Conceptual Approaches for Personal and Corporate Information and Knowledge Management

Andrew J. Slade & Albert F. Bokma
Centre for Electronic Commerce, School of Computing, Engineering and Technology, University of Sunderland, St. Peter's Campus, Sunderland, SR6 0DD, United Kingdom
{albert.bokma, andrew.slade}@sunderland.ac.uk

Abstract

There is a growing awareness in enterprises about the need for adequate Knowledge Management (KM) in the organisation as a key factor to sustainable competitive success of enterprises. As working patterns are rapidly changing and more flexible organisational structures are emerging, the individual "knowledge workers" have to keep track of information and knowledge in a fast evolving world characterised by information overload. What is required are generic and general purpose approaches that allow users to record a variety of knowledge together with its association to relevant information sources and to share it with colleagues in an intuitive way. An approach to this problem, using conceptual structures combined with the use of ontologies is proposed and discussed and an example of its application to the sharing of research knowledge in a University is presented.

1. Introduction

The almost exponential proliferation of information, evident in recent times, is starting to cause considerable problems both for individuals and organisations. As a consequence there is both a need to effectively manage existing collections of information and access and import new information relevant to a particular task. While a considerable amount of information tends to be text documents in proprietary formats, they do not lend themselves to easy and intuitive management as collections of documents. Although the use of proprietary text processing systems, and consequent file formats, is likely to continue to exist for some time, recent developments are offering an in many respects superior solution to more integrated information management in an organisation such as through the use of XML.

Recent times have seen dramatic changes in organisational patterns, as static and hierarchical structures are rapidly disappearing and more flexible project-team approaches are emerging that bring together a group of individuals for a project and disband after the project has finished only to reorganise with a different membership for other teams on other projects. This pattern of work, whilst perhaps new in most commercial organisations, is the standard approach in most Universities. The academic staff of a University are expected to be individual 'knowledge workers' and to share there work with the wider academic community. This is normally done through the usual academic publishing route but this process does not meet the corporate needs of the University. Every University needs to 'market' itself to a wider community who does not necessarily share the goals and approach of academic staff.

As knowledge sharing is important at the enterprise level, approaches are required to facilitate the recording of knowledge and encourage effective sharing of it with colleagues. To support all this adequately the tools have to be very intuitive and easy to use and the progeny and use of knowledge readily ascertainable.

Knowledge Management has received widespread interest and a number of approaches are emerging. Thus a significant effort for instance has been invested in process oriented solutions. The present work distinguishes itself and concentrates on an integrated information and knowledge management approach largely independent of processes that allows the individual to effectively create an information and knowledge map or network for their personal needs and as a basis for sharing with other users. In this paper ongoing work on a general-purpose knowledge management concept is being presented, drawing on important developments in conceptual graph and ontology technologies. An example is presented that illustrates the basic approach and highlights some of the difficulties faced by implementers of such systems.

2. Information and Knowledge Management Needs

A considerable amount of information average users deal with in their everyday activities is contained in document files and increasingly in HTML documents. In fact, we would contend that most users have several hundred document files on their computers, more than half of which they could not tell what they contained. In the course of their activities users typically reuse existing documents and import new ones. As the users engage with these documents and produce new ones a considerable amount of knowledge builds up around these documents. This knowledge can be about how these documents relate to one another as well as deeper insights about the content. This knowledge typically resides in the mind of the user and never gets recorded. Another interesting feature in this respect, we would argue is the fact that as time progresses people forget this knowledge unless they keep referring back to the same documents time and time again. They may remember some useful information for the task at hand but unable to quickly lay their hand on the relevant information and have to spend time and effort in order to retrace their thoughts. In addition, few people work in isolation and need to interact with colleagues in the organisation. In order to effectively collaborate and avoid duplication of effort they need to communicate or maintain comprehensive inventories to ensure that the same work has not already been done elsewhere. A number of approaches and technologies exist to help users manage the storage and retrieval this type of information. The most basic of these is the filing hierarchy, as well as searching algorithms and document management systems.

In devising a filing hierarchy, the organising principles are typically a combination of thematic and organisational categories, and few users are able to devise and correctly use failsafe hierarchies that allow them subsequently to effortless locate documents. The filing hierarchy also suffers from the problem that it forces the user to create only a single categorisation of the entire set of files. What is frequently needed is a set of alternative strategies, where a particular document is classified by subject (what it is about), organisational context (i.e. the project it originated in) and by author, to name but a few. In this way the user could locate their document thematically by browsing the subject hierarchy, or find it by looking in the organisational hierarchy and the project it is associated with, or just look at only their documents by following the "people" viewpoint. Another advantage of such a strategy is that it potentially allows the user also to see other documents that fall under the same category and thus to find related documents.

Managing documents is not just a matter of correct classification and cataloguing and nor is the management of associated knowledge. Though search-based information management and retrieval techniques have their uses it often depends on the skill of the user to correctly locate and retrieve documents. One cannot help but feeling that this is analogous to geologists drilling for oil, who, while being able to rely on some scanning techniques, in the end also need a bit of luck to hit on a productive site. Website searching facilities are a typical example of this problem and who frequently provide bad and patchy results leaving the more ingenious user to navigate the site manually in order to get to the documents he knows to be there.

Document management systems have a number of benefits for maintaining large collections of documents especially where these are shared amongst a workforce and which help to serve a varied need such as archival, document tracking, locking etc. to ensure that documents are protected and accessed in an organised fashion and that there is nor proliferation of copies of the same document that could lead to confusion. At the same time they do not adequately support our needs for providing alternative access paths nor do they provide for intuitive ways to record the interrelation of documents and associated knowledge. In order to manage documents in a more intuitive fashion, there is a need to go beyond considering documents in isolation and to represent the context in which these documents fit. This would allow both direct access to any associated materials as well as making explicit any relations between them. This approach would also be able to convey the context in which documents and records of knowledge fit and which is vital for their correct interpretation in a shared environment.

The need for supporting both the *individual* and the *organisation* is an important aspect frequently missed by existing solutions as they tend to address the needs of the organisation at the expense of the individual, thus necessitating more drastic organisational strategies such as incentives and policies to provide incentives to the individual to actively participate in knowledge capturing and sharing. This is far from satisfactory and any adequate solution should be actively and substantially support the individual in their work and allow them to record information sources and the knowledge generated around them. At the same time, the effective sharing with colleagues also needs to be supported to capitalise on the assets created by individuals and teams. Shared components should be stored in a generally accessible location and in an intuitive navigable form that allows easy location and retrieval of useful information.

3. Ontologies and Conceptual Graph Based Approaches

Philosophers and psychologists have for some considerable time recognised that human memory is associative [1]. Allied to this is the suggestion that human beings tend to understand and manage reality by devising abstract models and classifications that are used to categorise observations and interpret them [2]. Following this direction, information and associated knowledge should be categorised in a conceptual fashion through the use of appropriate classifications, where relations amongst items of information (e.g. documents, files database records) and any knowledge associated with them can be recorded. Following our earlier point about the need to maintain the context in which both information and knowledge occur as a necessary basis for their correct interpretation, especially by third parties, this would also allow expression of the context in which these documents and associated knowledge fit. This lesson is increasingly becoming evident in the field of knowledge management as can be seen from [3]. Consequently, it is proposed that the information and knowledge management strategies of the future should follow conceptual principles and allow the user to catalogue and access in a graphical way. This requires a model of the organisation and of the respective domain and to associate documents with elements in this model. In this way the context in which the documents fit and how they relate to other documents can be expressed.

In recent years there has been a growing community working in the field of ontology and conceptual graph research [4] and [5] leading to interesting developments which, in our view, can usefully contribute to a solution to the type of problem outlined. Ontologies, as used in AI, are conceptualisations of a domain that allow the user to build a model and reason about it [6]. Formal models of a domain can be built and shared amongst users. A number of ontology based approaches have been developed (see amongst others [7], [8], [9], [10]) where the main motivation has been to support the sharing and reuse of bodies of knowledge in a computational form [11].

Ontologies can be used to develop a detailed model of a chosen domain (fine grained). However, we would like to use them in a more schematic fashion (coarse grained) and which should result in a much smaller overall size of the ontology and the resulting conceptual structure whose complexity could otherwise get out of hand. We suggest instead to develop sets of smaller ontologies to represent different classification viewpoints such as an abstract subject hierarchy, or an organisational hierarchy and which will offer alternative access paths to documents and items of knowledge. The resulting ontology would therefore serve as a document organisation principle and as a model to represent the world of the organisation. Some work in this direction is already under way in the KA2 initiative that is currently developing a number of ontologies used for annotation of documents amongst a research community [12]. However, while their approach will offer an ontology-based information and knowledge management functionality the connection between ontologies and related documents and knowledge is not encoded and available in a persistent graphical-browsable form. Instead, connections need to be made by an intelligent web-crawler following a user query. Coming back to our previous point, this is again a query-based approach whereas we propose to allow users to create maps/webs of information and associated knowledge and make this available in this form to other users in a shared environment. At the same time, ontology-based approaches are in principle appropriate for the purposes of developing organisational specific or domain specific classifications and could be used to categorise and associate items of information and knowledge.

For our purposes it is important to be able to create arbitrary connections between items, to represent relations between them, and which the user can annotate. This should be supported by a graphical navigation capability where the user can browse the information and knowledge space so created by navigating a graph-like structure rather than by conventional searching techniques and presentation of results in tabular form. We foresee that the problem for the average user will be that the formalisms are rather technical assuming familiarity with frame-based approaches and therefore not sufficiently intuitive. Using ontologies to devise detailed models of in a fine-grained form will require specific training and a logic/formal mindset to develop and will be impracticable for the average use case. Although there is a place for them, this goes beyond what we believe is strictly required here, and which will be too laborious to build up and maintain. We therefore propose to use a pre-specified custom-built ontology, developed for a domain or enterprise, and which provides a suitable core taxonomy to which the user can attach items of information and annotations containing knowledge (knowledge objects), linked to the documents they relate to and likewise attached to the core taxonomy.

Ontology research has originally been driven by the motivation to jointly develop definitive global and detailed conceptualisation of a domain. It has been questioned by some such, as [13], whether this is in fact realistic and achievable. Instead they propose to allow the user to develop personal, evolving ontologies and not only make changes but annotate these with the rationale behind the changes. APECKS is an example of this and provides the user with tools for the incremental building of

ontologies of this nature. In line with this we suggest the use of extendable taxonomies as a basis for categorisation of items that allows the user to correctly classify any new items of information and to annotate them with any reflections, criticisms or knowledge gained. The taxonomies should include a standard repertoire of cross-referenced organisational and subject based categorisations. There is at the same time a need for a common core ontology specific for an organisation to facilitate the effective sharing of information amongst workers and which should be capable of being browsed graphically by the users. The common taxonomy will also provide for a sharing strategy between users, who are familiar with the taxonomy and thus able to easily explore the taxonomy and associated information and knowledge of their colleagues.

Though we favour the use of a core enterprise taxonomy as opposed to a full-scale ontology based system, this is not to say that an ontology-based system would be wholly inappropriate at some point. A distinction needs to be made between using a core ontology or taxonomy (coarse-grained) for the purpose of cataloguing from the use of ontologies of a detailed conceptualisation of a large domain (fine-grained) and which would be unworkable for the majority of users in an average organisation as envisaged here. For information management purposes it is also important that the conceptualisation is not isolated from the domain it is modelling, to allow the user to manage a set of information sources. Though it could be argued that the resulting framework is in its totality an ontology we prefer to make a distinction between those parts of the conceptual framework that are used by the user for the classification of items from the items themselves, their interrelations and associated knowledge objects. While the former fulfil the role of an abstract model in which to order items in the universe of discourse the latter are not abstract and normative but real and descriptive and are arbitrarily connected by the user and thus of a different type. It would therefore appear to be reasonable to keep that distinction.

Conceptual graphs are an alternative approach that allow models to be created. While ontologies are usually expressed in a frame-based formalism, conceptual graphs can generate equivalent models by specifying a number of nodes representing concepts and which can be interconnected by a number of arcs representing how these concepts are related to one another. Conceptual graphs as computational models were first proposed by [4] and more recently revisited in [5]. These have the benefit that the models generated can be encoded in predicate logic, KIF or a display format that graphically shows the respective concepts and their relations. The latter has the benefit of a graphical representation that would appear more intuitive for the average user to comprehend and has the advantage that automated reasoning techniques can easily be employed on the predicate logic format. This will in principle be implemented in the conceptual framework but with some modification as for example there is a need to directly connect the model to the set of files that are to be modelled. On the other hand it may also be useful to restrict the constructs that can be added by the user so as to ensure consistency and reduce confusion and ensure that the final tool is intuitive and easy to use by the average user.

4. Integrated Information and Knowledge Management

We argue that human beings tend to understand and ultimately manage domains by devising abstract classifications and to associate items observed in the world with one or more classes [3] and make links between items of discourse. It would therefore appear that the most intuitive way to organise the information and associated knowledge in the first instance through a domain classification in which the different aspects of the domains handled by the individual are divided in a class hierarchy. This is shown in the top half of figure 1, with a sample ontology for a University Research School. The subject taxonomy provides a basic classification for the objects within the research domain (this could be more detailed if required) that is linked to the organisational taxonomy that describes the divisions and sub-divisions of the University (other departments such as finance, personnel etc. not shown). At the top the ontology shows how the University is organised in terms of its departments and divisions and subgroups. This includes for instance such activities as externally funded research projects and the personnel associated with them. At the bottom the diagram shows the subject taxonomy and proceeds from the concept of research subdivided progressively to its constituent parts in an iterative fashion.

Figure 2 shows a more detailed picture of the use of the subject hierarchy as an organising principle for a number of files or database entries that contain staff details, research interests and quality indicators. Also shown are the authors of these documents and some reflections/rationale/knowledge they have about some of the documents. It is expected that the user will have this view on their screen and will navigate and edit the graph through a point-and-click interface. The user will use the taxonomies to record new items of information (documents, rationales, structures, records, etc.) by linking them to one or more of the concepts in the ontology. For

the purposes of simplicity only a partial subject taxonomy is shown but classification into the other ontologies will also be required. The core ontology can then be used as a cataloguing system for information contained in documents stored elsewhere (such as word processing documents, spreadsheets, presentations, databases, etc.). The user can also add reference links to other items in the system which the new item is related to, such as to indicate another document containing background information. Alternatively, the user can record annotations containing knowledge about any item which can represent a critical appraisal, rationale for the item in question or any additional knowledge about the item and its use for a particular purpose. These will be in free text form recorded in separate documents whose reference will be inserted in the conceptual structure. In the diagram the USER1 has identified the book as a possible item for inclusion in annual research report and has linked the item to the Report document.

The benefit of this approach is that the user will be able to browse and explore the graph on-screen by progressing through nodes and arcs where facilities will be provided to control the desired detail shown (depth of neighbourhood, types of documents and links, only author's documents or including colleagues etc.). When looking for a specific document it is envisaged that the user will progress through one of the available ontologies to proceed to the subject that match the content of the document or through the projects they are associated with (or a match of a number of these).

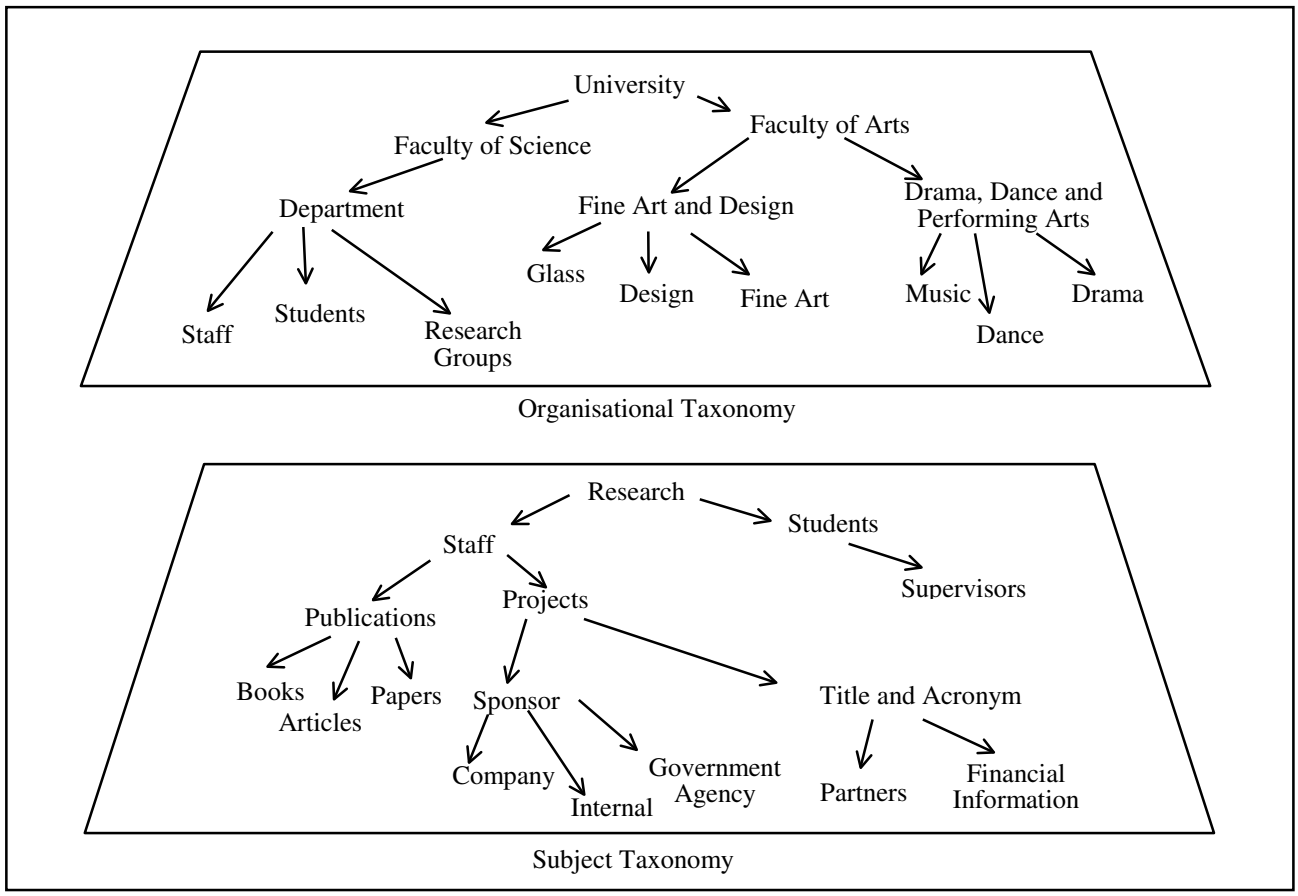

Figure 1: Alternative Ontologies

As argued earlier, document organisation approaches rely on a meaningful document hierarchy in the form of a filing hierarchy with meaningful directory and file names. Few users are methodical enough to develop meaningful and failsafe approaches for the structuring and naming that allow the user to effortlessly locate files as time progresses. An ontological approach, like the one presented here, has the advantage that the information can be accessed directly in a thematic fashion with the added benefit that other items with the same "theme" are

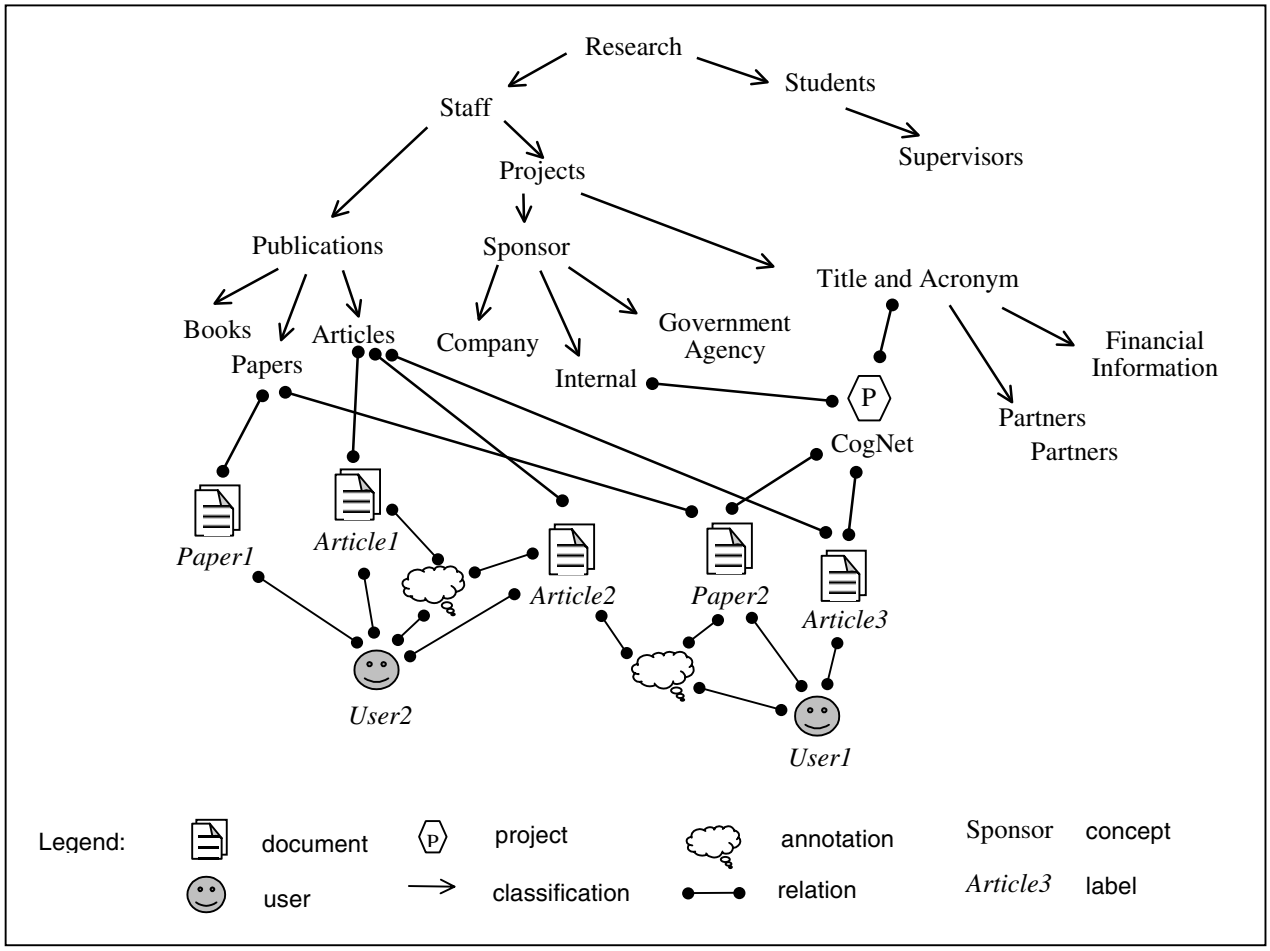

Figure 2: A Sample Use Case

immediately available as well by inspecting the immediate neighbourhood or by moving up to the subject taxonomy to view other items connected with it. This will allow context sensitive browsing and enable related information and knowledge to be located with ease.

One disadvantage of many ontology approaches and tools is that they tend to present ontologies in a tabular or frame-based form and which may be convenient for the technically versed user but which is difficult to comprehend and operate by a technically less literate user. In this form it is also difficult for the user to effectively browse the ontology so created and to easily comprehend the overall picture and where a given concept fits into the order of things. Consequently a more intuitive access mechanism, involving a graphical representation is needed. We believe that our approach remedies these concerns.

5. Benefits of the Proposed Approach

We have argued for an intuitive approach that is easy to pick up and use, capable of graphical representation and navigation rather than tabular and we believe that this concept can deliver this need. In our view it also satisfies the requirement of allowing items to be easily categorised through the provided taxonomies. These taxonomies will have to be built for the users in an organisation (and may be adapted and expanded as the need arises). The proposed system allows the managing of both information and knowledge in the same framework and to create relations between knowledge and the underlying information/data upon which it builds, as indicated in the diagrams above. This allows items of knowledge and indeed items of information to be seen in their context and which may be crucial for subsequent reinterpretation or use on a new task or project. The latter is particularly significant in a distributed environment where distributed

workers, maybe even in a virtual enterprise, have to create a shared information and knowledge space in order to collaborate more effectively since they are not able to rely on a shared understanding and culture that results from more intensive personal interaction in a common office.

The type of iconic representation shown in the diagram is in the process of being specified in detail and will be provided in a tool that will have editing and inspection capabilities. This will include iconic representations of items such as users, documents, projects, annotations etc. to represent items added to the shared collection by the users. This will be used in conjunction with existing applications and productivity tools as documents and files are being referred to rather than being incorporated into the information and knowledge management system. The documents and files will continue to be processed and edited by their respective application and remain in their local or remote location. The approach we propose operates at a less detailed and sophisticated level of detail as is the case with some examples of ontologies that have been developed for instance with Ontolingua such as the enterprise ontology [14] or the mineral ontology developed with APECKS [13]. At the same time the system will be easier to operate by the average user; speed and ease are essential if users are to be convinced to use the system. In this respect any significant effort required to maintain such structures are detrimental. Also this level of detail may not always be strictly necessary for general-purpose use. The benefit of using a formal system like conceptual graphs will provide rigor to allow the future use of reasoning techniques to be employed, such as a combination of conceptual mapping and case based reasoning as presented by [15]. The user should be provided with a basic repertoire of building blocks with which to quickly, and safely, incorporate new pieces of information and knowledge into the conceptual framework and the editing facilities should guide the user through this process.

The system described above has been developed as far as the core ontologies and data storage mechanisms. The data includes details on 600 academic staff, over 3000 publications and more than 1500 research projects. The approach detailed in [16] is being used to develop a mapping between ontology concepts and XML elements by way of DTDs. As they state this allows the creation of XML documents that can represent facts compatible with the domain model (ontology). Currently the user can exist at one of nine different levels of authority, each of which allows for different methods of access and for different viewpoints. In a sense this could be viewed as nine pre-packaged user types with fixed ontologies for each. The adoption of the new techniques described above will allow for reasoning over the data.

The user's experience can be assessed from the reaction of academic staff and from the views expressed by external viewers of the data. One of the levels of access is for users external to the University and we restrict them to a view mode only, they cannot for example annotate any document or record although in principle there is no reason why they could not and it might be valuable to have external comment on some of the research conducted in the University. The internal users are able to annotate and add records to various parts of the data store and to develop their own ontologies by building on the core provided. An assessment of this is being conducted which will guide the authors for future developments and also provide data on usability and utility. One difficulty found so far is that this type of system needs careful explanation and some training before users can really begin to explore the extent of its powers.

Irrespective of the question of ontology extension and refinement we believe that the proposed approach will provide a standard interface understood by all and which allows users to easily locate appropriate information and associated knowledge from colleagues and allow easy aggregation of relevant information and knowledge across the enterprise. The sharing modalities have not been addressed in detail in this paper, but the system is currently hosted on a central server with Internet access via standard web browsers. An ORACLE database contains all shared components that are accessible by all and can selectively be updated by the authors, who maintain editorial control. This will restrict possible side-effects in a purely distributed set-up when individual users are not logged on.

It may be argued that the level of detail of the information and knowledge management approach presented at a document level is too crude for many applications where more detail is needed. There exist approaches that allow the semantic analysis and categorisation of natural language texts. The problem is that the application of techniques to proceed to a full semantic analysis will quickly lead to the creation of huge conceptual structures which again will make the resulting system difficult for the user to use. It is however important to create an approach that is capable of eventually being extended to allow the analysis of documents and files at a more detailed level. The key is however that it must be controllable by the user of how far to go in this respect. One area of investigation that has considerable promise is the use of XML and has attracted considerable interest since its inception. XML if used for text documents in an organisation would provide an ideal ground for automated categorisation on one hand and easy specific referencing and annotation by the user. By separating the formatting from the semantic structuring of

documents the defined semantic structure can be used even to automatically generate document ontologies. The use of XML for the purpose of knowledge management is starting to attract attention and [17] is one example of this and presents an approach for an enterprise ontology guided approach for searching through XML documents. We believe that the use of XML however goes further than searching and that the ability to locate and annotate XML documents will add to this and which will require further investigation.

References

1 Tan TC, "The Development of an Intelligent Conceptual Storage and Retrieval System", PhD Thesis, University of Sunderland, UK, 1993

2 Bokma A., "A Source Modelling System and its Use for Uncertainty Management", PhD Thesis, University of Durham, UK, March 1993Bokma 93

3 Sellens C. & Wilson O., "The CMG Knowledge Intranet", Proc. of the 2nd Int. Conf. on Practical Aspects of Knowledge Management (PAKM98), Basel Switzerland, 29-30 Oct 1998 (U. Reimer eds.)

4 Sowa J F, "Conceptual Structures: Information Processing in Mind and Machine", Addison-Wesley, Reading, Massachusetts, USA, 1984

5 Sowa J F, "Knowledge Representation", Brooks/Cole Publishing, Pacific Grove, CA, USA, 2000

6 Guarino N, "Understanding, Building and Using Ontologies – a Commentary to 'Using Explicit Ontologies in KBS Development'" by van Heijst, Wielinga and Schreiber" Int. Jour. of Human and Computer Studies 46(3/4), pp 293-310

7 Farquhar A, Fikes R, Pratt W & Rice J, "Collaborative Ontology Construction for Information Integration", Research Report No. 63, Knowledge System Laboratory, Stanford University, Stanford USA, (ftp://kls.stanford.edu/pub/KSL-95-63.ps)

8 Dominique J, "Tadzebao and WebOnto: Discussing, Browsing and Editing Ontologies on the Web", Proc. of the 11th Knowledge Acquisition Workshop KAW'98, Banff, Alberta Canada, April 1998Dominique 98

9 Erikson H, Ferguson R, Shahar Y & Musen M, "Automatic Generation of Ontology Editors, Proc. of the Knowledge Acquisition Workshop KAW'99, Banff, Alberta, Canada http://ksi.cpsc.ucalgary.ca/KAW/KAW99/

10 Fernández M, Gómez-Pérez, Pazos J & Pazos A, "Building a Chemical Ontology Using Methontology and the Ontology Design Environment", IEEE Intelligent Systems, Jan/Feb 1999, pp. 37-46

11 Duineveld A, Stoter R, Weiden M, Kenepa B & Benjamins V, "Wondertools? A Comparative Study of Ontological Engineering Tools", Proceedings of the 12th Knowledge Acquisition Workshop (KAW'99), Banff, Alberta, Canada, 16-22 October 1999

12 Benjamins VR, "The ontological engineering Initiative $(KA)^2$, in Formal Ontology in Information Systems", pp. 287-301, IOS Press, Amsterdam

13 Tennison J and Shadbolt N, "APECKS: A Tool to Support Living Ontologies", Proceedings of the 11th Knowledge Acquisition Workshop KAW'98, Banff, Alberta Canada, April 1998

14 Uschold M (ed.), "Knowledge Level Modelling, Concepts and Terminology", Knowledge Engineering Review, vol 13/1 1998, pp. 5-29, Cambridge University Press, Cambridge, UK

15 Cañas A, Leake D and Wilson D, "Managing, Mapping and Manipulating Conceptual Knowledge", Proceedings of the AAAI-99 KM/CBR Workshop, 19 July 1999, Orlando, Florida, USA

16 Erdmann, M., Studer R., "Ontologies as Conceptual Models for XML Documents", AIFB, University of Karlsruhe, D-76128 Karlsruhe, Germany. http://www.aifb.uni-karlsruhe.de/WBS/publications/1999/kaw_merrst_1999.pdf

17 Rabarijoana A, Dieng R &Corby O, "Exploitation of XML for Corporate Knowledge Management", in "Knowledge Acquisition, Modeling and Management", Fensel & Studer (Eds.), LNCS 1621, Springer Verlag Berlin, 1999