

9-2010

INTRANET SEARCH PATTERNS IN A COMPLEX ORGANIZATION – THE HYBRID INFORMATION MODEL

Ravit Weisman

Bar-Ilan University, Israel, ravitws@bezeqint.net

Judit Bar-Ilan

Bar-Ilan University, Israel, barilaj@mail.biu.ac.il

Follow this and additional works at: <http://aisel.aisnet.org/mcis2010>

Recommended Citation

Weisman, Ravit and Bar-Ilan, Judit, "INTRANET SEARCH PATTERNS IN A COMPLEX ORGANIZATION – THE HYBRID INFORMATION MODEL" (2010). *MCIS 2010 Proceedings*. 88.

<http://aisel.aisnet.org/mcis2010/88>

This material is brought to you by the Mediterranean Conference on Information Systems (MCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in MCIS 2010 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

INTRANET SEARCH PATTERNS IN A COMPLEX ORGANIZATION – THE HYBRID INFORMATION MODEL

Ravit Weisman
Department of Information Science
Bar-Ilan University, Ramat Gan, 52900, Israel

Judit Bar-Ilan
Department of Information Science
Bar-Ilan University, Ramat Gan, 52900, Israel

Abstract

The Internet and the Web have become a major information and communication source. The technologies used on the Internet have found their ways to the Intranet of organizations as well. The organizational portal contains huge quantities of information, especially when the organization is large and complex and relies heavily on existing information, as in the case of a learning organization. It turns out that simply applying Web technologies “as is” for the enterprise is not the solution. Surprisingly, there are not many studies that tried to explore information behavior within organizations.

In this study our aim is to develop model for searching and retrieving information from the organizational portal within the intranet. As a case study we looked at the information searching patterns of users in a large and complex organization. Three focus groups were held with 13 users of the portal at the end of 2008. In addition in-depth interviews were conducted with eight knowledge workers in the organization.

The users participating in the study were shown two search result interfaces based on previous models in addition to the interface proposed by us, based on the results of the pilot study and on our “hybrid information model”. The users in the study preferred the “hybrid information” interface. Two focus groups with the expert users, the knowledge workers, dealt with finalizing the discussion regarding their interface preference. During these discussions we suggested an improvement to the suggested search result interface, which was endorsed by the expert users.

The study shows that within a complex organization it is crucial to employ an integrative solution for information search that satisfies the specific needs of the users of the organizational portal. The model developed for presenting search results has practical implications on the effectiveness, efficiency and relevance of searches conducted within the portal. Even though the case study was conducted within a specific organization, we believe that the models can be applied in other organizations as well.

Keywords: Intranet, portal, search interface, search results interface.

1 INTRODUCTION

The Internet and the Web have become a major information and communication source. One of the most popular activities on the Web is searching (see for example Fallows, 2008). Search has become almost synonymous with Google (StatCounter, 2010; McGee, 2010). One of the best-known features of Google is its clean and simple search interface, with (almost) only a search box on its homepage where people can enter their simple searches. This interface is often mentioned as one of the reasons for Google's high popularity.

The technologies used on the Internet have found their ways to the Intranet of organizations as well. The organizational portal contains huge quantities of information, especially when the organization is large and complex and relies heavily on existing information, as in the case of a learning organization. Li et al. (2005) claim that enterprises have tens or hundreds times larger data collections than the information available on the Internet. Many organizations provide access to their employees to internal information and communication through a portal, since it provides a single gateway to all information and knowledge resources in an enterprise (Dias, 2001; Anonymous 2005). Enterprise portals are more than simple information dissemination tools, they need to provide access to content and allow for communication and coordination (Detlor, 2000; 2004). Choo, Detlor and Turnball (2000) discuss intranets and the design of intranets and organizational portals that support knowledge work.

It turns out that simply applying Web technologies "as is" for the enterprise is not the solution (see for example Broder & Ciccolo, 2004; Hawking, 2004; Mukherjee & Mao, 2004). As Mukherjee and Mao (2004, p. 37) state in their paper entitled "Enterprise search is tough stuff": "Although these employees seek Web-like experiences in the enterprise, the Internet and enterprise domains differ fundamentally in the nature of the content, user behavior, and economic motivations".

In this study our aim is to develop model for searching and retrieving information from the organizational portal within the intranet. As a case study we looked at the information searching patterns of users in a large and complex organization. Knowledge workers were also consulted in the process.

To much of our surprise, there are not many studies that tried to explore information behavior within organizations. Wilson (2000) defines *information behavior* as "the totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use". Additionally, he defines *information seeking* and *information searching*, where information seeking is "the purposive seeking for information as a consequence of a need to satisfy some goal" and information searching is the 'micro-level' of behavior employed by the searcher in interacting with information systems of all kinds".

The lack of research on information behaviour and search was noted by Stenmark (2005, p.2): "In other words, since intranets are well established and widely used technology that has received little academic interest, there is a need for more in-depth studies of intranet searching". In addition to the lack of research on user behavior in intranets, there are almost no existing models for search and search results interfaces that are based on the information behaviour of users.

The rest of the paper is organized as follows: in the next section we discuss information searching models within organizations, then present the study methodology, followed by results and discussion in section 4, and finally we present some conclusions.

2 INFORMATION SEARCHING MODELS WITHIN ORGANIZATIONS

In this section we considered several models of information search within organizations, and here we describe the three models that focus on the presentation of search results and have relevance to our study. Our basic assumption, based on a pilot study within the organization was that users are not interested in complex search forms, and prefer the basic search interface comprised of a single search box, and thus we concentrated mainly on models that discussed effective and efficient ways to present search results.

2.1 The “information extraction” model

This model was proposed by Mukherjee and Mao (2004). They claim that one of the advantages of the intranet as compared to the Internet is that in the intranet taxonomy, navigation, classification, and categorization can be utilized. In addition, metadata available for semi-structured documents also have great value for content search. They provide a list seven characteristics of information and information retrieval within the enterprise:

- Diversity of content sources and formats
- Secure access
- Combining structured and unstructured search
- Flexible scoring and ranking mechanisms
- Federated and peered results (single point of access)
- Content generation processes
- People/roles/behaviors – utilizing the knowledge and experience of the employees

Mukherjee and Mao (2004) explain the difference between the notion of a “good” search on the Internet and within the organization. On the Internet “[b]ecause a large number of documents are typically relevant to a query, a user is often looking for the ‘best’ or most relevant document. On an intranet, the notion of a ‘good’ answer is often defined as the ‘right’ answer. Users might know or have previously seen the specific document(s) that they are looking for” (Mukherjee & Mao, 2004 p. 38).

2.2 The “information desk” model

This model was proposed by Li et al. (2005) and focuses on conducting searches and displaying results of these searches in the intranet, or more precisely within the organizational portal. They assume that that information needs can be categorized into searches for information of different types (e.g. definitions, experts, homepages). For each information type they employ extraction technologies in advance, and they combine the results for the different types of information together and present it in a unified system. The research was conducted within Microsoft and was comprised of four stages. The first step in developing the model was: understanding search needs. The analysis was based on search logs and was supplemented by a small user survey. The top level categories are identical to the categories proposed by Broder (2002), and the second level resembles the second level categories proposed by Rose and Levinson (2004), with one important new category: “tell me about”, which was identified by the authors as relevance search. The categories are as follows:

- Informational
 - When (time)
 - Where (place)
 - Why (reason)
 - What is (definition)

- Who knows about (expert)
- Who is (person)
- How to? (manual)
- Tell me about (relevance)
- Navigational
 - Group
 - Person
 - Product
 - Technology
 - Services
- Transactional

As a result of the search need analysis, Li et al. (2005) set up an “information desk” that provides search results for four categories for which are supported by information extraction methods:

- What is (definition)
- Who is (person)
- Where is home page of (group, product or technology)
- Who knows about (experts)

When submitting a query, the user checks the information type(s) he/she is looking for. Another option is to introduce facets when creating the query, see for example Hearst (2006) or Ben-Yitzhak et al., (2008).

2.3 Results in context

Dumais, Cutrell and Chen (2001) developed and evaluated different search result interfaces that integrate semantic information with the search results. The interfaces were evaluated for web search, but the results are highly relevant for searches conducted with the intranet as well. They compared interfaces where the results were categorized into hierarchical categories with interfaces where the results appear in a simple list. Four category and three list interfaces were evaluated:

- Category hover – summaries available as hover text
- Category inline – summaries are inline
- Category with no category names
- Category with no page titles (browse)
- List hover – no categories, summaries available as hover text
- List inline – no categories, summaries available inline
- List plus category names

The effectiveness of the interfaces was evaluated by users who had to find answers to queries using a specific interface. In all cases the category interfaces were faster than the list interfaces. In addition the users were asked to fill in a satisfaction survey. The user almost unanimously preferred the category interfaces over the list interfaces. Their findings are quite interesting if we take into account the extreme popularity of the list interface of Google.

3 METHODOLOGY

In this research, our aim was to study what types of information employees search for and retrieve from the organization portal for their work tasks. The second, more practical aim was to provide a model for presenting the search results that facilitates the execution of user tasks.

We chose to study a large, complex (Bar-Yam, 2005) and learning organization (Druker, 1964, Senge, 1990) from the Israeli public-government sector. The organization is comprised of several subunits and has thousands of employees. The employees access information through the organizational portal,

which is comprised of subportals and searchable through a Microsoft based search engine. The internal databases are searched through the search engine provided by these databases. The organization has a well-developed and extended taxonomy. The organization taxonomy has six levels with 16 subunits at the top level and the subject taxonomy has two levels, with 12 topics on the top level and about 10 topics each at the second level, altogether more than 120 subject categories.

Li et al. (2005) tried to deduce the employees' information needs from search logs. We preferred a user study where employees were interviewed or participated in focus groups. These methods allowed them to freely express their needs and concerns. These qualitative methods (Creswell, 2003) allowed the users to convey their opinions on the suggested search results interfaces as well.

The user study was conducted at the end of 2008 and included 13 employees and 15 knowledge workers (knowledge managers and librarians). The ten knowledge managers came from seven different organizational units, nine of them mid-range and one in a managerial position. The five librarians were the directors of the five largest libraries within the organization. The "regular" workers worked at three different organizational units and used the organizational portal either everyday or at least several times a week.

The knowledge managers were asked about their information needs both as content providers and as users. In depth interviews were conducted with eight knowledge workers. During the pilot phase attempts were made to conduct one-on-one interviews with the "regular" employees as well, but it turned out that the regular employees had difficulties in expressing their needs and concerns, thus we decided to hold focus group meeting with the regular employees instead, where the group dynamics motivated their participation. Three focus group meetings were held with the regular users, one in each of the three organizational units chosen. In these meetings the users were asked about their information needs and uses.

In addition two focus group meetings took place with the experts – the knowledge workers – to present them with the different, suggested search result interfaces and to discuss the users' reactions to the suggestions. Most of the participants were also observed while trying to retrieve information from the portal. The observations allowed us to gain insights on the users' search experiences and the relations between the searches and their tasks (Detlor, 2001) and supplemented the data gathered through the interviews and the focus groups.

After conducting a small pilot and based on previous research a preliminary model of information behaviour and a model for search result display were developed. In addition to the interface based our model, the *hybrid information model*, two additional interfaces were presented to the users and knowledge workers and their opinions were requested.

In the following we describe the three models for search result display. The first model is the well-known Google model, where search results are presented in a simple list (see Figure 1).

The second model is the focused search result display, suggested by Levy (2006). The model is similar to the categories model with inline summaries evaluated by Dumais et al. (2001) as described in section 2.3. This model proposes that when the user submits a query to the simple search interface, he/she receives categorized results. The categories proposed by Levy are somewhat different from the categories discussed by Dumais et al. The user will receive between four to eight "windows", each one corresponding to a category. These windows will be fixed for all the users in the organization and will be based on the characteristics of the central themes relevant for the organization. In Figure 2, we see four such windows: results from the Internet, organizational documents, contact persons and dictionaries.

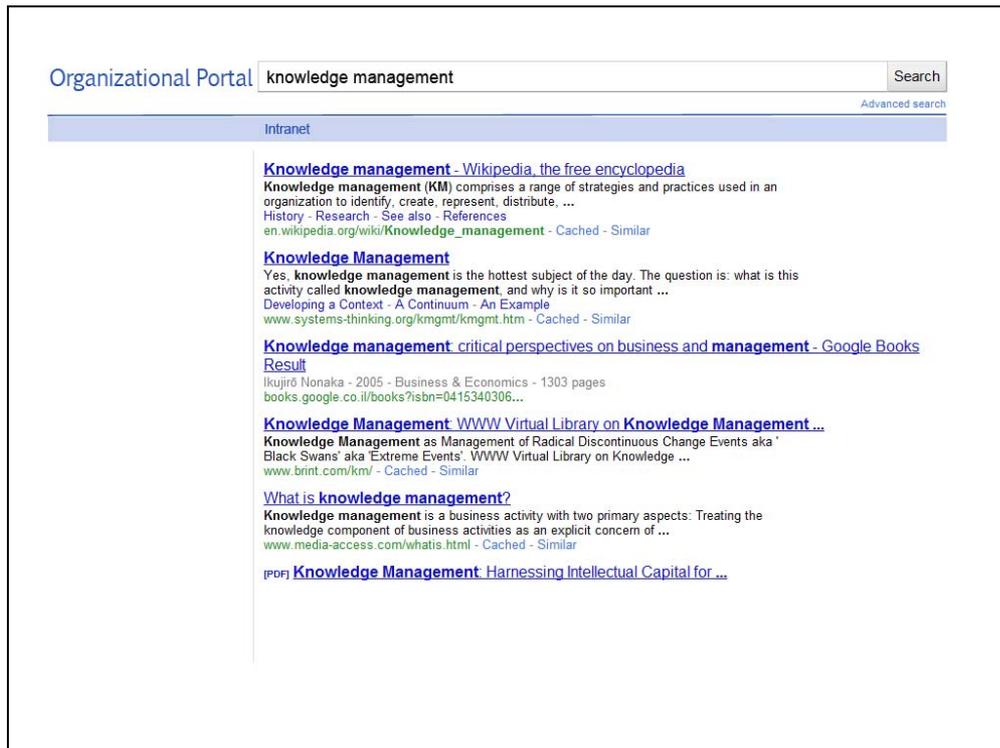


Figure 1: The Google-style interface

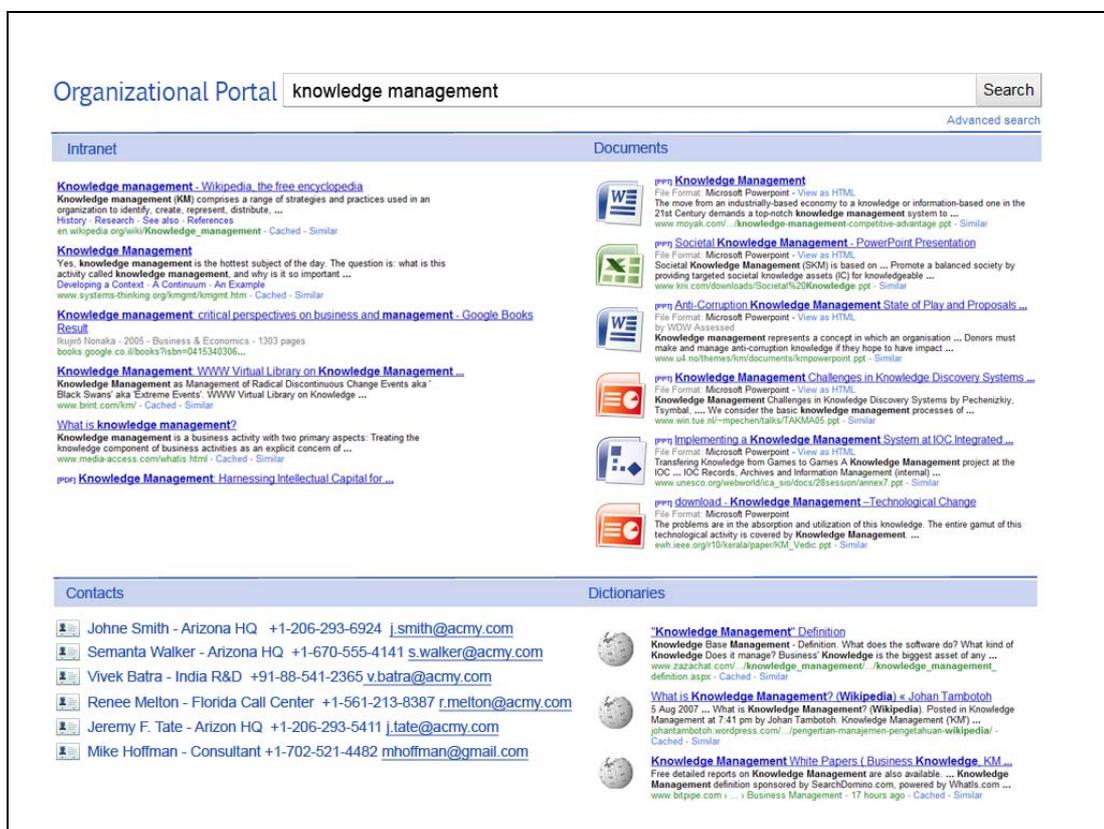


Figure 2: Focused display

The third model, the one proposed by us, is based on and integrates a number of previous models and on the findings of the pilot. This model is called the *hybrid information model*. In Figure 3 the basic version is displayed, as a result of the study an additional, improved version was developed, as can be seen in Figure 4. The rationale for the improvements is explained later. Here are the major points guiding us in proposing the basic version:

- The model should be based on the information behavior of the users (Li et al., 2005; Choo, Detlor and Turnball, 2000), see section 2.2.
- The search interface should be a simple search interface, similar to the Google search interface – users rarely use advanced search interfaces (Levy, 2006), advanced search features (Spink & Jansen, 2004), and visit most of the time only the first and perhaps the second search results pages (Spink & Jansen, 2004).
- The search results interface, on the other hand, should be more elaborated than the google-like interface. The interface should enable:
 - Filtering the results, and should include both a list component and a categorized component of the type suggested by Dumais et al. (2001), see section 2.3. The users are accustomed to the list component from the Web (Anonymous, 2005).
 - The categories should be based on the characteristics and the rules of the organization, which are enforced when new content is added to the portal (taxonomies, metadata). The categories should allow navigation by taxonomy (classification) and filtering by metadata which provides better results than google-like interface (Mukherjee & Mao, 2004), see section 2.1. Thus the first group of categories will include the organizational taxonomy, the second group allows to filter information by metadata (Content providers in the organization are required to provide metadata when uploading information to the portal e.g. title, document type), and the third group will point to virtual information – knowledge communities, experts and contact persons.
 - The scrolling list in the central part of the display contains summaries.
 - Selecting one of the categories changes the central part: a list of items belonging to the chosen category will be displayed in the center.
 - Only the top-level categories are displayed initially, but by clicking on the + sign in Figure 3, a further level of the relevant hierarchy appears.

The model is displayed in Figure 3.

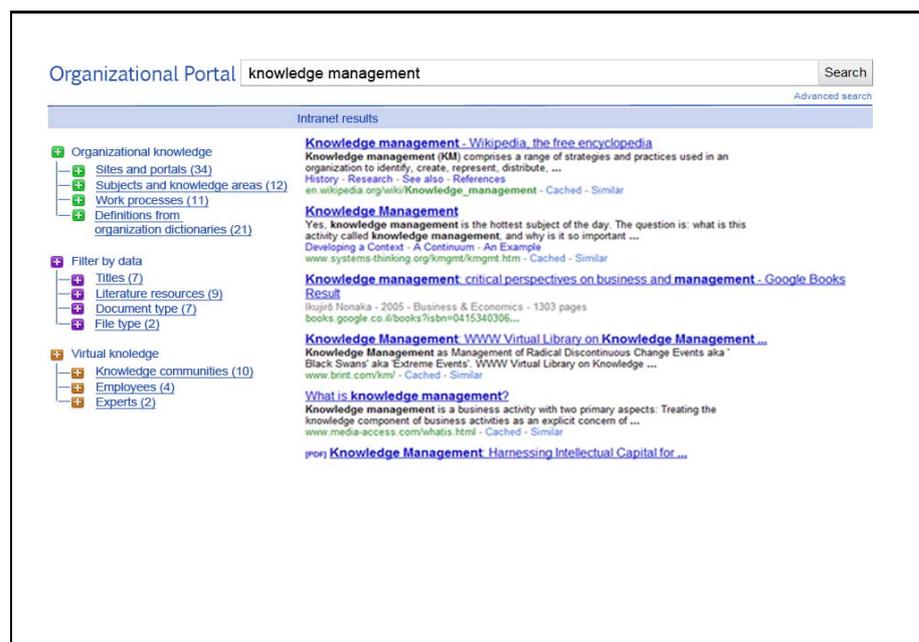


Figure 3: The initial hybrid information model – search results interface

The analysis of the data followed the grounded-theory method (Strauss & Corbin, 1990): the units of analysis were defined, emerging themes (categories) were identified and the units ascribed to the themes. At first only top-level categories were defined later a second level was added in order to refine the categorization. The initial categorization of the users' information behavior was informed by Choo et al.'s conceptual framework (2000), and the top-level categories were defined as "information needs", "information seeking" and "information use". For categorizing the users' perceptions regarding the search interface, the top-level categories were the four principles of the "hybrid information model": integrating different presentation interfaces (categorical & scrolling); types of information to appear in the categorical presentation; categorical presentation – with or without snippets and presentation display format.

4 RESULTS AND DISCUSSION

The interviews, the focus group discussions and the observations were analyzed based on Choo et al.'s (2000) integrated model of human information seeking, according to which the three main processes involved are information needs, information seeking and information use.

In terms of information needs, the user search for different types of information in the portal, both for their personal and professional needs. The major types of information within the organization are: working papers, forms, rules and regulations, lessons learned, professional sources, news and summary info that appear in the portal, info on experts, contact info and information from knowledge communities, information extracted from databases and metadata (e.g. the expiry date of some documents). For the regular employees, generally the portal is the primary source, except for information on people and communities. However the librarians often rely on additional sources, included printed material as well. There are huge amounts of information in the organization and the users feel overloaded. Their requests from the information system are as follows:

- Accessibility
- Relevance
- Filtering of results
- Personalization
- Effectiveness
- Displaying semantic relations between terms
- Improved interface design
- Single point of access for all types of information
- Updated info

The major finding **in terms of information seeking** is that the preferred method of information seeking is through the use of simple search – entering the search terms into a search box ("google-like" search), supporting the assumption in our hybrid information model. Users do not use the existing "advanced search" interface, some of them were not even aware of its existence. In addition with the exception of the librarians they do not use any advanced search features (e.g. Boolean search) either. There are two additional information seeking methods: taxonomy navigation and filtering by metadata (without entering free text search terms). The users perceive taxonomy navigation as complex and this method is rarely utilized. There is some use of the metadata filtering method. Although searching is the preferred method by far, the users have problems with choosing the appropriate query terms.

The major problems surfaced when discussing the search results interface. The current approach, the Google-like scrolling list overwhelms the users, because of the huge amounts of results. They rarely look beyond the first page of results, and almost never beyond the third page. They work under time constraints, and thus often abandon their searches, become frustrated and try other ways of acquiring the needed information or try to manage without it. Some of the workarounds we saw during the observations included entering metadata information into the simple search box and the use of

favorites to quickly reach the relevant sub-site in the portal. Because of the lack of integrated and structured information in the system on experts and communities, the portal was rarely used locating such “virtual” information. Thus the portal misses one of its goals: to enhance information sharing within the organization.

These findings further support the needs expressed by the users for accessible, relevant and timely information, the need for effective filtering and personalization.

In terms of information use we found that the major uses of information are:

- Supporting existing and known information
- Refuting existing and known information
- Use as background material
- Use as a starting point for additional searches
- Filing for office use

The portal is used to retrieve information both for personal and for work-oriented needs. Work-oriented uses were for office use: forms, rules and regulations and for fulfilling specific tasks: preparing documents or professional update.

All the users were shown the three search interfaces (Figures 1 to 3). All of them preferred our approach, although there were comments on the exact categories and their ordering. It was suggested to allow personalization in selecting the displayed categories. Further suggestions included adding semantic relations maps, allowing the user to choose between the Google-like and the hybrid information interface, adding additional sources and providing relevance weights of the results.

During the discussion with the expert groups we raised the idea of pushing additional useful information related to the search in a designated area. This idea was accepted by the expert groups, and it was decided to propose an interface with pushed information in five areas: forms, procedures, lessons learned, definitions & terms and literature. The pushed information is provided with short summaries to allow the users to quickly assess its usefulness (see Figure 4).

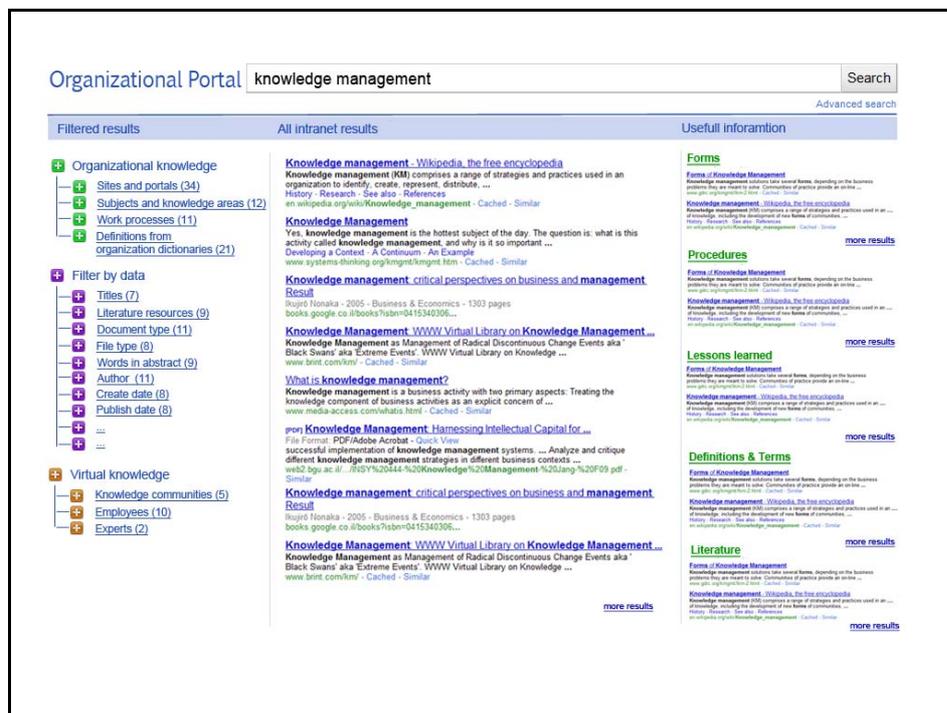


Figure 4: The hybrid information model – search results interface

Figure 5 summarizes our hybrid information model, and depicts the information retrieval process of the users. This model was inspired by the interviews and the observations. Users can abort the search process either because they are satisfied and their information needs are fulfilled or they give up, because of lack of motivation, frustration of time constraints. Our hope is that when using the suggested search results interface of the hybrid information model, the user satisfaction and search experience will be greatly improved.

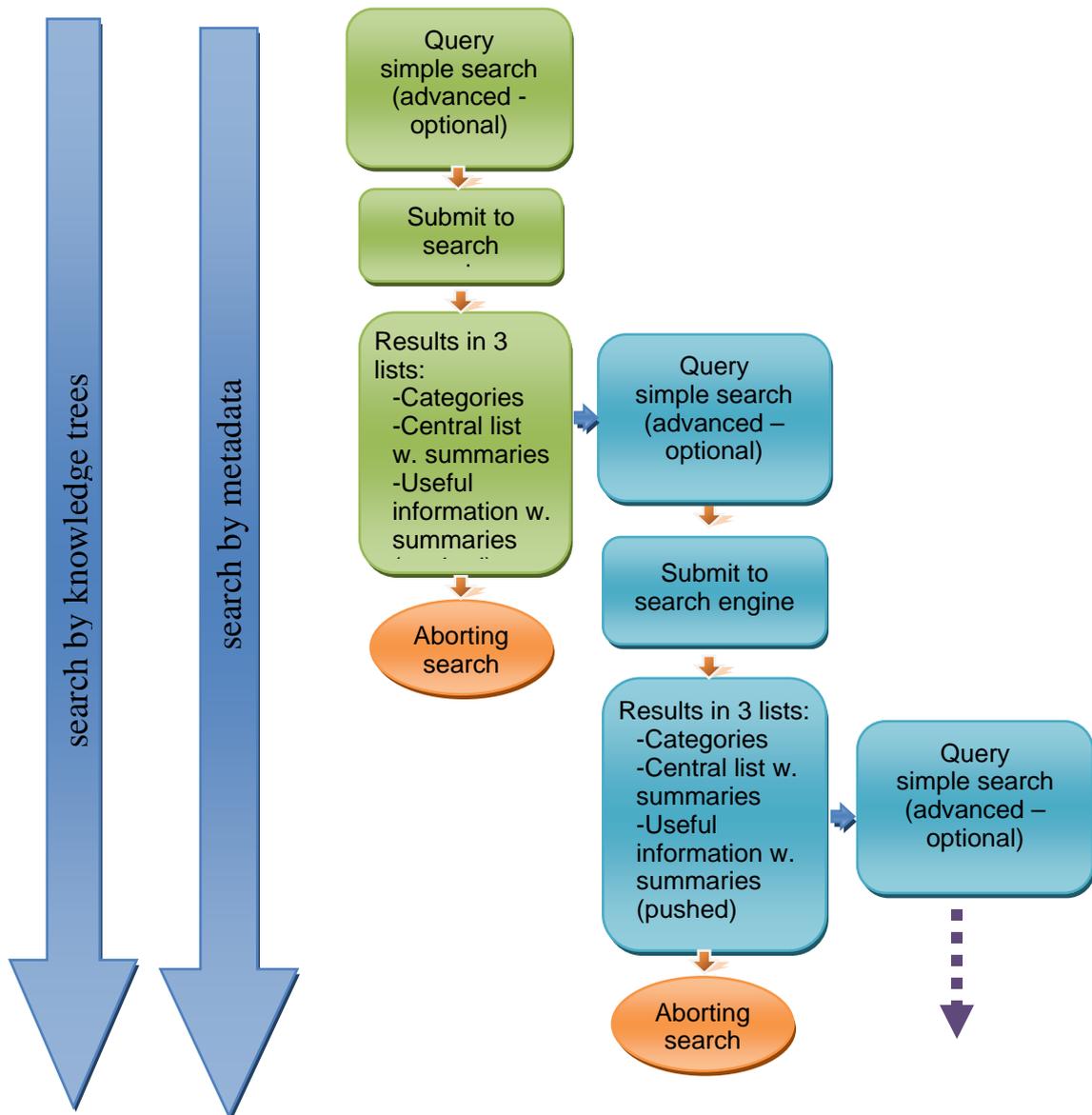


Figure 5: The hybrid information model – process view

5 CONCLUSIONS

Studying information needs, information seeking and information use within the organizational portal is of utmost importance, because organizations choose this technology to provide information to their employees. It has already been shown that technologies applicable for the Internet are not directly applicable to the intranet. However, not many user studies have been conducted to this day in order to gain insights on information behavior within the intranet.

The study touched upon additional issues that we were unable to discuss in detail in this paper due to space limitations: the similarities and differences between searching in the intranet and on the Web; characteristics of complex, learning organizations and their relation to information search and the influence of search interfaces on the information behaviour of the users of the system.

Our case study shed light to the challenges for locating information through the organizational portal. The information seeking process was analyzed, and modelled by our hybrid information model. It was shown that users do not want elaborate search interfaces, but they welcome improved search results interfaces that are better suited for the organization. We provided a model for the search results interface that was welcomed by the users and experts in the study. When our suggested model will be adopted and implemented, we will be able to evaluate its effectiveness and the users' satisfaction with the new interface. Although the study was confined to a single organization, we believe that at least some of its conclusions are applicable to other organizations and enterprises as well.

References

- Anonymous (2005). Web portals growing once again. *International Journal of Productivity and Performance Management*, 54(1), 61-65.
- Bar-Yam, Y. (2005). *Making things work: Solving complex problems in a complex world* (1 ed.). n.d. , U.S.A.: Knowledge Press.
- Ben-Yitzhak, O. et al. (2008). Beyond basic faceted search. In *Proceedings of WSDM 2008*, Palo Alto, CA, pp. 33-43.
- Broder, A. (2002). A Taxonomy of Web search. *SIGIR Forum* 36(2).
- Broder, A. Z. And Ciccolo, A. C. (2004). Towards the next generation of enterprise technology. *IBM Systems Journal*, 43(3), 451-454.
- Choo, W. C., Detlor, B. And Turnbull, D. (2000). *Web work. Information seeking and knowledge work on the World Wide Web*. Dordrecht, Netherlands: Kluwer.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative and mixed methods approaches* (2nd ed.) Thousand Oaks, Ca: Sage.
- Detlor, B. (2000). The corporate portal as information infrastructure: towards a framework for portal design. *International Journal of Information Management*, 20(2), 91-101.
- Detlor, B. (2001). The influence of information ecology on e-commerce initiatives. *Internet Research*, 11(4), 286-295.
- Detlor, B. (2004). *Towards knowledge portals: From human issues to intelligent agents*. Dordrecht, The Netherlands: Kluwer Academic Publishers
- Dias, C. (2001). Corporate portals: a literature review of a new concept in Information Management. *International Journal of information Management*, 21, 269-287.
- Drucker, P. (1964). *Managing the results*. New York, NY: Harper & Row Publishers.
- Dumais, S. T., Cutrell, E. and Chen, H. (2001). Optimizing search by showing results in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Seattle, Washington, pp. 277-284.
- Fallows, D. (2008). Almost half of all internet users now use search engines on a typical day. *PEW Internet and American Life Project*. Retrieved May, 13, 2010, from http://www.pewinternet.org/~media/Files/Reports/2008/PIP_Search_Aug08.pdf.pdf
- Hawking, D. (2004). Challenges in enterprise search. In *Proceedings of the 15th Australasian Database Conference*, pp.15-24.

- Hearst, M. (2006). Design recommendations for hierarchical faceted interfaces. In ACM SIGIR Workshop on Faceted Search. Retrieved May 14, 2010, from <http://flamenco.berkeley.edu/papers/faceted-workshop06.pdf>
- Levy, M. (2006). Focused search – the new-old search (in Hebrew). Retrieved May 14, 2010, from <http://www.kmrom.com/Site/Articles/ViewArticle.aspx?ArticleID=771&P=59>
- Li, H., Cao, Y., Xu, J., Hu, Y., Li S. and Meyerzon, D. (2005). A new approach to intranet search based on information extraction. In *Proceedings of the 14th ACM International Conference on Knowledge Management*, pp.460-368.
- McGee, M. (2010). Google regains search share in April. *Search Engine Land*, Retrieved May, 13, 2010, from <http://searchengineland.com/google-regains-search-share-in-april-hitwise-says-41367>
- Mukherjee, R. and Rao, J. (2004). Enterprise search is tough stuff. *ACM Queue*, 2(2), 37-46.
- Rose, D. E. and Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th international World Wide Web Conference on Alternate Track Papers & Posters*, 2004 New York, USA.
- Senge, P. (1990). *The fifth discipline: The art and practice of learning organization*. New York: Doubleday.
- Spink, A. and Jansen, B. J. (2004). *Web search: Public searching on the Web*. Dordrecht, Netherlands: Kluwer.
- StatCounter (2010). *Top 5 search engines from Apr 09 to May 10*. Retrieved May, 13, 2010, from http://gs.statcounter.com/#search_engine-ww-monthly-200904-201005
- Stenmark, D. (2005). One week with a corporate search engine: A time-based analysis of intranet information seeking. In *Proceedings of the Eleventh Americas Conference on Information Systems*, Omaha, Ne, pp. 1-12.
- Strauss, A. and Corbin, J. (2000). *Basics of qualitative research: Grounded theory procedures and techniques*. New Bury Park, CA: Sage.
- Wilson, T. (2000). Human information behavior. *Informing Science*, 3(2).