Knowledge sharing agents over the World Wide Web

N J Davies, R S Stewart and R Weeks

Large and increasing amounts of information are now available both on the Internet and on corporate intranets. With the availability of these vast networked information resources comes a requirement for tools to manage the information and provide users with the information they want, when they want it.

This paper describes a system which facilitates and encourages the sharing of knowledge between groups of users within (or perhaps across) organisations. KSE (knowledge sharing environment) is a system of information agents for organising, summarising and sharing knowledge from a number of sources, including the World Wide Web, an organisation's internal intranet or from other users. Users are organised into closed user groups or communities of interest with related or overlapping interests. Such groups could be members of a project team, students studying the same subject (perhaps at different institutions), members of an organisational department, and so on. As well as sharing explicit (codified) knowledge, the sharing of tacit knowledge is encouraged via the automatic suggestion of, and support for, contact between people with mutual concerns or interests.

1. Introduction

There are now more than 60 million documents on World Wide Web (WWW) servers. Every day millions of people trawl the Internet for information using any one of a dozen or more different search tools. However, whether they find what they are looking for sometimes depends not only on their skill, but also on their luck. Studies indicate that many existing search engines do not meet the needs of users [1].

Recently, we have seen increasing interest in the use of Internet technology within organisations and the emergence of intranets [2]. More and more information is being stored on intranet networks and this has led to the realisation that these intranets are valuable repositories of corporate knowledge. Organisations are increasingly realising the importance of using knowledge and information residing on their networks for competitive advantage. It is forecast that annual spending on corporate intranets will top $2.1bn in the UK in 1998; for western Europe the total is expected to be over $7.7bn [3].

Key to the usefulness and viability of both the public Internet and corporate intranets, however, will be the ability to manage the information and provide users with the information they want, when they want it.

Of course, raw information in large quantities does not by itself solve business problems, produce value, or enhance competitiveness. Knowledge solves a problem. Knowledge is information transformed into a capability for effective action — that is to say, it produces competence leading to effective action. Thus today we see a great emphasis on the development of methods for turning information into knowledge. Information as a growing part of a company's asset base is often inefficiently captured, stored and sometimes simply lost. Knowledge management [4] is a strategy that turns an organization's information into greater productivity, new value, and increased competitiveness.

The first step in implementing a knowledge management solution is to provide seamless access to an organisation's entire collection of information. Many organisations have begun this process through the provision of an intranet. The logical progression for organisations that have implemented intelligent information access software will be to extend it to deliver additional value to knowledge workers. Organisations need to look beyond their formal organisational structures to recognise communities of interest — informal networks of people with common interests who can share knowledge across departments, as well as perhaps continents and time zones.

This paper describes a system which facilitates and encourages the sharing of information between communities of interest within (or perhaps across) organisations and which encourages people — who may not previously have known of each other's existence in a large

organisation — to make contact where there are mutual concerns or interests.

KSE (Knowledge Sharing Environment) is a system of information agents for organising, summarising and sharing knowledge from a number of sources, including WWW, an organisation's internal intranet or from other users. Users are organised into user groups or communities of interest with related or overlapping interests. Such groups could be members of a project team, students studying the same subject (perhaps at different institutions), members of an organisational department, and so on. KSE extends and enhances the earlier work on the Jasper information agent [5]. The Jasper agent automates the sharing of WWW-based information among a community of users with related interests. The additions and enhancements embodied in KSE, described below, are informed by the concerns of the knowledge management community and by the results of trials of the Jasper system.

2. Storing and organising knowledge in KSE

KSE agents are used to store, retrieve, summarise and inform other agents about information considered in some sense valuable by a KSE user. This information may be from a number of sources:

- it can be notes typed by the users themselves,
- it can be an intra/Internet page,
- it can be copied from another application on the user's computer.

Each KSE user has a personal agent which holds a user profile based on a set of key phrases which models that user's information needs and interests. As will be shown below, KSE agents modify a user's profile based on their usage of the system, seeking to refine the profile to better model the user's interests.

Given the vast amount of information available on the WWW, it is preferable to avoid the copying of information from its original location to a local server. Indeed, it could be argued that this approach is contrary to the whole ethos of the Web. Rather than copying information, therefore, KSE agents store only relevant meta-information. This meta-information is then used as an index to link to the actual information when a retrieval request is made.

When users find information of sufficient interest to be stored by KSE, a 'store' request is sent to KSE via a menu option on their WWW client. KSE then invites the user to supply an annotation to be stored with the information. Typically, this might be the reason the information was stored and can be very useful for other users in deciding which information retrieved from the KSE store to access. The user can also specify at this point one of a predefined set of interest groups to which to post the information being stored.

In the case of WWW-based information the uniform resource locator (URL) of the WWW page is then added to the KSE store. Similarly, when users wish to store some information from a source other than WWW, they can enter the information in a text box on their WWW browser and can again supply a relevant annotation. The information thus entered could be from a document in another format or might be a note or snippet of knowledge which the user wishes to enter directly themselves. This information is converted to a WWW hypertext mark-up language (HTML) page on the user's KSE server and stored as before.

Essentially, the KSE store is a simple term-document matrix M, wherein:

$$M_{i,j} = \begin{cases} 0 \text{ if term } j \text{ does not occur in document } i \\ n \text{ if term } j \text{ appears in document } i \ n \text{ times} \end{cases}$$

At storage time, KSE agents perform four tasks.

- An abridgement of the information is created, to be held on the user's local KSE server. This summary is created using the ProSum text summarisation tool. The summariser extracts key theme sentences from the document. It is based on the frequency of words and phrases within a document, using the technique of lexical cohesion analysis [6]. Access to this locally held summary enables a user to quickly assess the content of a page from a local store before deciding whether to retrieve remote information.
- The content of the page is analysed and matched against every user's profile in the community of interest. If the profile and document match strongly enough, KSE informs the user by e-mail of the page which has been stored.
- The information is also matched against the storer's own profile. If the profile does not match the information being stored, the agent will suggest phrases which the user may elect to add to their profile. Thus KSE agents have the capability to adaptively learn their user's interests by observing the user's behaviour.
- For each document, an entry in the KSE store is made, holding keywords, an abridgement of the document, document title, user annotation, URL, storer name and date of storage.

In this way, a shared and enhanced information resource is built up in the KSE store. Given that users must make a conscious decision to store information, the quality of the information in the KSE store is high — it is effectively pre-filtered by KSE users. Furthermore, each user leverages the assessment of the information made by all the other users.

3. Tacit and explicit knowledge in KSE

3.1 Sharing and retrieving explicit knowledge in KSE

It has been shown in the section above how KSE allows a user to store information of interest using an enhanced, shared bookmark concept. This facility goes well beyond the bookmarks familiar from WWW browsers such as Netscape Communicator, in that in addition to the reference to the remote WWW document, a summary of the document, an annotation, date of storage and the user who stored the information are recorded in a shared store. Furthermore, KSE can be used to store and organize information from many sources and in many formats (rather than only WWW-based information).

This section discusses the various ways in which KSE facilitates access to, and the automatic sharing of, the knowledge thus stored.

E-mail notification

As described above, when information is stored by a KSE agent, the agent checks the profiles of other agents' users in its 'local community' (the set of users who contribute to that particular KSE store). If the information matches a user's profile with a score above a certain threshold, an e-mail message is automatically generated by the agent and sent to the user concerned, informing them of the discovery of the information. Thus in cases where a user's profile indicates that they would have a strong interest in the information stored, they are immediately and automatically e-mailed about the appearance of the information.

Keyword retrieval — accessing information and people

From their KSE home page, a user can supply a query in the form of a set of key words and phrases in the way familiar from WWW search engines (see Fig 1). The KSE agent then retrieves the most closely matching pages held in the KSE store, using a vector space matching and scoring algorithm [7] as follows:

$$\mathrm{sim}(q, d) = \sum_{i=1,n}(t_{iq} * t_{id}) \Big/ \sqrt{\left[\sum_{i=1,n} t_{iq}^2 * \sum_{i=1,n} t_{id}^2\right]}$$

where t_{iq} is the weight of the ith term in the query q, and
t_{id} is the weight of the ith term in the document d, and
n is the number of unique terms in the combined profiles of user u_1 and user u_2.

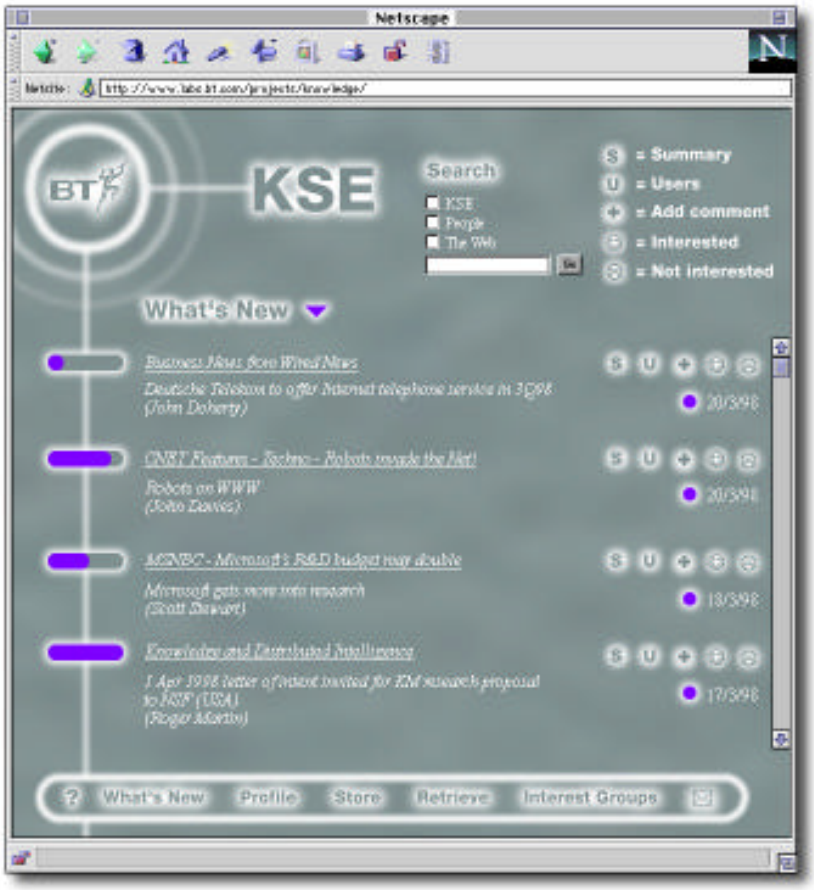

Fig 1 A typical KSE home page.

A simple Boolean weighting schemeis currently used, where the weught of a term is 1 if it occurs in the query or document under consideration, and 0 otherwise.

In addition to these pages from the KSE store, the agent can also retrieve a set of pages from an organisation's intranet and from the WWW. The agent then dynamically constructs an HTML page with a ranked list of links to the pages retrieved and their abridgements, along with the scores of each retrieved page. In the case of pages from the KSE store, any annotation made by the original user is also shown. It should be noted that the user can ask their agent to search for other users by selecting the appropriate check box (see Fig 1). More about this capability to identify other users as well as information is contained in section 3.3 which looks at accessing explicit and tacit knowledge using KSE.

What's new?

A user can ask his KSE agent 'What's new?' The agent then interrogates the KSE store and retrieves the most recently stored information. It determines which of these pages best match the user's profile. A WWW page is then presented to the user showing a list of links to the recently stored information, along with annotations where provided, date of storage, the storer and an indication of how well the information matches the user's profile.

In addition, a series of buttons are provided so that the user can:

- add their own comment or annotation to information stored by another user,
- indicate interest or disinterest in a particular piece of information — this feedback will be used to modify the user's profile,
- examine a locally held summary of the information before deciding to download all the information,
- ask their KSE agent to identify other users with an interest in the information under consideration (see section 3.2).

This 'What's new?' information is in fact displayed on the user's KSE home page, so that whenever they access the system, they are shown the latest information. For a typical KSE home page, see Fig 1.

Interest groups

As mentioned above, when storing information in KSE, a user has an opportunity to specify one of a predefined set of interest groups to which to post the information. Interest groups gather together pages of related information. KSE users can visit interest group pages which are dynamically constructed from the pages which have been posted to them and consist of a list of links to the pages and their abridgements, along with any annotation provided by the original storer of the page. Interest groups are similar to the 'list of links' pages found in many WWW locations, with the important extensions that multiple users can contribute to the list (automatically via the storage process) and that abridgements of the information and annotations by the original storer of the link are also available.

3.2 Adaptive agents

It has already been mentioned that KSE agents adapt to better understand their user's interests over time. There are two types of event which trigger the profile adaptation process.

As discussed above, when a user is storing some information, if the profile does not match the information being stored, KSE will extract the main themes from the information using a unique theme extraction algorithm [8]. The user's agent then suggests to the user new phrases they may wish to add to their profile. The user can accept or decline these suggestions.

Similarly, when information stored by another member of the community is retrieved by a user using one of the methods described in section 3.1, a feedback mechanism is provided whereby the user can indicate interest or disinterest in the information by clicking on a button (indicated by ☺ or ☹ as shown in Fig 1). Again, the agent will suggest to the user phrases which should be added to or removed from the profile.

3.3 Finding people and tacit knowledge in KSE

Section 3.1 focused on the technical aspects of KSE and on explicit knowledge. This section now looks at the social aspects of the system and tacit knowledge.

A large amount of the knowledge within an organisation may of course not be codified — it may be personal, context-specific and difficult to write down. Such knowledge is referred to as tacit knowledge [9]. When tacit knowledge is difficult to make explicit (codify), we need to find new ways of transmitting the knowledge through the organization. Failure to do so can lead to loss of expertise when people leave, failure to benefit from the experience of others, needless duplication of a learning process, and so on.

One way in which a system such as KSE can encourage the sharing of tacit knowledge is by using its knowledge of the users within a community of interest to put people who would benefit from sharing their (tacit) knowledge in touch with one another automatically.

User profiles can be employed by the KSE system to enable people to find other users with similar interests. The user can request KSE via their WWW client to show them a

list of people with similar interests to themselves. KSE then compares their profile with that of every user in the store and returns to the WWW client for viewing a list of names of users whose interests closely match their own. Each name is represented as a hypertext link which when clicked initiates an e-mail message to the named user. Profiles in KSE are a set of phrases and the vector space model can be used to measure the similarity between two users u_1 and u_2:

$$\text{sim}(u_1, u_2) = \Sigma_{i=1,n}(t_{i1} * t_{i2}) \Big/ \sqrt{\left[\Sigma_{i=1,n} t_{i1}^2 * \Sigma_{i=1,n} t_{i2}^2\right]}$$

where t_{i1} is the weight of the ith term in the profile for user u_1, and
n is the number of unique terms in the combined profiles of user u_1 and user u_2.

As before (see section 3.1), a Boolean weighting scheme is used.

A threshold can then be used to determine which users are of sufficient similarity to be deemed to 'match'.

This notion is also extended to allow a user to view a set of users who are interested in a given document. When KSE presents a document to the user via their WWW client using the 'What's new?' facility (see above), there is also a hyperlink presented which when clicked will initiate a process in the KSE system to match users against the document in question, again using the vector cosine model. KSE determines which members of the community match the relevant document above a predetermined threshold figure and presents back to the user via their WWW client a list of user names. As before, these names are presented as hypertext links, allowing the user to initiate an e-mail message to any or all of the users who match the document.

In addition, as already discussed in section 3.1, a user can carry out a keyword search on other users and thus identify users with an interest in a particular subject.

In this way, KSE, while not claiming to actually capture tacit knowledge, provides an environment which actively encourages the sharing of tacit knowledge, perhaps by people who previously would not otherwise have been aware of each other's existence.

4. Conclusions

This paper has described KSE, a tool for encouraging the exchange of tacit and explicit knowledge among a community of interest. KSE builds on earlier work with the Jasper system [5] and incorporates a series of extensions including:

- the ability to store information from multiple sources (rather than simply WWW pages),
- the facility for multiple users to annotate information,
- the ability to search for other users (as well as information) in a variety of ways,
- the facility for users to provide feedback on retrieved information in order to modify their profile.

Figure 2 shows the main features of the KSE system of agents, people and information — users can create, store and organise information, they can locate relevant information stored by others in their community of interest, and they can contact other users with overlapping interests.

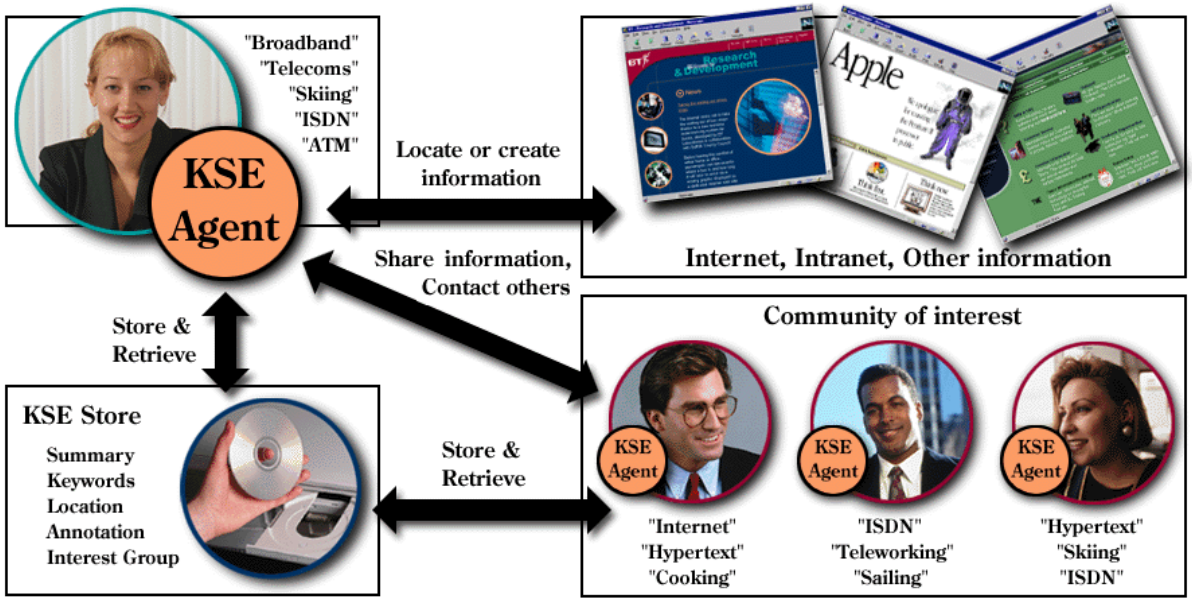

Fig 2 The KSE.

In his seminal article, Bush [10] describes a tool to aid the human mind in dealing with information. He states that previous scientific advances have helped humans in their interactions with the physical world but have not assisted humans in dealing with large amounts of knowledge and information. Bush proposed a tool called a 'memex' which could augment human memory through associative memory, where related pieces of information are linked. Trails through these links could then be stored and shared by others. WWW itself fulfils Bush's vision in some respects — Bush's associative memory can be seen in the hyperlinks of WWW. What is lacking is a way of organising this vast 'memory' of WWW pages into coherent 'trails' which can be saved and communicated to others. Currently, only relatively simplistic bookmarks and menus are available.

KSE goes some way to addressing these problems by providing agents which, as has been shown, can store meta-information about WWW pages that can then be used to retrieve relevant pages quickly and easily and share the information contained in those pages with other users with the same interests. In addition, KSE can put users in touch with other users who share their interests in general or who share their interest in a specific document. In this respect, KSE is moving into the area of tacit knowledge in the sense that it is suggesting useful person-to-person interactions wherein tacit knowledge can be exchanged. So as well as addressing the issue of how best to search WWW for information, KSE is an attempt to address the complementary problem of how best to store information once it has been found and how to share information with others with the same interests. KSE is a step along the road towards the original vision for WWW [11] as a network which supports fully co-operative working and the sharing of knowledge.

Acknowledgement

Thanks are due to Ian Matthews for visual design consultancy on the KSE project.

References

1 Pollock A and Hockley A: 'What's wrong with Internet searching?', Designing the Web: Empirical Studies conference, Microsoft Usability Group, Microsoft, Redmond, US (October 1996).

2 Cochrane R: 'Unleashing the intranet', BT Technol J, 15, No 2, pp 107—113 (April 1997).

3 INTECO Corporation (1997), at http://www.intecoresearch.com/

4 Nonaka I and Takeuchi H: 'The Knowledge-creating Company', Oxford University Press (1995).

5 Davies N J et al: 'Information agents for the World Wide Web', BT Technol J, 14, No 4, pp 105—114 (October 1996).

6 Davies N J and Weeks R: 'ProSum — profile-based text summarization', First Automatic Text Summarization Conference (SUMMAC-1), Virginia, USA (May 1998).

7 Salton G: 'Automatic Text Processing', Addision-Wesley, Reading, Mass, USA (1989).

8 Davies N J, Weeks R and Crabtree I B: 'Document theme extraction', to be published (1998).

9 Polyani M: 'The Tacit Dimension', Routledge & Paul, London (1966).

10 Bush V: 'As we may think', The Atlantic Monthly (July 1945), also available as http://www.isg.sfu.ca/~duchier/misc/vbush/

11 Berners-Lee T: http://www.w3.org/Summary.html

John Davies graduated in Computer Science and Physics at the University of London in 1981. Following two years in industry, he obtained an MSc in Computer Science from Essex University. After four years with GEC, he returned to Essex where he received his PhD in the area of artificial intelligence. He joined BT in 1990 and currently leads BT's knowledge management research group. Technical interests include on-line information management, agent technology, intranets and their use in knowledge management applications. He is an author of and contributor to many books and papers in the areas of information retrieval, agents, AI and knowledge management. He is a Chartered Engineer and a member of the British Computer Society, where he sits on the Information Retrieval Specialist committee.

Scott Stewart attended Aston University as a BT sponsored student where he spent 4 years studying for a Master of Engineering in Electronic Systems Engineering. In 1996 he graduated with first class honours, and took up a position in the Applied Research and Technology department at BT Laboratories. Currently his main technical interests lie in Internet agents and information management. He is a member of the Knowledge Management Research group where he primarily fulfils a development role but also promotes the work of the group through presentations to both internal and external customers and has co-authored papers on user profiling and knowledge management.

Richard Weeks joined BT as a student in 1967. He graduated in Electronics Science at Southampton University in 1971, joining BT Laboratories. After a period in the New Materials division he moved into Visual Telecommunications, establishing computer-based techniques for the design, synthesis and analysis of high-quality passive networks for video applications. Subsequent areas of software experience include VLSI CAD, telephony, and speech-dialogue prototyping. He is now involved in knowledge management research, with an emphasis on text summarisation, search-engine technology, and other WWW-related tools.