Knowledge Management and Information Retrieval:

Some New Challenges

Eng. Ab L.M. Helderman

PHEIDIS CONSULTANTS

Zoetermeer, Netherlands

Email: helderman@pheidis.

Abstract

Change is one of the stable factors in many organisations. Therefor agility is essential for success and good management of knowledge is one of the key factors for success in such organisations. This document describes a model for the organisation of knowledge management. This model is based on a distinguishing of concepts of knowledge: the stock concept and the flow concepts. The importance of a knowledge management policy is stressed. In such a policy the key strategic knowledge requirements of an organisation are defined based upon the organisations mission, primary goals and strategy. Also the knowledge management infrastructure to implement such a policy is elaborated.

Further the paper focuses on the requirements of information retrieval software for such a knowledge management infrastructure. It deal with the concept "document" and the need for user friendly functionality. It concludes that many of today's search engines are too complicated and it advises to invest in the development of software that helps the user to use advanced functionality that retrieval systems have.

One of the main deficiencies of modern retrieval software is that it is based on the presumption that the meaning of a document is completely and only depending on the document itself. It does not take the user into account and the domain from which he is searching. A document however should be interpreted in the terms of the user and his organisation. Modelling knowledge from the perspective of the user is one of the key challenges for the coming years.

1 Introduction

The subject of the presentation is "knowledge management and information retrieval".

I think that the knowledge management 'hype' will have an enormous influence on the utilisation of Information Storage and Retrieval technology in the coming years. Until a few years ago ISR-technology was the domain of specialists in libraries and documentation centres. The text database – that is what I am talking about - was not part of the "real" IT-world. Relational databases and transaction processing systems were the issue there.

But since a number of years we see a growing need to deal with text in the ICT-world. That is caused by the fact that more and more knowledge is created and condensed in documents and secondly also because more and more of these documents are digitised. A growing number of documents are produced, stored, and exchanged with computers and more and more documents are made available for the computer. The Internet and Internet technology plays an important role in this.

Managing these document and its contents becomes a major issue. The knowledge management 'hype' is one of the phenomena indicating that. People and organisations have problems coping with the enormous amount of information that is available. I think that this development will make or made ISR-technology a common technology.

In our consultancy practice we help organisations managing their knowledge. One of the aspects in that is that we assist these organisations in selecting and applying the proper information storage and retrieval technology.

For today I developed a few models on some aspects of the way knowledge management could be organised within an organisation. Based on that I want to focus on the information retrieval functionality that is required within such an environment. Maybe this gives you ideas on the research and development that is required in the coming years to make IR-system better for knowledge management purposes.

2 Knowledge in an agile organisation

It is not refined but you could say that in the early days you could do your job after you finished your education,

and read your daily newspaper and professional magazine. Maybe for some professions you needed 2 or maybe 3 magazines, but that was it. More was the real exception.

Things have changed. Organisations are operating in a fast changing environment. Competition grows also caused by deregulation. Market demands are changing, higher quality is needed, mass production is replaced by production on specification and demand, the influence of the society on the organisation changes, environmental protection, etc. all influences that put organisations in a turbulent environment.

Technology, specifically information and communication technology, plays an important roll in this. The speeds in which developments follow each other are – for a great deal - made possible by ICT. Think about the Internet for instance.

Change is the only stable factor in many organisations, these days. Agility¹ has become one of the main aspects in a modern organisation. Today, becoming and staying agile is the most fundamental factor for success.

Such developments require the utmost of the organisation to be able to adapt to the continuous changing circumstances. Knowledge plays an important roll in this. It is an essential element in the learning process in the organisation. But it is more. In an increasing number of organisations it is the fourth production factor, that is essential to be able to survive in the changing environment I described.

A conclusion is that the organisations should manage the way they deal with knowledge. That is what knowledge management is about.

I will not attempt to give a definition of knowledge. A lot has been written on it and I have read too much of that to think that I can add something here. But I know that knowledge is something between the ears and it has something to do with information and the ability to create and apply it.

In its essential sense knowledge is a “flow concept” (“Information in motion”). It is always in motion. Knowledge is a quality of people and is related to their daily handling. In this concept working and communicating people create and exchange knowledge. Knowledge management has to do with managing aspects such as communication, learning, habits, etc. The flow concept of knowledge is a valuable concept also for information and communication technology. ICT can be used to support working and communicating people.

In another sense however you can look at knowledge as a “stock concept”. Knowledge is something concrete and tangible. It can be exchanged and it can also exist on its own. You can “condense” it to a medium, record it or commit it to paper. In this concept you can store knowledge and buy and sell it. I would call it information then or - if you want - data.

In this area information storage and retrieval is important. For that reason I want to concentrate on the ‘stock concept’ of knowledge.

Figure 1: The flow concept and the stock concept of knowledge

Knowledge management in the context in which we discuss it here is managing and administering the knowledge-related processes directed towards the mission, goals, and strategy of the organisation. It is done by means of planning, implementing and evaluating the knowledge-related processes.

Because I want to concentrate on the “stock concept” of knowledge, for me knowledge management is the domain in which you manage and administer the knowledge-related processes based upon the exchange,

¹ Luc Chalmet, Agile competition and virtual organisations, PA Consultants, 1996

storage and retrieval of information, directed towards the goals, etc.

It is the word “directed” that has to be emphasised here. Knowledge management is not something on its own. For the agile organisation it is the added value of knowledge management and information retrieval that is important. Particularly when you look at knowledge management from the stock concept – where it is seen as a separated process – this notion is very important.

3 A model for a knowledge management organisation and infrastructure

To further visualise the stock concept of knowledge I developed the model in figure 1. I emphasise that I do not intend to present a model of the reality. It is a thinking model. It is meant to help us understand the stock concept of knowledge management and the infrastructure and organisation we need to implement that concept. This model contains the following 3 elements

- The area where the knowledge is created and used and where you find primary sources of information (internal and external) in which this knowledge is condensed.
- The knowledge management policy
- The knowledge management infrastructure

Let us look at the last two elements in some more details.

3.1 Knowledge management policy

Essential in the model is the knowledge management policy. It is the policy in which the organisation makes clear which knowledge is required for its well being. It is based on the organisations mission, primary goals, and strategy. What does it describe:

- Related to information needs: It describes on which issues the scarce resources that are available for knowledge management are to be spent.
- It describes which knowledge is of interest to the organisation. It classifies this knowledge in – what they call – the ‘corporate knowledge’ (the strategic highly important knowledge), and other types of knowledge.

Next it describes the policy related to these different categories of knowledge. It describes for instance that the corporate knowledge is classified and that measures should be taken to prevent others from using it, or that the corporate knowledge should be owned (between the ears) by employees of the organisation.

For a company like Honda for instance the knowledge related to engines and engine technology is corporate knowledge. On this issue they want to be number one in the world. Knowledge related to engines and engine technology is protected and classified. In other area’s, for instance related to the application of this engine technology in automobiles, lawn mowers, etc. it is acceptable to buy knowledge or use it from others (for instance Rover).

Of course the plan also deals with marketing knowledge and information. It states which market research is done internally and which knowledge can be bought from outside. It classifies and plans the activities related to the maintenance of knowledge on competitors, etc.

- It translates the operational plans for production, export, research and development, etc. to the specific information needs that could help to implement these plans. It states for instance that information is required on export markets and habits in South America or that information is required on product legislation in countries in these parts of the world.
- Related to planning and organisation: it describes the organisation, resources, planning, architecture and infrastructure to implement the knowledge management infrastructure. A lot of words describing who and how en when the activities related to knowledge management should be implemented.

Of course, drafting and maintaining such a plan is a complicated and difficult job. Wrong choices can have enormous implications. But – if knowledge is a production factor – then the consequences of no choices will be bigger. In fact the plan describes in what way the scarce resources that are available for acquiring and processing knowledge are spend related to the primary mission, goals and strategy of the organisation.

Figure 2: the Stock concept of knowledge

It is clear that drafting such a plan requires a lot of involvement from top management of the organisation. Choices should be made on that level.

3.2 The knowledge management infrastructure

This knowledge management infrastructure describes the facilities and procedures the organisation created to implement the knowledge policy. It contains (see figure 2 and 3):

- The facilities for the selection, the recording, the enrichment and storage of relevant knowledge: With facilities I mean the organisation, the procedures and the technical functionality that is required. Some important aspect of these facilities are:
 - The quality of the information is highly essential. Without this, advanced retrieval functionality for example is of no use.
 - Of course, IR functionality could be used for the selection of external information for inclusion in the corporate knowledge database. I think that intelligence – based on for instance knowledge models derived from the knowledge management policy - could be used to select that information
 - For the editorial and quality control process the support by a workflow environment could be required. With the help of such an environment the procedures of the editorial and quality control process can be implemented.

Figure 3: Infrastructure of knowledge in stock

© Phcidis Consultants

- The facilities for the retrieval and dissemination of information from the databases. The information from the database will be used by individual users. For this they will make use of a search interface with all the retrieval facilities and intelligence it may have.
You should be aware however that it is also required to publish information from the database on a regular basis to groups and individuals. The retrieval profiles that feed this dissemination process could be straightforward but also intelligent.
Also here the assistance by workflow functionality could be required to support the retrieval and publishing process. In some environments – such as the aeroplane industry and the petrochemical industry – this workflow can be necessary to completely automate the process for the retrieval and publishing of for instance updated procedures and policies stored in the knowledge database.
- The information storage facilities. Related to the storage facility I want to make 3 remarks:
 - The storage facility contains both internal and (linked) external databases. It is a distributed environment.
 - To enable the publishing process on different media the storage of data should be as far as possible media independent. The use of standards like SGML and XML is obvious here.
 - I think that the functionality related to the management of the meta data that defines the data structure and contents (thesauri, classifications, etc.) is highly essential. Such meta data is required to support fast and precise retrieval. However the flexibility that is required – see my remarks on the agile organisation – could be hindered by the concrete structure that databases based on meta data tend to have. This is an interesting area for improvement.

4. Some new challenges for IR in knowledge management

In this chapter I want to focus on some of the key features that should be supported by the IR infrastructure of today and tomorrow in the knowledge management environment I described. It is primarily based upon practical experience and only secondary on some research I have done on the state of the art on science in this area.

There are many ways to evaluate IR-technology for knowledge management. I read an article² for instance where two aspects of some approaches were brought together in one diagram. On one axe the externalisation efficiency (“what does it cost to classify a document”) was scaled and on the other the internalisation efficiency (“how easy is the retrieval”). It resulted in a diagram like in figure 4.

I have a problem with such evaluations. They compare approaches and products and that is useful but it does not give answer to the question: do these approaches and products fit the needs of the user. I think the user should be brought into the picture. We should look at the way the knowledge worker wants to work or works in an IR environment. The first 3 issues I want to speak about are on this issue.

Figure 4: Evaluation of some IR approaches

4.1 Document as a concept

My first remark is on the concept “document”. I think that for knowledge management – and maybe for IR in general - “document” is not a useful concept anymore. This technology is not meant for the disclosure of paper collections anymore.

In our information age we do not browse through documents but we browse through “objects” of knowledge. We do not go from page 1 to page last in a book but we “travel” through the information space and we are satisfied when our information needs are fulfilled. In many cases it is unclear, even for the user himself, how we gained our knowledge and which documents played a role in that. The user jumps from one page or object in one document to another page or object in another document. He does not see documents anymore.

This is an important observation also because it will have a dramatic influence on the way information will be produced and published in the future by the publisher but also in the knowledge environment I spoke about.

Think for instance about the influence this will have on the publishing market for educational material. I have to leave this subject aside and limit myself to my remark on the concept “document” related to IR. I think it is useful to think about the way we are going to produce and structure knowledge in the future and the way we are going to classify documents and the objects they contain.

4.2 Easy in use

IRS becomes common technology. Today, every knowledge worker uses it, and that made a big problem visible.

A lot of research has been done on information retrieval. Next to the straight forward functionality's

² Stanoevska, K, A. Hombrecher, S. Handschuh, B. Schmid, Efficient Information Retrieval for Knowledge management, www.netacademy.org

based on inverted files and Boolean operators we developed rule based systems, self learning technology and other intelligent ways that could help the user to find his way through the ever increasing pile of information that is available today.

When we look at the Internet however – the information retrieval system of this time – we see that nothing or almost nothing of that advanced functionality is used. The result is that most queries on the Web produce an overwhelming result.

The capabilities that search engines have (if they have them) to provide accuracy and relevance in a query are not used. For many users it is too difficult or it takes too much time to find out how this advanced functionality works and can be applied.

My conclusion is that today's search engines are too complicated. Therefore, where possible, software intelligence should be used to help the user to use the advanced functionality that retrieval systems have.

This is particularly important because knowledge management is one of the primary processes in the modern organisation. It is a key process that will not be left in the hands of information retrieval specialists or intermediaries. It is business managers business. It is end user business and the end user him- or herself should be able to use the technology involved. No long or supportive activities are acceptable to help the end user interpreting and modifying the question under discussion.

4.3 Interactive knowledge modelling

Important is the fundamental aspect that most retrieval systems are based on the presumption that the meaning of a document is completely and only depending on the document itself. The document is understood in the domain in which it was made. Most document retrieval systems - also the modern ones - try to understand what the author meant when he wrote the document. The emphasis of the functionality in the system is directed toward a representation of that meaning.

This is particularly true with functionality as hypertext. The relations made are the relations of the author, or at least, of another user, while in fact we are interested in relations that should be made based upon the thinking model of the actual user. Or even further, in a knowledge organisation, we are looking for new relations and new models to represent new ideas.

It is true that if we use free text searching, then we search through the database based upon search items of the user. But we should be able to go much further. We should be able to understand and interpret a document or a document collection in the context of the information need of the user or its organisation. We should be able to analyse a document and represent it based upon a model of a specific information need and its context. A document should be interpreted – as my former colleague John Mackenzie Owen³ mentioned– in terms of the organisation, the individual or the particular question.

And this process should be highly interactive (see figure 5) because it is not true that the user knows where he is looking for. If he knew we would not require all this intelligence in IR systems but just a simple database. The problem is that the user does not know what he is looking for and does not know how to interpret the document collection he is searching in. We are in a dilemma and the IR intelligence should help the user out of this situation. It should help the user to build its model of the information need, it should help to model the document collection based upon this user model and it should help to compare and interact these models so a useful set of documents could be retrieved from the database collections. Further it should help to evaluate these documents in the context of the user model so the game could be played for a second and a third time, until the information needs of the user are satisfied.

3 Prof. J.S. Mackenzie Owen, Kennis in Veelvoud, Vossiuspers AUP, 1988

Figure 5: Interactive knowledge modelling

In fact the needs of the user are satisfied when this new knowledge model is created, and not when relevant documents are retrieved. The documents only play an important roll in that. I call this process 'interactive knowledge modelling'.

How could models like I describe look like? I do not know. It is an interesting question for research I presume. It should be based on knowledge structures existing in the organisation or the mind of the user. Such a model could be based on a workflow model of a business process of the organisation or of one of the competitors (best practice models). It could be based on a model of the market. It could be a model derived from lessons learned in the past, or from an internal procedure. The model could be a linguistic model or a rule model or a model based on learning theory or knowledge engineering. But it could also be very simple, a model in the form of a classification, or a thesaurus.

It is always a model of the knowledge domains we want to look at and the documents are – in real time – linked to the model and the changes we make to it.

Of course this modelling should be based upon a set of (Meta-) models that are available for the user. In this way we can integrate the existing knowledge structure, and by doing that we minimise the effort every time we make use of the infrastructure. In fact these models represent a basic asset of the organisations.

These models are also an important instrument to implement and manage the knowledge management policy I spoke about in chapter 3.1. I will come back to that later during my presentation. In line with this I think it is required to have facilities to store knowledge models so users are able to reuse them in new queries. These facilities should be simple to use.

It is important here to remark that the user is not stupid. We should make use of all the knowledge of the user related to the question and combine it with the knowledge the system has on IR, knowledge modelling and the document collection. It should combine the intelligence of the system with the intelligence of the user.

The dynamic aspect - I cannot repeat it enough - is particularly important in the knowledge management domain because of the dynamic environment of the modern organisation. This continuously changes the context in which documents are used and combined to new knowledge and applications of knowledge. It is important because one of the main issues of knowledge management is the interpretation of existing knowledge in new and changing environments.

4.4 Instrument to implement and maintain a knowledge management policy

The last but one aspect I want to discuss is directly related to the model I presented in chapter 3.1. As I said, the knowledge management policy is an essential aspect in the knowledge organisation.

Having such a policy requires instruments to implement it. A new statement related to a specific core competence for instance is only useful if such a statement would result in changes in the organisation.

Such a statement would result for instance in the linking of new information resources to the Intranet. It would also lead to new models to represent and link the existing information to. It must have an influence on all the relevant knowledge activities of the organisation.

I think it is interesting to think about the way you could link the knowledge management policy explicitly

to the facilities in the infrastructure and in such a way that you can manage the knowledge management efforts directly from that policy.

4.5 Easy to integrate: standards and hybrid

In this presentation I focused on the “stock model” of knowledge. Many people will say that this stock model is a fake. In its essence, that is true. You cannot keep knowledge in stock. Knowledge is a flow concept, by definition. Knowledge only exists in minds of people.

During the writing of a paper like this you are confronted with this ‘fake-aspect’ every time you write down the word ‘knowledge’, because many times you talk about simple data, instead.

But this conclusion does not make the stock concept of knowledge nonsense. In the contrary, it is a useful concept to help us solve the knowledge problem we have in a modern knowledge organisation. The stock concept is a useful thinking model.

But we should draw another conclusion from this too. The “stock knowledge infrastructure” should be highly integrated into the infrastructure of the real organisation, specifically the infrastructure related to ICT. That is an essential condition to make the results of this ‘fake environment’ available in the real world of knowledge workers. That is required to bring the thinking model on the stock concept of knowledge to reality. It is required to be able to create the added value for the organisation.

The knowledge infrastructure should be highly integrated into the world of communication, data storage and individual work. In fact this means adapted to universal standards for electronic communication, database technology, GroupWare, knowledge modelling, and other areas. This means that infrastructures should be bases upon XML/SGML, Internet technology standards, SQL, etc.

4.6 Conclusion

I am coming to an end.

My conclusion for information retrieval systems is that they:

- 1 Should be easy in use. At this moment they are too difficult for most users.
- 2 Apply intelligence to help the user to use all the advance retrieval functionality.
- 3 Help to interpret the information needs and its context into some kind of model from the perspective of the user.
- 4 Should analyse the documents in real time in the context of that user model.
- 5 Should help to modify the model based upon retrieval results.
- 6 Should be highly interactive
- 7 Should be an instrument to implement a knowledge management policy
- 8 Should be easy to integrate and hybrid.

5. Concluding remark

In my presentation I spoke mainly about IR-technology in relation to knowledge management. Certainly, IR-technology is an important aspect in a knowledge organisation. However I do not want to suggest that IR is the only important aspect. There are others that are at least as important or even more important.

I think that the key aspect is related to culture and habits of people. You can introduce as much technology and procedures as you want, it is of no use if people are not aware of the strategic value of knowledge for their organisation and of the activities they should perform to select, share and exchange knowledge.

I think however that good tools, like IR-systems, can make it easy to select, share and exchange information, and in that way, could have an enormous positive influence on the knowledge culture and habits of people. To my opinion that is the challenge for the coming years: to develop storage and retrieval systems that fit into the knowledge culture and habits of the knowledge worker.