INFORMATION RETRIEVAL ALGORITHMS FOR KNOWLEDGE MANAGEMENT – THE CHALLENGE CONTINUES

Elaine Ferneley

Information Systems research institute, University of Salford, Manchester, M5 4WT, UK
E.Ferneley@salford.ac.uk

Brendan Berney

Information Systems research institute, University of Salford, Manchester, M5 4WT, UK
B.T.Berney@salford.ac.uk

Yacine Rezgui

Information Systems research institute, University of Salford, Manchester, M5 4WT, UK
Y.Rezgui@salford.ac.uk

This paper considers Information Communication Technology (ICT) support for the knowledge creation process that takes place by the interaction of both tacit and explicit knowledge with the knowledge creating entities of the individual, group and organisation (or organisations). Attempts to provide ICT support for this process have tended to focus on two stages in the knowledge evolution cycle, firstly extraction and representation and secondly dissemination. In order to extract and represent knowledge a number of approaches have been used, these include: the use of knowledge bases and ontologies, the use of filtering and categorisation mechanisms to extract key terms and the development of various weighting mechanisms in an attempt to prioritise or cluster related entities. To support dissemination various approaches to user profiling have been used which usually incorporate some form of adaptive information filtering mechanism. This paper presents a critical evaluation of a number of the more well know extraction and representation techniques. It then presents a set of user profiling techniques appropriate for use in intra-organisation knowledge management portal applications.

1. INTRODUCTION

The need for organizations to encourage collaborative working through knowledge sharing in order to better exploit their intellectual capital is now recognized [Davenport & Prusak, 1998], [Sieloff, 1999]. Much of the work to date suggests that despite the intuitive appeal of a collaborative approach significant knowledge remains locked away. It is argued that the problem is both technological and cultural. Whilst technologically mature, sophisticated information communication technologies (ICTs) exist, providing a technological medium to support a collaborative culture in which knowledge can be elicited, stored, shared and disseminated is still elusive [Scarborough et al, 1999], [Swan et al, 2000]. Early research into supporting informal collaborative information exchanges through technology were optimistic, inspired by the rationalistic view of computerisation [Sproull & Kiesler, 1986], [Kling, 1980]. Later research recognised that the technological solutions were context dependent, operating and adapting in specific social contexts [Kling, 1996], [Orlikowski et al, 1995]. More recently [Mantovani, 2001] has introduced the co-construction metaphor, proposing that a socio-technical relationship emerges in each scenario as it employs social and technical dimensions.

This paper reports on the technology review work that has been undertaken as part of the C-Sand project and introduces the project's approach to the development of a socio-technical solution. The technical focus of the C-Sand project is to develop software environments that support the creation and dissemination of knowledge related to the development of sustainable constructs. In this rapidly evolving domain there is a need to disseminate best practice both within and between companies, the C-Sand project aims to provide technological architectures and mechanisms to do this.

Organisations are become more flexible and willing to share across organisational boundaries. There is an increasing trend towards flatter organisational structures, the encouragement of individuality, the decentralisation of control and the development of heterogeneous inter-organisational groups brought together to solve specific tasks – the construction industry, which is the focus of this project, can be viewed as an early-adopter of such alternative structures. ICTs are increasingly being used in order to facilitate communication and co-ordination as communities become dispersed across time and space. In practical terms accessing and distributing knowledge via ICTs is difficult, ICTs frequently become inhibitors rather than enablers of knowledge transfer [Swan et al, 1999], [Sorensen & Snis, 2001]. The C-Sand project aims alleviate some of these difficulties by adopting an approach which moves away from the notion of technology managing information and toward the idea of social processes and technological tools evolving reciprocally – the notion of co-construction [Mantovani, 2001].

This paper identifies two levels of support that ICTs could give to a knowledge management initiative. Firstly, from a high level view, in order to support innovation and facilitate knowledge dissemination and leveraging KM ICTs should support innovative interaction between disparate communities. By encouraging social networking multiple views on common knowledge can be explored, creating a greater depth of understanding and encouraging individuals to take ownership of the entire project rather than their specific tasks. This paper argues that, whilst a high level view can be supported, this should not be interpreted as the ability to codify tacit knowledge, tacit knowledge tends to be situated and may be too difficult to explain, uncertain, irrelevant to others, subject to change, politically sensitive or too valuable to its owners to disseminate [Gardner, 1998]. Therefore, rather than attempt to evolve tacit knowledge into explicit form by codification it is argued in this paper that the more useful ICT solution is to create environments which, within the metaphor of co-construction, support a non-confrontational collaborative environment that both supports explicit knowledge sharing and encourages the socialization of tacit knowledge, thereby stimulating the SECI (Socialization, Externalisation, Combination, Internalisation) cycle and providing a *ba* environment [Nonaka and Treece, 2001]. Secondly, from a lower level view, if ICTs are to be used to support the management of explicit knowledge then formalized structures and representations are inevitable – the classification, retrieval and dissemination of information via an ICT, at the lowest level, relies on some form of pattern matching and, with the aid of more recent innovations, extrapolation exercises. Specific problems in the articulation and dissemination of explicit knowledge relate to how such knowledge is stored and the formats in which it is stored. Classic extraction mechanisms focus on keywords, term frequency and various linguistic analysis techniques. From a technology perspective, more advanced technologies incorporate various techniques including meta-data representations, taxonomies or ontologies and probabilistic models [Baeza-Yates, 1999]. Such technologies largely rely on the user actively engaging in a search activity, there is a tendency to assume that the user can formulate their information needs into an appropriate query in a non-collaborative environment with limited query reformulation activity. In such isolation the user has little motivation to disseminate experiences and knowledge thus adding to the domain's knowledge corpus. More recently research has focussed on proactively supporting the user in their query formulation and information retrieval activities and facilitating the social interaction processes that have resulted in serendipitous information retrieval. This paper recognises that, if ICTs are to be used to support the lower level KM processes, algorithmic techniques are a prerequisite. Therefore the paper considers some of more mature algorithmic approaches to information and knowledge dissemination within the context of supporting the social processes that are imperative if externalisation of explicit, let alone tacit, knowledge is to occur.

The remainder of this paper is structured as follows: section 2 will review the more widely known strategies that are employed in the search for knowledge and information. Section 3 provides a critical evaluation of a number of indexing, abstraction and retrieval algorithms. Section 4 introduces some algorithmic approaches to support knowledge elicitation and dissemination that the C-Sand project intends to trial. Section 5 presents conclusions and future directions.

2. KNOWLEDGE AND INFORMATION SEEKING STRATEGIES

In order to design KM systems it is important to identify the differing types of knowledge seeking strategies that are employed. Researchers have investigated information and knowledge seeking strategies from a many perspectives, identifying differing levels, tactics and approaches [Bates, 1979a, b]; [Shute & Smith, 1993]. This paper argues that it is impossible to develop ICT support for IR, let alone KM, that supports all strategies across differing domains. Rather, the relationship between a users interactive goals and their information seeking strategies should be identified and, from within this context, a tailored ICT environment can evolve.

2.1. Theoretical framework

KM processes should aim to support both high and low level user goals. [Xie, 2002] presents a hierarchical view of co-dependent user goals. Firstly, long-term goals which relate to an individual's personal aims; secondly, learning search goals which are a user's current task related goals; thirdly, current search goals which are a user's specific search requirements and finally, interactive intention goals which refers to the set of sub-goals that a user has to perform in order to achieve their current search goal. In order to achieve these goals a range of constantly evolving information seeking strategies are employed, as information is retrieved so the user's goals may change, this evolution should be supported by any ICT implementation.

To satisfy their various goals users engaged in the search process have recognized the importance of user collaboration, typical actions are to confer with peers or engage in a reference interview with a subject librarian [Marchionini, 1995]. Such collaborative activity allows users to address problems that go beyond their individual capacity to solve them. Traditional ICT supported IR activities work under the assumption that the user can recognize their information needs and formulate them into an appropriate query. In reality users may only be able to provide vague search terms. Indeed the term *browser* as applied to internet search engines is an equally appropriate title for users engaged in knowledge acquisition from electronic sources, the process is frequently opportunistic, reactive and unplanned, the aim of which may be uncertain and evolve during the browsing process. It is argued that it is such informal social interaction that will be lost by the move to remote, computerized searching [Bishop & Star, 1996]. Twidale et al [1997], building on the work of Bates [1979a, b], have proposed a classification of social interaction approaches that are, usually collaborative, informal communications that promote serendipity: *consultation* - to ask a specific colleague for help; *wandering* - being receptive to new sources of information and new search ideas triggered by the materials that come into view; *brainstorming* - although this can be done in isolation, it is an activity, which is usually more successful when undertaken as a group and *bibbling* - to utilize previous searches, as in consulting a bibliography. By its very nature the Internet encourages browsing but isolates the user creating a barrier to social interaction. This paper recognizes the importance of social interaction during the browsing process when using computerized information retrieval tools. Specifically, the importance of *serendipitous discovery* is acknowledged, the identification of useful, although perhaps not directly relevant information, during even highly focussed information searches [Twidale et al, 1997].

2.2. Supported Strategies

The socio-technical implementation under development within the C-Sand project regards support for knowledge elicitation and dissemination at both the individual and the community as interrelated. From the individual's perspective an ICT should provide a bespoke view that evolves with the user over time. Such an environment should provide support for a number of differing monitoring and searching tactics, from passive knowledge push to active knowledge discovery, be that via serendipity or highly targeted search strategies. Such a range of strategies can benefit from support from the wider community. To this end we propose an environment that supports IR via keyword search which:

- trials innovative abstraction and retrieval algorithms with the aim being to support *knowledge* as well as information retrieval;
- allows users to incrementally develop *spheres of interest*;
- proactively augments query terms to simulate individual *brainstorming* activity;
- supports the notion of *communities of practice* by introducing the user to peers with similar interests [Wenger, 1998];
- provides mechanisms to facilitate knowledge *push*.

3. INDEXING, ABSTRACTING AND RETRIEVING

Indexing and abstracting are the usual mechanisms to create characterisations of text, the result is a representation of the text [Rowley, 1988], [Moens, 2000]. Indexing extracts a set of words and phrases, which are commonly used as identifiers for the underlying text. Abstracting creates a coherent summary of the text's content. In both cases human intervention and interpretation is still the norm, although increasingly the process is being automated. Retrieval of information usually follows the model of returning a ranked set of documents in response to a user query. Based on Luhn's seminal work, this is usually undertaken by comparing a representation of the query against representations of documents within the search set [Luhn, 1957].

3.1. Indexing and Abstraction Mechanisms

Index terms can either be drawn directly from the text, this is known as *extraction indexing* (often supported by some weighting technique to provide some indication of *importance* [Salton, 1975]). Extraction indexing can be undertaken from different perspectives, providing a multi-facetted view of the text however this increases the possibility of homonyms. Furthermore, it may be difficult to reuse the extraction set in other searches – the underlying concepts may not have been captured. Alternatively the index can be drawn from a controlled vocabulary - *assignment indexing* which can incorporate semantic structures that identify hierarchical and associative relationships, assigned terms are referred to as *descriptors*. The key advantages of the assignment approach are the elimination of ambiguity, the associated increase in precision and the incorporation of associative relationships that significantly increase the richness of the indexing approach. There is a symbiotic relationship between indexing and abstraction mechanisms, indeed an abstract can be viewed as a narrative index.

3.2. Retrieval Mechanisms

The usual approach to information retrieval involves a relatively static document set being interrogated by a changeable set of queries, as an extension to this approach filtering mechanisms can be employed. Filtering is usually based on descriptions of an individual or group of users preferences – commonly known as *user profiling*. The most common retrieval models are the Boolean model, the vector space model, the probabilistic model and the network model [Moens, 2000].

Using a purely Boolean approach a document can be either relevant or not to a given query. The main criticism of the Boolean model lies in its binary evaluation system. Variants on the Boolean model include the ability to rank the retrieved document set based on partial matches to the query conditions. To increase precision, recent extensions to the pure Boolean approach include index term weighting. The vector space model represents both documents and queries as vectors in a vector space; the relevance of a document to a query is calculated by the application of a distance measure [Salton, 1989]. The measure, known as Term Frequency * Inverse Document Frequency (TF/IDF) has the effect of giving keywords that occur frequently in a particular document (and that are peculiar to that document) a high weighting whilst lowering the weights of universally occurring keywords such as 'the' or 'and'. As queries and documents may both be prepared and represented in this way, it is possible to measure the similarity between queries and documents trigonometrically; the query vector is compared with each document vector in the collection using the formula:

$$Similarity(Q_{i,}D_{j}) = \frac{\sum_{k=1}^{t}(w_{jk}w_{ik})}{\sqrt{\sum_{k=1}^{t}(w_{ik}^{2}) X \sum_{k=1}^{t}(w_{jk}^{2})}}$$

Where Q_i is a query vector comprising a set of weights w_{ik} and D_j is a document vector comprising a set of weights w_{jk}. The vector space model is very popular due to its simplicity and suitability for application across different domains and text types. It has been criticised for not recognising any correlation between terms and assuming that term vectors are pair-wise orthogonal. The probabilistic model approaches IR by estimating the probability of a document representation matching or satisfying a query representation, documents can then be ranked according to these probabilities. User feedback is used recursively to evaluate the relevance of the retained set. Documents are ranked in decreasing order of their probability of being relevant however the method does not take into account the frequency in which an index term appears within a document. The network retrieval model represents documents and queries as networks. Typically, each document has text nodes that correspond to a specific text representation - identifiers, concepts or index terms. Multiple text nodes can be generated from differing perspectives or by applying differing indexing techniques. The relationships between nodes can be defined by applying probabilistic or weighting approaches. A similar representation can be generated for a query. The two networks can then be connected by their common concepts to form an inference network, the most common inference approach used for retrieval being Bayesian.

4. ALGORITHMS FOR KNOWLEDGE MANAGEMENT

The C-Sand system employs a networked approach to the retrieval of documents, whereby documents are related to keywords by a double network of weighted links. These weights allow the significance and position of both document and query keywords to be used in retrieval. This approach provides both highly accurate ranked retrieval, a suitable platform for a novel document summarisation and indexing technique and intrinsic support for interactive user level components of the system, such as query by reformulation and user profiling.

4.1 Indexing and Abstraction Mechanisms

The C-Sand approach to summarisation is based upon the TF/IDF method [Salton, 1989]. The summarising method described below is not intended to provide a comprehensive abstract of the document, but rather an indication of its main salient content. Whilst there are methods of document summarising technology that are able to provide more effective summaries or abstracts of text documents, these tend to involve linguistic processing which makes them domain/language dependent. The approach provides a phrase signature comprising an ordered set of weighted keywords representing the *average* of the phrases contained within the document. It is believed that this method provides for each document, an indication of the major scope or *essence* of its contents. The method has the following stages:

a) Segmentation of the document into sentences. The document is broken down into sentences using punctuation and layout as a guide. This produces a set of abstract phrases.
b) Conversion of each phrase into a *phrase neuron* which represents the keywords within that phrase (minus closed-class keywords such as 'and' and 'the'). Term weights are allocated based upon their frequency within the phrase, capitalisation and the overall status of the phrase within the document (titles, headings etc.). Position weights are allocated according to the order of the words within the phrase. Where a term occurs more than once in a phrase, the position weight is the average of its absolute positions. In line with standard neural network practices, and to prevent long sentences from gaining a weight advantage over shorter phrases, both frequency and position weights are scaled to between 0 and 1.
c) All phrases extracted from the document are clustered into sets of similar phrases. Each phrase is used to trigger every other phrase within the document. Thus each phrase will produce a variably sized set of *similar* phrases. The largest of these sets is taken to be an indication of the *average content* of the document. Finally, these phrases are sorted into their original order within the document.
d) The final task in indexing the document is the production of the signature itself. This involves producing a set of weighted keywords representing the aggregate of the phrases in the summary set.

Variables that may be used to affect the above described method include varying the trigger threshold of the phrase neurons to produce differently sized summary phrase sets, influencing the phrases contained in the phrase sets by centring the clustering around a *centre phrase*.

4.2 Collaborative User Profiling

The C-Sand system aims to support multiple users, unite users with similar *interests* and, by presenting the differences between these similar *interests*, to demonstrate to them subtly different IR techniques, as well as providing the results of previous searches. The *interest sharing* process is carried out in two ways. Firstly, pre-search collaboration is used. During query formulation, the system attempts to retrieve a user's interests based on the keyword entry, if the user currently has no relevant interests, the system attempts to trigger spheres of interest in other users' profiles. Furthermore, the interests returned are compared with the assistant's existing interests and may be retained for future use if they are deemed similar enough. This approach allows the system to *bootstrap* itself in order to start providing a service more quickly. The second way in which the *interest sharing* process is carried out is via post-search collaboration. Whilst pre-search collaboration provides *emergency help* for a user, post-search collaboration provides a mechanism for a more generalised learning enhancement. Under this approach, whenever the system is idle, it will attempt to augment each user's profile with interest nodes from other users' profiles by using each interest node in a user's profile to trigger similar interests in other users' profiles. This form of collaboration is intended to provide the opportunity to unite similar users, present ideas for different searches and to determine whether the search proposed by a user has already been carried out by another user (by offering the results of previous searches).

Additionally, when the system is not in use, a user's set of interests is used in order to perform a proactive search using simple genetic algorithms. A cross-section of the interest set is taken by extracting the highest weighted keywords from the set as this reflects the subjects in which the user is most interested. The system may then carry out a search using these keywords and presenting the resulting documents for review when the user next logs in. Various constraints are proposed in order to avoid repeated recommendation of the same documents. Successive proactive searches may be made to sample keywords from different subsets of the interest spheres, by either cycling through them or by random selection.

4.3 Retrieval Mechanisms

The C-Sand project proposes that accuracy of document retrieval may be improved greatly by extending the retrieval system to consider the relative positions of the keywords. Rather than providing a relevance measurement based solely upon a *bag of words* (i.e. unordered keywords), the proposal is to measure the relevance of documents on the basis of similarities between phrases representing queries and phrases representing documents. Consider the following single-phrase documents:

- *US government pursues Microsoft under their anti-trust laws.*
- *Microsoft pursues the US government under their anti-trust laws.*

The query 'Who pursues Microsoft?' will produce the same relevance ranking for both documents using a *bag of words* system. The approach multiplies the relevance weighting of each word of the query with the relevance weighting of each word within the document e.g. assume that the query keywords 'pursues' and 'Microsoft' have relevance weightings of 0.7 and the query keyword 'who' has a relevance weighting of 0.1. The standard approach fails to identify which of the documents is most relevant to the query, because the relative positions of the words within the phrases are not taken into account – see table 1:

Query 1		Who	Pursues	Microsoft				
Weight		0.1	0.7	0.7				
d1	US	Government	pursues	Microsoft	Under	Anti-Trust	Laws	Score
Weight	0.7	0.7	0.7 (x0.7)	0.7 (x0.7)	0.7	0.7	0.7	0.98
								0.98
d2	Microsoft	pursues	US	Government	Under	Anti-Trust	Laws	
Weight	0.7 (x0.7)	0.7 (x0.7)	0.7	0.7	0.7	0.7	0.7	0.98
								0.98

Table 1

Alternatively, we propose a scaling method where positional weightings are given to keywords between –1.0 and 1.0, using the following formula:

$$w_i = w_i * \frac{1.0}{|w_{\max}|}$$

Where w_i is the weighting, which may be negative, given to the ith keyword of the phrase, and w_{max} is the number of keywords in the phrase. The relevance weightings given to keywords are scaled in the same way. Generally, known vector-space analysis methods and document similarity measurement methods, normalise the weights of keywords by using a formula which produces vectors in which the sum of the keyword weights = 1.0. However, this formula affects individual weights depending upon the number of keywords within the keyword vector. If a document or interest node contains many keywords, the individual weights of keywords are reduced unnecessarily. Thus, if a small query were used to retrieve documents with keyword vectors of varying lengths, those with few keywords would be retrieved with higher relevance scores than those with large numbers of keywords, thus penalising larger documents. This normalisation method is therefore not used, and the system instead uses the scaling method described above. The method presented in Table 2 may be expressed as follows: Given a query $Q_j=(w_j1,p_j1,w_j2p_j2,...,w_jt,p_jt)$ and document $D_i=(w_i1,p_i1,w_i2p_i2,...,w_it,p_it)$, where w_j (and p_j) and w_i (and p_i) are the weights (and positions) of the query and document keywords respectively the similarity is given by:

$$Similarity(Q_i, D_i) = \sum_{k=1}^{t} w_{jk} w_{ik} + \sum_{k=1}^{t} (p_{jk} p_{ik})$$

Query 1		'Who'	'pursues'	'Microsoft'				Score
Weight		0.1	0.7	0.7				
Pos		0.1	0.5	1.0				
Doc1	US	Government	pursues	Microsoft	Under	Anti-Trust	Laws	
Weight	0.7	0.7	0.7 (x0.7)	0.7 (x0.7)	0.7	0.7	0.7	0.98
Pos	0.14	0.28	0.42 (x0.5)	0.56 (x1.0)	0.7	0.84	1.0	0.77
								1.75
Doc2	Microsoft	pursues	US	Government	Under	Anti-Trust	Laws	
Weight	0.7 (x0.7)	0.7 (x0.7)	0.7	0.7	0.7	0.7	0.7	0.98
Pos	0.14 (x1.0)	0.28 (x0.5)	0.42	0.56	0.7	0.84	1.0	0.28
								1.26

Table 2

5. CONCLUSIONS

This paper has introduced the early work being undertaken as part of the C-Sand project into developing ICTs to support KM activity from a socio-technical perspective. There is a paradoxical relationship between the notion of developing KM initiatives and employing technological support for such initiatives. By its very nature tacit knowledge ceases to be tacit when embodied within an ICT environment. The C-Sand project aims to employ the notion of co-construction; social processes and technological tools evolving reciprocally [Mantovani, 2001]. Furthermore, we acknowledge that ICT support for KM, at a low level, requires some form of pattern matching and extrapolation algorithms. Therefore this paper has critically evaluated several of the more mature indexing, summarisation and retrieval approaches and, based on this analysis, has presented a number of new approaches. Firstly, we present an approach to summarisation and indexing that, whilst based on TF/IDF, focuses on extracting the *essence* of a document before engaging in the inevitable identification of keywords. The argument here being that it is more useful from a KM perspective to understand the raison d'être of a document than focus on document storage mechanisms. Secondly, we present an approach to user profiling that encourages both pre and post-search collaboration against specific search requests. An additional feature of this user profiling approach is off-line proactive knowledge elicitation. The structure of these three features have their theoretical foundations in the work of Twidale et al [1997] who proposed a classification of social interaction that occurs during the information and knowledge retrieval processes. This classification mechanism highlights the relevance of *consultation, wandering, brainstorming, bibbling* and the serendipitous discovery process. We believe that ICTs that can support such processes will support the completion of the SECI cycle and create a community where knowledge can be created – the notion of *ba* [Nonaka and Treece, 2001]. Finally, the retrieval mechanisms we employ are based on mature IR strategies (primarily TF/IDF) but extend this approach to consider phrase structure and term positioning. From a technical perspective, the next stage of the project is to focus on the ICT interface this being the essential explicit representation of the KM environment.

REFERENCES

Baeza-Yates, R., Ribeiro-Neto, B. (1999). Modern information retrieval, Addison Wesley

Bates, M.J. (1979a). Information search tactics. *Journal of the ASIS,* 30 (4), pp. 205-214

Bates, M.J. (1979b). Idea tactics. *Journal of the ASIS,* 30 (5), pp.281-289

Davenport, T. & L. Prusak (1998). Working Knowledge: how organisations manage what they know. Boston, Harvard Business School Press.

Gardner, D. (1998). Knowledge that won't fit in a database - people, *InfoWorld*, 20(14), pp. 98

Kling, R. (1980). Social analysis of computing: theoretical perspectives in recent empirical research*, Computing Surveys*, 12, pp. 61-110

Kling, R. (Ed.) (1996). Computerization and Controversy, Academic Press, San Diego, CA.

Scarbrough, H., Swan, J. and Preston, J. (1999). Knowledge management: a literature review, Issues in People Management, Institute of Personnel and Development, London.

Luhn, H.P. (1957). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research & Development,* 1(4), pp. 309-317

Marchionini, G. (1995). Information seeking in electronic environments. Cambridge: Cambridge University Press

Mantovani, G. & Spagnolli, A. (2001). Legitimating technologies: ambiguity as a premise for negotiation in a networked institution, *Info. Tech. & People*, 14(3), pp. 304-320

Moens, M-F. (2000). Automatic indexing and abstracting of document texts. Kluwer

Nonaka, I. and Treece, D.J. (2001). Managing industrial knowledge: creation, transfer and utilization. SAGE Publications, London

Orlikowski, W.J., Yates, J.A. & Okamura, K. (1995). Shaping electronic communication: the metastructuring of technology in the context of use, *Organization Scienc*e, 6, pp. 423-44

Rowley, J.E., (1988). Abstracting and Indexing (2nd Edition). London: Clive Bingley.

Salton, G. (1975). A theory of indexing. Bristol, UK: J.W. Arrowsmith

Salton, G. (1989). Automatic text processing: The transformation, analysis and retrieval of information by computer, Reading, MA: Addison-Wesley

Shute, S.J. & Smith, P.J. (1993). Knowledge-based search tactics. *Information Processing and Management,* 29(1), pp.29-45

Sieloff, C. G. (1999). If only HP knew what HP knows": the roots of knowledge management at Hewlett-Packard. *Journal of Knowledge Management* 3(1): 47-53.

Skryme, D.J., (1999). Knowledge networking: creating collaborative enterprise. Butterworth Heinemann.

Sorensen, C. & Snis, U. (2001). Innovation through knowledge codification, *J IT*, 16 (2)

Sproull, L. & Kiesler, S. (1986). Reducing social context cues: electronic mail in organizational communication, *Management Scienc*e,. 32(11), pp. 1492-512

Swan, J., Robertson, M. & Newell, S. (2000). Knowledge management: when will people enter the debate, in Sprague, R.H. Jr (Ed.), *Proc of HICSS-3*, Maui, HI.

Swan, J., Newell, S., Scarbrough, H. & Hislop, D. (1999). Knowledge management and innovation: networks and networking, Journal of Knowledge Management,. 3(3), pp. 262-75.

Wenger, E (1998). Communities of practice: Learning, meaning and identity. Cambridge, Cambridge University Press

Xie, H. (2002). Patterns between interactive intentions and information-seeking strategies. *Information Processing and Management,* 38, pp.55-77