The International JOURNAL of KNOWLEDGE, CULTURE & CHANGE MANAGEMENT

Pitfalls and Potentials of Knowledge Management

Sunil Choenni
Saskia Harkema
Debora Moolenaar

VOLUME 5

COMMON GROUND

www.Management-Journal.com

INTERNATIONAL JOURNAL OF KNOWLEDGE, CULTURE AND CHANGE MANAGEMENT
http://www.Management-Journal.com

First published in 2005 in Melbourne, Australia by Common Ground Publishing Pty Ltd
www.CommonGroundPublishing.com.

ISSN: 1447-9524 (print), 1447-9575 (online)
Publisher Site: http://www.Management-Journal.com

The INTERNATIONAL JOURNAL OF KNOWLEDGE, CULTURE AND CHANGE MANAGEMENT is a peer refereed journal. Full papers submitted for publication are refereed by Associate Editors through anonymous referee processes.

Typeset in Common Ground Markup Language using CGCreator multichannel typesetting system
http://www.CommonGroundSoftware.com.

Pitfalls and Potentials of Knowledge Management

Sunil Choenni, Research & Documentation Centre, Ministry of Justice
Saskia Harkema, Utrecht University, Netherlands
Debora Moolenaar, Research & Documentation Centre, Ministry of Justice, Netherlands

Abstract: Organizations are interested in actively using knowledge / past experiences to improve their business processes. In this paper, we address some of the major issues in obtaining this goal. We discuss the importance of codifying and representing knowledge in computerized information systems, the distribution of knowledge to knowledge workers, and the interaction between knowledge workers via computer systems for knowledge management. We signal the pitfalls and potentials involved in these issues for knowledge management.

Keywords: (Perspectives on) Knowledge management, Codifying knowledge, Representing knowledge, Knowledge distribution, Information retrieval, Data mining

Introduction

CURRENTLY, THE FIELD of knowledge management receives a lot of attention from the business as well as the academic community. From an academic point of view the field of knowledge management raises many challenging research questions. Within companies the growing importance of knowledge management emphasizes the strategic value of knowledge. Organizations want to exploit their knowledge / experiences to improve their business processes. In order to achieve this goal, a number of steps, all evenly important, can be distinguished: the collection of knowledge / experiences, codifying knowledge/experiences in a formal system in order to store them in a computer system, the distribution of information / knowledge, and the application of information / knowledge in business processes.

In (Swan 1999), the author provides a comprehensive definition of knowledge management incorporating the above-mentioned steps. In this definition two perspectives on knowledge management are distinguished. One focuses on what is termed a cognitive model, the other on a community model of knowledge management. The first perspective stresses the technical issues that are of importance for knowledge management, while the second definition stresses the social issues. From the first perspective, knowledge is conceived as being captured and codified from individuals, packaged, transmitted, and processed through the use of Information and Communication Technology (ICT) and subsequently disseminated and used by other individuals in new contexts. The second definition focuses on social interaction and negotiation, and emphasizes the idea of supporting interaction and collaboration in order to manage knowledge. In this case knowledge is regarded as socially constructed through interaction within communities of practice. Knowledge is considered to be situated and contextualized. In this paper we mainly discuss a number of issues touched upon by the first perspective, though the second perspective will also shortly be discussed.

We focus on the potentials and pitfalls of codifying and representing knowledge in computerized information systems, the distribution of knowledge to knowledge workers, and the interaction between knowledge workers via computer systems. In addition, we extend our discussion to what extent technology supports these three aspects of knowledge management. Note that although the notions of data, information, and knowledge are subjective, it is widely accepted that a sensible distinction can be made (Bocij et all 1999).

Before knowledge can be codified and represented it needs to be collected. The collection of knowledge/experiences is a typical task of knowledge engineers. In general, a knowledge engineer performs this step by means of literature review, interviews, and protocol analysis. The codification of knowledge/experiences into a formal system is a task of computer specialists. There has always been a tradeoff between the simplicity and the expressive power of a formal system. The business community requires formalisms with an expressive power that is easy to understand, but it appears that an efficient implementation of this type of formalisms is not easy. Another question that still needs to be answered is: given some body of knowledge represented in a formalism,

http://www.Management-Journal.com, ISSN 1447-9524 (print), 1447-9575 (online)

how can we efficiently update this body of knowledge if this is needed. Often such updates lead to inconsistency in the knowledge as will be shown in this paper.

Tools that support the effective dissemination of knowledge are growing due to research and development in the fields of information retrieval (Baeza-Yates & Ribeiro-Neto 1999) and data mining (Fayyad et all 1996). In the field of information retrieval, effort is put into building systems that are capable of handling information needs of a user. Information needs formulated by a user are not exact, as they are in traditional (database) applications, but rather vague and incomplete. Often an information need is expressed by a set of keywords. Suppose that we have a system containing a digital library and a user needs to gain some information about information systems. He / she consults this system by typing the keywords "information systems" in order to find all relevant documents that deal with this subject. There may be many documents about this subject, making it non trivial to select the proper documents for this user. Furthermore, a document dealing with information systems might be quite interesting for one user but not for another user, while both might express their information need by the same keywords. Another issue that should be taken care of by information retrieval systems is that keywords that are not entered by the user might still be interesting for him / her. For example, since a database is a major component of an information system, documents dealing with databases might be interesting for a user who entered "information systems" as keywords, even though database is not mentioned as a keyword. By means of an interactive session with the user, an information retrieval system attempts to discover what precisely the information need of a user is and to meet this need of the user. The basics concepts and techniques that are used in the processing of information needs of a user will be discussed in this paper. The basic of another topic that will be covered in this paper is data mining.

The fields of data mining and information retrieval both aim to meet information need of a user. The difference between them lies in the type of information need both fields deal with. An information need expressed in the context of data mining has a higher degree of vagueness and incompleteness than an information need expressed in the context of information retrieval. The goal of data mining is to extract implicit, previously unknown, and potentially useful knowledge from large data sets. The extracted knowledge may support or be used in strategic decision-making. A typical mining question for instance -- in the context of our supermarket example -- is: find me interesting profiles of clients that have not been discovered so far. It should be clear that although we discuss data mining as a knowledge distribution mechanism, it might be considered as a knowledge creation technique as well.

As noted, from a community model perspective on knowledge management interaction is a primary importance. Tools that support interaction between knowledge workers are growing and becoming more advanced. Tools like email, instant messaging, chat boxes, etc. contribute to the communication and collaboration of knowledge workers. The role of these tools in knowledge management will be discussed in this paper as well.

The remainder of this paper is organized as follows. Next, we discuss the fundamental issues of capturing knowledge in formal knowledge representation systems and the maintenance of knowledge. Then, we focus to the basics of information retrieval and data mining and the importance of these fields for knowledge dissemination. Then, we provide a brief overview of the state of ICT tools that support knowledge management from a community model perspective. Finally, we conclude the paper.

Capturing Knowledge

In order to store knowledge in a computer system, this knowledge should be expressed in a formal representation model. A knowledge representation model consists of two parts, 1) a knowledge model that provides the possibility to represent knowledge in formalism, and 2) an inference scheme that consists of a set of rules/operators that can be used to manipulate knowledge. For example, logic can be considered as a knowledge model and the inference rules in logic can be used to derive new knowledge. Suppose we have collected the following piece of knowledge: all human beings are mortal and Joe Sixpack appears to be a human being. This can be modeled by the following two premises:

(1) ∀x: human-being(x) → mortal(x) (the symbol ∀means **for all** and x is a variable)

(2) human-being(Joe Sixpack)

Now an inference rule, the so-called "modus pones", allow us to derive from (1) and (2) a new fact namely, mortal(Joe Sixpack), which means that Joe Sixpack is mortal. We note that the "modus pones" draws conclusions on the basis of the form of a set of premises.

Through the years many knowledge representation models have been developed, such as rule-based models, decision trees, object frames, and so on. In general, the simpler a formal system is to understand by human beings, the less expressive it is. An example of such a formalism is the rule based model, which consists of a set of rules, which have the form *IF* condition *THEN* action_x *ELSE* action_y. Such a model is very intuitive and easy to understand, however, only a limited number of problems can be expressed in this way. For example, modeling the relation that Y is an ancestor of X whenever Y is a parent of Z and Z is an ancestor of X gives all kind of looping problems if we use rule based model. Although this problem can be expressed in a very elegant way in first order logic as follows: Ancestor(X,Y) $\leftarrow$ Parent (Z,Y); Ancestor(X,Z), this is in general considered as difficult to read.[1]

In general problems with a limited number of variables that are well understood can be modeled and handled by rule-based models. However, if a condition in an IF clause is dependent on the THEN clause of another rule; this considerably complicates the modeling process. So, knowledge representation models that capture interaction between variables lead to knowledge representation models that are not easy to understand. Other notions, like uncertainty and vagueness, which are inherent to many real-life systems, also lead to complicated models. For example, efforts have been reported in the literature to extend if-then-else rules with uncertainty, in which an uncertainty measure is assigned to the action part. It has been shown that not only these models lead to counter intuitive results in some cases, but its applicability is limited due to the complexity. A recent approach to capture uncertainty in the context of databases is discussed in (Choenni et all 2004).

As stated in the foregoing, it is hard to express knowledge in a knowledge representation formalism, and therefore to capture it in a computerized system. Through the years it has been shown that it is even harder to update knowledge in an effective and efficient way. This is due to the fact that machines are equipped with monotonic reasoning techniques. If a rule T $\rightarrow$ Z holds, then T $\cup$ Q $\rightarrow$ Z should also hold according to a monotonic reasoning system. We illustrate by means of a small example the limitation of monotonic reasoning systems for real-life business applications. Suppose we have a knowledge base (KB) that contains the following rule:

(1) Birds can fly

Suppose we want to add that parrots can fly in KB, then KB can be extended to KB+, as follows: (1) Birds can fly; (2) Parrots can fly.

But if we want to add rules that pertain to an ostrich in KB+, this will lead to an inconsistent knowledge base KB+ since an ostrich is a bird but it cannot fly. To add facts and rules about ostriches in KB+, rule (1) needs to be updated in order to keep our knowledge base consistent. For example, we modify rule (1) into: All birds, **except** ostriches, can fly resulting into KB++: (1) All birds, **except** ostriches, can fly; (2) Parrots can fly; (3) Ostriches.

If we would like to add knowledge about baby birds that may not fly; then rule (1) needs to be modified again. It is clear that since there are many exceptions, it is infeasible to include all exceptions in a system. This implies that whenever we add new knowledge in a knowledge base, all knowledge in the knowledge base needs to be checked and possibly updated in order to keep the knowledge bases consistent. This is a laborious and tedious task.

Since in many business applications knowledge evolves over time, e.g., business rules and procedures in organizations change due to new laws, maintenance of knowledge is of vital importance. As noted before, this is at best a tedious and laborious task. Therefore, we need non-monotonic reasoning techniques that are able to (semi-) automatically maintain knowledge bases. Since there is no significant progress in non-monotonic reasoning techniques, the maintaining of knowledge bases may become a bottleneck in knowledge management. Our observation that the development of knowledge bases is restricted to small and well understood domains is confirmed in (Davenport & Glaser 2002) by means of a case study. We expect that knowledge management provides more opportunities in business applications where we avoid the modeling and maintaining of abstract knowledge, such as information retrieval and data mining applications. In the next section, we discuss information retrieval and data mining as tools for knowledge management.

Knowledge Dissemination

The explosive growth of storage capacity in computer systems has led to the storage of vast amounts of data in, e.g., databases and on the web. An often heard complaint in the business community is that we are drowning in data, but starving for information. The challenge is to provide users the information that they are looking for from the enormous amount of data stored in computerized information systems. This challenge is studied in the fields of information retrieval and data mining. In information retrieval we consider a set of documents, e.g., papers, books, web pages and so on, stored in a system, with the goal to select those documents that match the information need of a user. In data mining, we take as

[1] The predicate Ancestor(X,Y) should be read as Y is an ancestor of X

starting point large databases, which often consist of well-structured and true data, with the goal to extract potentially useful information for a user. In the following two subsections, we discuss the basics of information retrieval and data mining. Furthermore, we discuss the usefulness of these fields for knowledge management.

Information Retrieval

The growth of the web has entailed a boost in the development of modern information retrieval systems. The challenge the user faces is finding the information he / she needs. Many modern information retrieval systems, like search engines, are designed to facilitate the user in this search for useful information. Compared to the traditional information retrieval systems, modern retrieval systems are designed for ordinary users, i.e. those unfamiliar with the available collection of documents in a system or on the web, the representation of documents, and the use of retrieval operators. This implies that requirements imposed on modern retrieval systems are different from those imposed on traditional systems, like data retrieval systems. In Table 1, we list some of the major differences between data and information retrieval systems. As we can see, for data retrieval systems, a question needs to be formulated in a formal query language, which is in turn used to search for data that exactly match the question. Therefore, a data retrieval system is capable of returning exact answers without errors to the user.

Table 1: Differences between Data Retrieval and Information Retrieval

Aspect	**Data retrieval**	**Information retrieval**
Matching	Exact	Partial & best
Model	Deterministic	Probabilistic
Query language	Formal	Natural
Answers to questions	Exact	Relevant
Output sensitivity to errors	No	Yes

In a modern information retrieval system, we have on the one hand the contents of an object, e.g. documents, represented in one or another (formal) way, and on the other hand we have an information need often represented in natural language. The goal is to find relevant and useful matches between the information need and the contents of the objects represented in the system. In order to implement these systems, a number of basic steps should be supported. In Figure 1, we have depicted these steps as discussed in (Croft 1993). In the figure, square boxes represent data and ovals represent processes. We assume that we are only dealing with documents.

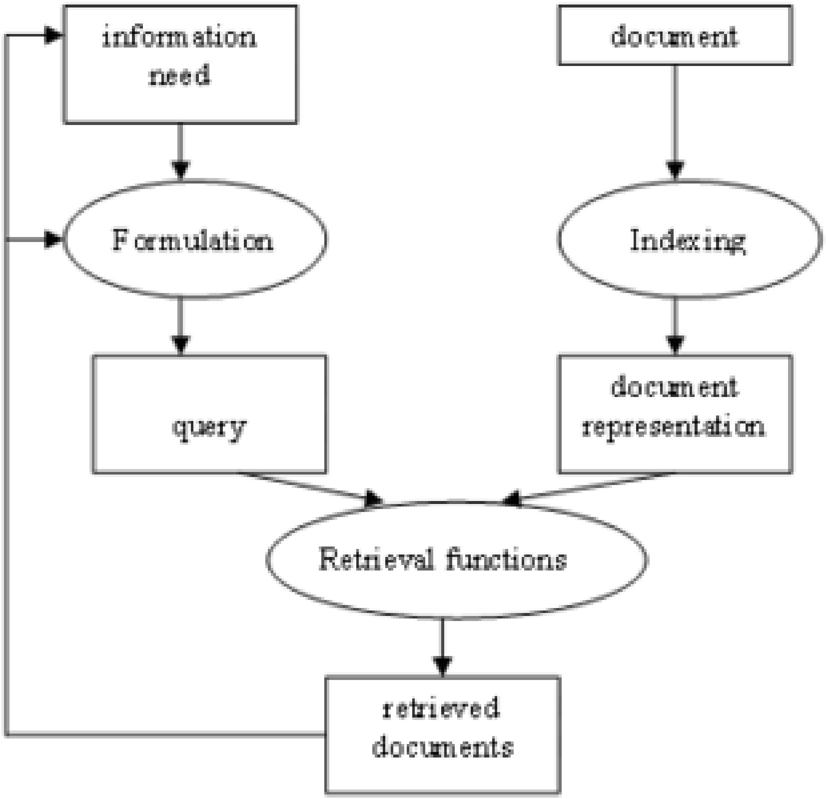

Figure 1: Basic process in an information retrieval system

Representing the documents is usually called indexing. The indexing process results in a formal document representation. For example, a document may be represented by a set of words that covers the content of a document or just by the title, abstract and storage location.

The translation of an information need into a set of queries is called formulation. On the basis of these queries and the document representations, retrieval functions determine the degree of matching between the set of queries and each document. The documents that match best are retrieved and delivered to the user, who in turn may provide feedback on the delivered documents. This feedback is considered as new input in the formulation process, which will lead to a new set of queries, and the matching between these queries and document representations will start again. In this way an interactive session between a user and the system is established, which leads to a better understanding of the information need for the system as well as for the user.

Information retrieval is one of the cornerstones of knowledge dissemination for the following reasons. First, the accessibility of information retrieval systems for ordinary users is high since it allows these users to express their information needs in a natural language. Second, development in this field attempts to deal with the subjective perception of information of individual users. An adequate handling of subjective perceptions is crucial in knowledge management, since different users may have different association with the same set of data or information. As a consequence, a same set of keywords may result into two different output for two different users. This can be obtained by extending the keywords with information about a user, e.g., by automatically extracting information from the homepage of a user. We believe that the development in the field of information retrieval may give a boost to the feasibility of computerized knowledge management systems.

Data Mining

The field of data mining has been developed rather ad hoc, sometimes using vague concepts. Many informal definitions can be found in the popular press about data mining, such as the search for knowledge, patterns, regularities and so on. But let us take a closer look at what formally happens in data mining.

Data mining algorithms induce models from large databases, which contain observations from the real world. The goal of inducing a model is to provide insight in a phenomenon of interest that is part of the real world. This insight may help in understanding the phenomenon, or it may help to predict the outcome of similar phenomena. Although data mining algorithms induce models from a large set of observations from real-life, this does not necessarily mean that these models are correct. The explanation for this fact is that an induction process is not truth preserving. We illustrate this by the following example. Suppose we have a database that records data about swans. In this database, the color of all swans appears to be white. Under the closed world assumption, the conclusion “all swans are white” is correct, but this might not be true in the real world. The fact that we have observed only white swans (until now in the real world) does not mean that all swans are indeed white. It is possible that there are black swans, but we have not observed them yet, and therefore they are not recorded in our database. This implies that a model obtained by data mining should be tested on its validity and to what extent it deviates from the real world. This can be done by comparing simulation results to real-life results. In Figure 2, the formal data mining process is depicted.

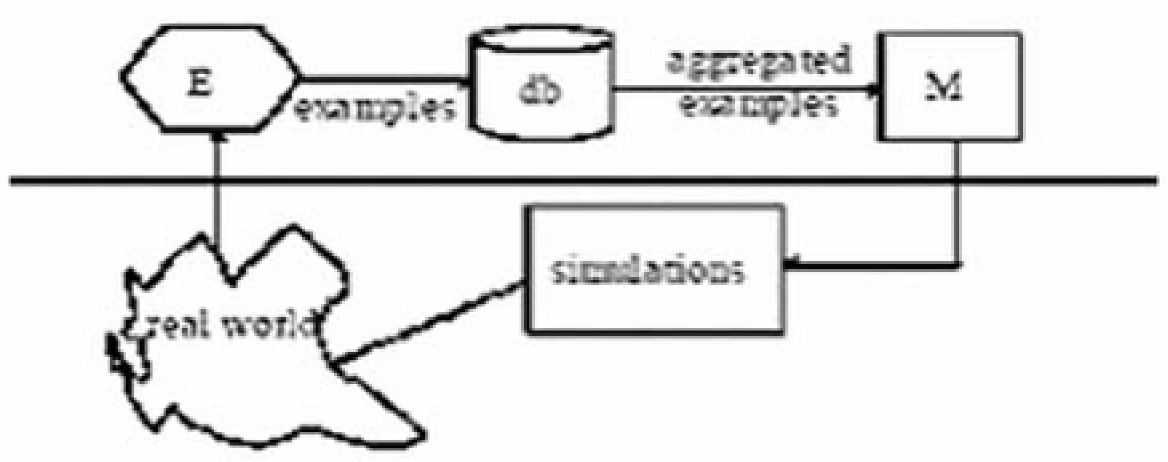

Figure 2: Basic Process of a Data Mining System

Applications of data mining technology can currently be found in a wide variety of business fields. Airline companies analyze historical reservation data in order to get a better profile of their customers. In the field of marketing, data mining technologies are used to decide which customers to send an advertisement and which not. Retailers analyze historical supply and demand data to detect trends that help in planning sales promotions and optimizing their purchasing. Supermarkets are looking for associations between items that improve the organization of the items in their shops. Data that pertain to the performance of large complex systems are analyzed to detect abnormal behavior. Insurance companies use data mining algorithms to discriminate between “good” and “bad” clients. And the list of data mining applications is still growing.

Data mining is in fact a step in a larger process, the so-called Knowledge Discovery in Databases (KKD) process (Fayyad et all 1996). In general, data mining is a highly interactive process (Wrobel et all 1996). In practice, users start with a rough idea of the information that might be interesting. During the mining session the user more explicitly specifies, based on, among others, the mining results obtained so far, which information should be searched for.

Data mining may become one of the driving forces behind knowledge management, since it focuses on the search of interesting strategic knowledge from operational databases. Since interestingness is a subjective notion, frameworks are in development to capture this notion in order to support the last step of a KDD process. Furthermore data mining is able to deal with knowledge that evolves over time in a natural way, which is a requirement for knowledge management. Suppose that in a certain period, people often buy diapers and beer together. This behavior is recorded in an implicit way in a transaction databases. Mining this transaction database will expose this implicit relation. Now, if the buying behavior of the people changes, e.g., they do not buy beer and diapers together any longer but chips and coke, the transaction database will record less combined selling of beer and diapers and more combined selling of chips and coke. If we mine the database we will find the new relationship between chips and coke, and we will no longer find the relationship between diapers and beer since this relationship has become obsolete. Given the major progress in data mining systems we expect that this type of systems will be a part of computerized of knowledge systems in future.

Interaction

As stated in the foregoing, social interaction is an important aspect of knowledge management. Distributing knowledge via social interaction eliminates the need for collecting and codifying knowledge. As we have seen, some of the major pitfalls of knowledge management can be found precisely in these two processes.

Given the increasingly global nature of firms, the role of ICT in supporting interaction will become more and more important, since meeting face-to-face often is not possible. Many tools that support interaction exist, for instance e-mail, instant messaging, video conferencing, discussion boards, file sharing, etc. Often these tools are not thought of as supporting knowledge management, but they are important as they help employees interact even when they are located far apart.

One consequence of the community model of knowledge management concerns the nature of knowledge, namely that it is dynamic and context dependent. This implies that an individual will need to learn the knowledge relevant to a situation / task at the time it is needed. We would therefore like to support employees in their learning. Supporting this learning with ICT is often called e-learning. We distinguish three main properties an e-learning tool should support:

- It should focus on the needs of the learner, given his or her background,
- Consequently, it should support a personalized learning process, giving the learner the option to choose his or her own learning pathway,
- It should offer several learning methods (e.g. a virtual classroom, collaborative tools, self-paces instruction, etc.) (Choenni et all 2002),

The current state of technological development does not allow these properties to be fully supported by ICT tools. Especially the personalization of tools (a consequence of properties 1 and 2) is not yet completely possible.

All in all social interaction has three roles to fulfill in knowledge management. First, it plays a role in individual learning (see property 3 of an e-learning tool above). Secondly, individual learning needs to be shared within the organization in order to become shared knowledge (distribution of knowledge). Thirdly, through interaction a community of practice is created in which the exchange of knowledge is made easier because of shared routines, words, tools, ways of doing things, stories, gestures, symbols, actions and concepts (Harkema 2004). In other words, there is a recursive effect in that interaction facilitates interaction.

In conclusion, we can state that ICT tools are particularly useful to support interaction within an organization. This will help both individual learning as well as organizational learning. Other tools to support individual learning need to be developed further before they reach their full potential.

Conclusions

Since the field of knowledge management is a broad one, many issues need to be considered for the implementation of computerized knowledge management systems. In this paper, we study the importance of a number of issues for knowledge management namely, codifying and representing knowledge in computerized information systems, the distribution of knowledge to knowledge workers, and the interaction between knowledge workers via computer systems. In addition, we study to what extent technology supports these three aspects of knowledge management. Our overall conclusion is that the implementation of some aspects of knowledge management is feasible and may be exploited by the business community. For other aspects of knowledge management we argue that there is hardly technological support, and therefore, the implementation of these aspects will probably lead to a failure or only restricted versions of these aspects can be implemented.

We argue that, e.g. other than monotonic reasoning techniques are required to update knowledge bases and more accessible knowledge representation formalisms are required by the business community in order to codify and maintain knowledge. With regard to the other two aspects, the distribution of knowledge to and the interaction between knowledge workers, we are positive. We observe that a major progress has been made to serve the business community with tools that support effective distribution of knowledge. Also many ICT tools exist that suffice in supporting communication / collaboration between knowledge workers. Research and development in the field of information retrieval, such as search engines, and data mining technology significantly contribute to an effective distribution of knowledge. Both fields pay amongst other attention to how to handle the subjective perceptions of users. Tools like email, instant messaging, chat boxes, etc. contribute to the communication and collaboration of knowledge workers.

References

Baeza-Yates, R., Ribeiro-Neto, B. (1999) *Modern Information Retrieval*, Addison Wesley Longman/ACM Press, USA.

Bocij, P., Chaffey, D., Greasley, A., Hickie, S., (1999) *Business Information System,* Prentice Hall, England.

Choenni, R., Blok H.E., Fokkinga, M. (2004) *Extending the Relational Model with Uncertainty and Ignorance*, Technical report, TR-CTIT-04-29, 16 pp., University of Twente, The Netherlands.

Choenni, R., Walker, R., Bakker, R., Baets, W. (2001) "E-learning as a Vehicle for Knowledge Management, *International Conference of Applications of Prolog 2001.*

Croft W.B. (1993) "Knowledge-Based and Statistical Approaches to Text Retrieval", *IEEE Expert 8* (2), pp. 8-12, 1993.

Davenport, T.H., Glaser, J. (2002) "Just-in-time Delivery comes to Knowledge Management", *Harvard Business Review*, pp. 107-111.

Fayyad U., Piatetsky-Shapiro, Smyth P., Uthurusamy, R. (eds)(1996), *Advances in Knowledge Discovery and Data Mining*, AAAI/The MIT Press.

Harkema, S. (2004), *Complexity and emergent learning in innovation projects: An application of complex adaptive systems theory*, Ph.D. thesis, Nyenrode University.

Swan, J., Scarbrough, H., Hislop, D. (1999) "Knowledge management and innovation: networks and networking", *Journal of Knowledge Management*, 3(4), pp. 262-275.

Wrobel S., Wettschereck, D., Verkamo, A. I., Siebes, A., Mannila, H., Kwakkel, F., Kloesgen, W. (1996) "User Interactivity in Very Large Scale Data Mining", *Proc. FGML-96 Annual Meeting of the GI Special Interest Group Machine Learning*, pp. 125-130, TU Chemnitz-Zwickau.

About the Authors

Dr. Sunil Choenni

Sunil Choenni is heading the department of Statistical Information Management and Policy Analysis at the Research & Documentation Centre (WODC), which is an institute of the Dutch Ministry of Justice. Prior to joining WODC he held several teaching and research positions at different universities and research centres. He published many papers in the field of Information Management and Business Administration and serves as PC member in a number of international conferences.

Saskia Harkema

Saskia Harkema received her PhD from Universiteit Nyenrode for her dissertation "Complexity and emergent learning in innovation projects". She is currently Assistant Professor at the Universiteit Utrecht and Associate Professor at The Hague University of Higher Education. Her research interests include innovation management, complexity theory and the learning organization.

Debora Moolenaar

Debora Moolenaar received her PhD from the University of Amsterdam for her dissertation "The Dutch market for agency work". She is currently working as a researcher at the department of Statistical Information Management and Policy Analysis at the Research & Documentation Centre (WODC), which is an institute of the Ministry of Justice in the Netherlands. Her research interests include econometrics, forecasting models, law & economics, labor market economics, benchmarking and efficiency analysis.

THE INTERNATIONAL JOURNAL OF KNOWLEDGE, CULTURE AND CHANGE MANAGEMENT

SCOPE AND CONCERNS
Visit: http://www.Management-Journal.com

SUBMISSION GUIDELINES
Visit: http://www.Management-Journal.com

INQUIRIES
Email: cg-support@commongroundpublishing.com