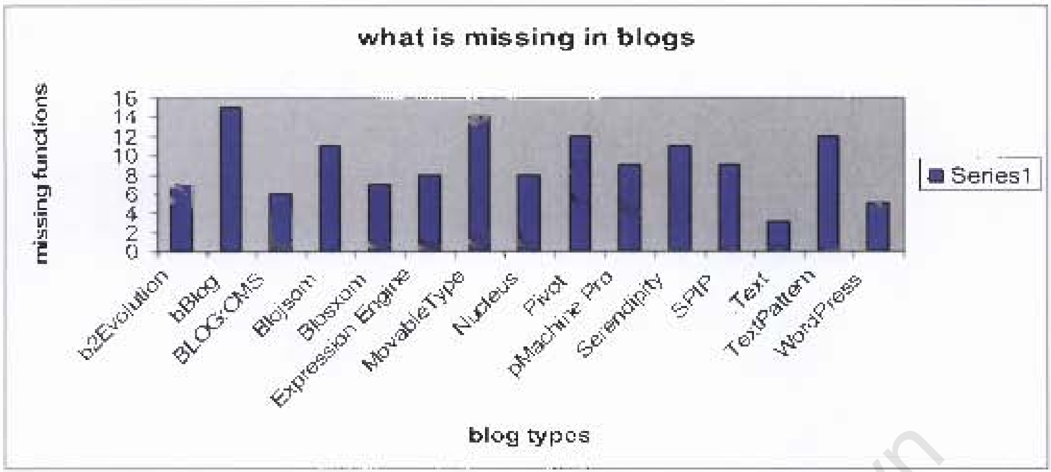
	B2Evolution	bBlog	BLOG:CMS	Blojsom	Blosxom	Expression Engine	MovableType	Nucleus
Current Version	0.9.0.8 Oslo	0.7.2	3.0 final	2.15	2	1	3	3
Multilingual	Yes		Yes†	Yes			Yes	No
Multiple Sites		No	No	No			Up to 3	No
Subcategories	Yes	No	No	Yes	Yes	Yes	No	No
Template Expert	No	No	No	No	No	No	No	No
File Editor	Yes	No	No	No	Yes†	No	No	No
CSS Editor	Simple	No	No	Simple	No	Simple	Simple	No
Cross-Post	Yes	No	No	No	Yes†	No	No	No
Password Posts	Yes	No		No	Yes†		No	No

	Pivot	pMachine Pro	Serendipity	SPIP	.Text	TextPattern	WordPress
Current Version	1.14	2.3	0.8.5-CVS	1.7.2	0.95	Gamma 1.18a	1.2 Mingus
Multiple Sites			No	Yes†	Yes		No
Keywords	No	No	No	Yes		Yes	No
Template Expert	No	No	No	No		Yes	No
Pingback	No	Yes	No	Yes		No	No
Cross-Post	Yes	No	No	No		No	No

Key:

The text below each blue box shows the functions supported by the fifteen popular blog software systems.

The chart below is a graphical representation of the analysis above. Even though .text and wordPress have more functions than BLOG:CMS, they do not support KM features such as RDF.



The figure above compares the missing functions from the different blog software systems that were considered. Manila belongs to Blog CMS (Content Management Systems).

Articles

1. SACLA 2006 pp 180-188

2. ELETE 2007 " An e-learning framework based on semantic blogging".