A Framework for Computer-Based Post-Graduate Education

S Berman and V Katoma

sonia,vkatoma@cs.uct.ac.za

Computer Science Department, University of Cape Town

Abstract

Human knowledge is now recognised as one of the most valuable assets of an institution, and in recent years a great deal of effort has been put into building knowledge management systems (KMS). The idea behind a KMS is to use technology to retain and leverage organisational knowledge. This can complement the ad hoc sharing of knowledge that occurs naturally when individuals work together, and can lead to the tacit knowledge (in people's minds) becoming explicit and sharable using an electronic system. At the same time, computer technology is common in education where it is particularly successful at managing and improving undergraduate teaching. This paper investigates how post-graduate education can be enhanced by utilising the latest developments in Knowledge Management Systems, and extending them specifically for the research training context. We propose a framework for postgraduate KMS, and describe a system we have developed around this. The framework relies on RDF semantics, a measure of knowledge quality and a simple inference mechanism to analyse and adapt the KMS automatically. Results of our initial evaluation are promising, and longer term studies are planned to obtain a better understanding of the system's potential.

1. Introduction

Knowledge Management Systems have become increasingly popular in recent years, with (Hylton 2002) asserting that over one trillion dollars have been spent on such enterprises. A knowledge management system (or KMS) 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" (Hofer-Alfeis 2003). This paper investigates how knowledge management can be used effectively as a research training tool.

We present a framework for using KMS to enhance research training in post-graduate education. Our framework is based i.a. on the use of semantic weblogs, a proposed notion of knowledge item quality, an inference mechanism and an adaptive system that stays in tune with the changing needs of its users. The framework facilitates the storage, management and sharing of knowledge. It also gives students practice in the skills needed by successful researchers – 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, etc. We believe that any institution that provides research training should offer some form of knowledge management software and include incentives to use this effectively as part of its award of the qualification.

The next section outlines the needs of postgraduate education as regards research training and we outline the foundations of KMS in section 3. We then present our framework for using and extending KMS to build a system to enhance the research training experience. In section 5 we describe the prototype system we have built using this framework, and feedback obtained from its initial usage. Section 6 concludes with an evaluation of the potential and drawbacks of our approach, and some suggestions for future work.

2. Research training in postgraduate education

Research training must impart a variety of knowledge and skills, some domain-specific and some general. Examples of the latter are reading, writing, summarizing, arguing, criticizing, finding and organizing information, etc. While this may be covered in coursework, postgraduates are expected to develop these skills primarily by tackling a project or thesis, and given few if any tools to perform these activities more effectively.

Typical problems that research students encounter are the inability to find relevant sources (such as books and articles), lack of opportunities to discuss their ideas with others, a general feeling of isolation, a lack of feedback on their work, uncertainty as to the quality and direction of their efforts, and a reluctance to write up or otherwise present their work despite the benefits of doing so. Stories abound of time wasted on poor ideas simply because these were not shared in good time with those who know better, of individuals who spend a great deal of time arriving independently at the same conclusions, and so on. While research laboratories may host seminar series and paper archives, participation by an individual is infrequent and there is little if any incentive for students to do more than the minimum in this regard.

3. Knowledge Management Systems

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.” (Davenport & Prusak 1997). Effective knowledge base construction is a hard problem (Simon 1996, Snis & Carstensen 2000). Knowledge management systems typically combine formal knowledge representations logic, semantic nets, frames, rule-based systems, concept hierarchies, neural nets, Bayesian nets, etc.) with free-form representation of knowledge suited to humans rather than automated deduction (e.g. Buckingham-Shum 1997; Fensel *et al* 1998). Such systems focus on the importance of representing multiple conflicting viewpoints, minimizing knowledge maintenance effort, incorporating meta-knowledge, etc (Easterbrook 1993, Dieng *et al* 1998). Knowledge in such systems comprises mainly documents, case studies, ontologies and formal propositions. In general, these systems tend to resemble document management systems rather than knowledge management systems.

Designing and implementing a workable KMS first requires assessment of current knowledge and needs (Coates 1999). This is called knowledge auditing (Hylton 2002). From the knowledge cycle in figure 1, the following points are identified as critical to KMS, and were the main drivers for our postgraduate-KMS framework:

- knowledge should be collected in a systematic way
- knowledge should be organised in flexible structures capturing all relevant attributes
- the system should be able to make inferences
- the system should have an ontology for common understanding
- knowledge retrieval should be easy and effective.

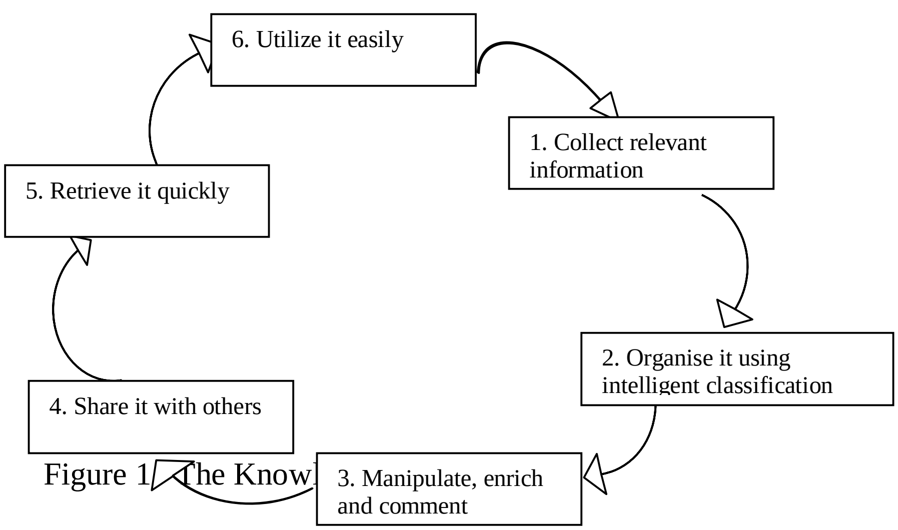

Figure 1 The Know

4. A KMS-based framework for research training

This section explains what our Postgraduate-KMS framework hopes to achieve, and then outlines the data model and 8-step process that together make up the framework proposed here. Finally, we present our notion of knowledge quality measurement and its use in making inferences that govern system adaptation.

4.1 Goals of the framework

In developing our framework for a postgraduate education tool, we strove to meet the requirements of a successful knowledge management system, while also incorporating the skills and addressing the problems involved in learning to do research. We therefore aimed at a system that would be easy to use, and that provided added value to make its usage worthwhile. The tool needed to embody the activities research students need to practice, particularly finding relevant information, organizing work effectively, reading, writing and critical analysis. The ability to measure or evaluate knowledge items as well as human contributors would be an added bonus in a teaching environment.

4.2 The data model

The framework is 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* (i.e. a posting or message, usually one fact or idea contributing tacit knowledge) and the *file* (representing either the creator's own work or a reference they found). 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.

Each of the three main types of object can be associated with secondary objects that add value to the object – *ratings* given by its readers, *comments* made by its readers, and *metadata* such as the individual who entered it into the system and the date on which this was done. These secondary objects all add value to the knowledge base, and are also useful in computing knowledge quality measures. They can be used to query the system in new ways, and for inferring how best to adapt the system, as described later.

Objects in the model are associated with each other via directional links. Any secondary object can be related to any source, snippet or file; and snippets can reference any other object in the model e.g. 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.

4.3 Weblogs as the basis for a KMS

For the tool to be as easy to use as possible, and to fit as naturally as possible into the working habits of research students, it was decided that a weblog (Kumar *et al* 1994) would be the best underlying technology. Weblogs or blogs are extremely popular (Kelvin, Desoze & Evaristo 2004), deceptively simple in appearance, and very inviting because of the informal manner in which contributions are handled. A blog typically comprises a list of snippets which can cross-reference each other. While blogs are predominantly about snippets, they can easily be extended to include the files and sources in our framework as well.

Exchanging summaries is a way of sharing across blogs. Published summaries are displayed in a common format called rich site summary (RSS). RSS is an export and interchange format (Downes 2002), the last two versions of which are also called resource description framework (RDF). While RSS is characterised with merely transferring information, RDF offers more structure through the triples (source, predicate and object). RSS feeds outline the knowledge resources of a blog and an aggregator collects these from different blogs and displays them in one list. Such aggregation not only allows relevant items from external sites to be easily incorporated, but also allows a community to have separate blogs that can feed off each other in this way. This can be useful for postgraduate education if, for example, different research groups have separate blogs and still share some snippets such as those relating to research methodology.

Weblog development tools permit rapid system development and easy customization. One particularly attractive feature is the ability to send and receive information via email. This makes the system unobtrusive, and allows users to quickly and painlessly switch between whatever they are working on and the KMS system.

Semantic blogging (Cayzer 2004) has recently received considerable attention. This essentially involves the capture of metadata along with snippets, and using this to enhance navigation, querying and display. Weblog construction tools support RDF generation, making semantic blogging a relatively easy extension to incorporate. The semantic weblog thus forms the foundation upon which our framework is built.

4.4 The framework

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

1. perform a knowledge *audit* to assess strengths and needs in the organisation

2. use a rapid-development *weblog construction* toolkit to create a simple weblog where snippets and files can be saved, shared and searched. Ensure that metadata including Dublin Core is captured and usage statistics maintained from the outset

3. *refine its interface* in an iterative fashion, based on feedback from users

4. extend the system to obtain commentary and *ratings* of snippets, files and sources

5. introduce *incentives* to encourage extensive use of the system

6. compute *quality measures* and use simple inference to predict future usage

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

The idea behind ratings in a KMS is to encourage those who consult the system to indicate the perceived accuracy and usefulness of snippets they read. A rating can simply be a number chosen from a semantic differential scale of 1-5, or can be selected from a predefined menu (e.g. agree, disagree, cannot-understand, irrelevant, etc.). Comments are textual, indicating how the ideas in a snippet could be expanded, improved upon or corrected. The name of the person making the comment should be included automatically, to ensure that comments are relevant and constructive.

One way of incentivising students is through a Research Methods module which contributes some percentage of their final mark, and/or has to be passed before their thesis is examined. Such a course can include a component which assesses how well they have used the KMS – e.g. how many snippets, comments and ratings they have contributed, and the quality of these and of their comments in the system evaluation. An adaptive system is one which learns from user behaviour over time, and reorganizes itself accordingly. This requires a quality measure and an intelligent software component. These aspects of the framework are outlined in the sections that follow.

4.5 A knowledge quality measure

Quality measures are needed by an adaptive system as a basis for understanding the current system and predicting how it can best be reorganized. We define the *ranking* of a knowledge item (snippet or file) as the average rating given to it by system users, normalized so that it is a fraction between zero and one. The *status* of a knowledge source is the average of its standing (expertise level as entered into the system when it was introduced) and its reputation, after each has been normalized to a fraction between zero and one. Reputation is the average ranking over all the files/snippets contributed by that source. The *popularity* of a knowledge item is the weighted average of its reading-quotient (the fraction of reads which have been reads of this item), its linking-quotient (the fraction of snippet-links which point to this snippet) and its comment-quotient (the fraction of snippet-comments/discussion-replies that are about this snippet). In a system where comments are not associated with a negative/positive indicator, comment-quotient may be weighted less than ranking or status, and possibly zero. In postgraduate education, staff can provide source status values, and their ratings of items can be weighted more heavily.

The *quality* of a knowledge item is the weighted average of its ranking, its popularity and the status of its source. A weighted average is used so that the system can be tuned to put greater or lesser emphasis on specific aspects. 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.

4.6 Inference and system adaptation

A KMS becomes more than a knowledge repository if an inference component is added, especially if the system is able to adapt itself based on what it learns. Even simple analysis can discover which knowledge items or collections are most popular, most contentious or of highest quality, and knowledge can be re-ordered on a page to take advantage of this. It can also detect users with similar usage patterns and add interest-groups to the KMS accordingly, notifying the people concerned by email when this occurs. With simple statistical inference, forward-looking predictions are possible and can be used to improve the system. For this we advocate the use of MLE or Maximum Likelihood Estimation (Crooke, Froeb & Tschantz 1999).

Given a known probability distribution with parameter p, and a sample of values from this distribution, MLE finds the value for p that is most likely to yield these sample values. As a simple example suppose we have a coin which has probability p of turning up heads when tossed, and a set of results of N tossings of the coin where H were heads (and N-H were tails), and we want to find p i.e. the bias of this coin. We take the probability distribution (which is in terms of p), maximize the likelihood of observing H heads from N tossings over all possible values of p, and find which value(s) of p yield this maximum. To find these value(s) of *p* which maximize *F(H heads from N tosses| p)* we differentiate *F* and solve for *p* where the derivative is zero.

To illustrate how MLE can be used to infer new information about a KMS, consider a simple situation similar to the coin example above. Suppose that all knowledge items are ranked on a Boolean scale i.e. as either helpful or unhelpful, and that a source has had *N* of its postings rated of which *H* were rated as helpful (and *N-H* were rated as unhelpful). We can differentiate *F(H helpful ratings from N ratings | p)* and solve for *p* when the derivative is zero to find the most likely value for *p* that would have given rise to these ratings – i.e. to determine the most likely helpfulness value of this knowledge source. A system with this facility could then adapt over time so that contributions from the most helpful sources appear above others, or are collected on a special page, etc. Using the same idea in more complex situations, it becomes possible to determine “hot topics”, “best items”, etc. and reorganize the system accordingly.

5. KC – A prototype implementation of the framework

This section outlines KC, the Knowledge Centre, a prototype we built based on the framework. Our first action was conducting an audit of 11 postgraduate students to discover current practice and needs. This showed that, while several structures were in place as a result of research laboratory activities, many students felt isolated, sharing was limited, knowledge was mostly saved in unstructured or very poorly organized ways, and the overwhelming majority had difficulty finding relevant knowledge. All 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.

The Manila content management system was used to create the initial Weblog. Its repository is an object-oriented database system called Frontier, in which all contributions and scripts are saved. This software is free and facilitated rapid system construction, while also providing many sophisticated features we required. KC currently has several different kinds of page, an example of which can be seen in figure 2. These range from blog-like chronological lists of postings, to discussion pages, search pages and directory pages (where files can be arranged in hierarchical structures at the discretion of their owner and be readable by all). Blog pages include syndicated items (contributions from our users that are also syndicated elsewhere in the public domain via RSS) and aggregator pages (comprising summaries of items on other sites to which the KC editor subscribes). Links to bibliographic databases and useful sites also appeared.

The system was made available to Masters students who were part of a coursework-and-dissertation programme, and one of their assignments included a mark for usage and critique of KC. The incentive of just a few marks in just one assignment proved most effective in ensuring that students collected their work on KC, looked at the contributions of their classmates, entered into discussions and commented critically on what they saw. This could perhaps be because this part of the assignment was perceived by students as one where marks could relatively easily be accrued. However, since part of the marks were for quality of critical comment, this was in fact not the case. The exercise thus proved successful both as an incentive to use KC and as a method of assessing students.

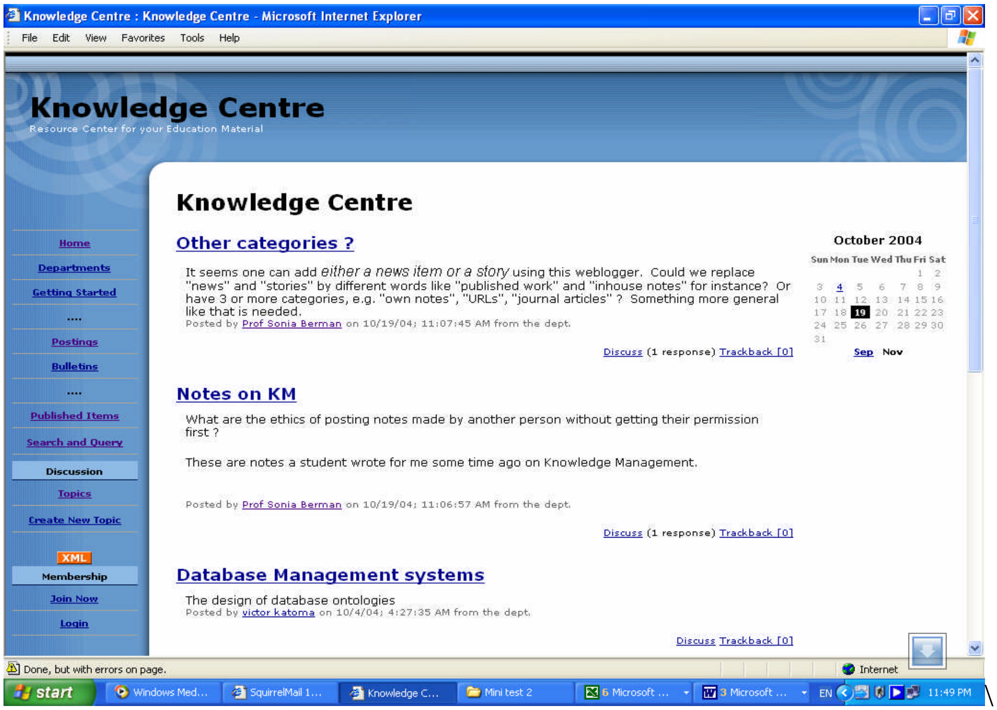

Figure 2. Interface showing calendar and syndicated items.

Our weblog-based KMS was subsequently extended to capture Dublin Core metadata along with all postings, and the query facility was extended to incorporate this. Since Manila provides for RDF generation automatically, this was easily achieved. The computation of quality measures, and their use in maximum likelihood estimation to infer new knowledge, has been tested in two ways – to automatically determine groups of students with similar interests, and to determine the ranking of individual students, in a manner similar to the example given earlier where helpfulness is determined based on ratings of items posted. The final evaluation by the nine student users indicates that they found the system easy to work with (all respondents) and useful (all but one were certain that they would use KC in future), but it is clear that substantially more usage is needed before major benefits can be evaluated, since in particular the maximum likelihood estimator needs enough real data to work on before its impact can be determined.

6. Conclusion

We believe that a knowledge management system (KMS), coupled with appropriate incentives to use it wisely, should be part of all research student training. To make it quick and easy to use, a weblog is the best basis for such a system. To deliver added value that makes it worthwhile, knowledge quality measures, inference and selfadaptation are advocated. This paper has presented a framework for constructing such a knowledge management system for research training of post-graduate students. It also described KC, a prototype implementation of this framework.

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. Our 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.

References

Buckingham-Shum, S. (1997). *Balancing Formality with Informality: User Centred Requirements for Knowledge Management Technology*. AAAI '97 Symposium on Artificial Intelligence in Knowledge Management 1997, Stanford University, CA, AAAI Press.

Cayzer, S. (2004). Semantic blogging and decentralized knowledge management. *Communications of the ACM 47(12*), 47-52.

Crooke, P., Froeb L., & Tschantz, S. (1999). Maximum Likelihood Estimation. *Mathematica in Education and Research 8(1),* 17-23.

Davenport, T.H. & Prusak, L. (1997). *Working knowledge: How organizations manage what they know*. Harvard Business School Press.

Dieng, R., Corby, O., Giboin, A., and Ribiere, M. (1998). Methods and tools for corporate knowledge management in Gaines, B.R. & Musen, M. (Eds.), *Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW'98*). Banff, Alberta, Canada, April 18-23.

Downes, S. (2002). *An introduction to RSS for Education Designers.* www.downes.ca/files/RSS_Educ.htm.

Easterbrook, S. (1993). *Domain Modelling With Hierarchies of Alternative Viewpoints*. IEE International Symposium on Requirements Engineering.

Fensel, D., Erdmann, M., & Struder, R. (1998). *Ontobroker: The Very High Idea*. Proceedings of the 11th International Flairs, Conference (FLAIRS-98), Sanibal Island, Florida.

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

Hylton, A. (2002). *A knowledge management initiative is unlikely to succeed without a knowledge audit*. 1 – 13, www.annHylton.com.

Desoza, K. C., & Evaristo, J. R. (2004). Managing knowledge in distributed projects. *Communications of the ACM 47(4*), 87-91.

Kumar, R., Novak, J., Raghavan, P., & Tomkins, A. (2004). Structure and evolution of blogspace: *Communications of the ACM 47(12*), 35-39.

Snis, U., & Carstensen, P.H. (2000). *On Knowledge Management: A Field Study in Internet-Based Organisational Memory and Knowledge Management*, Edited by David G. Schwartz, Monica Divitini, Terje Brasethvik, Idea Group Publishing .