

Information Seeking is Social

Kaisa Soininen

Sonera Zed

Posbus 8557

3009 AN Rotterdam

The Netherlands

+358 40 7572115

kaisa.soininen@sonera.com

Eija Suikola

Teamware Group

P.O. Box 135

FIN-00381 Helsinki

Finland

+358 9 5128 2660

eija.suikola@teamware.com

ABSTRACT

This study aimed to clarify how social knowledge is used in information seeking in the Internet. The issue was studied with tasks without one correct answer. The verbal protocols were categorised according to the subprocesses of information seeking and indication of use of social knowledge. We found clear evidence that social knowledge was used when selecting a source or place where to start looking for the information, while examining the results and extracting information, and when reflecting and making the decision to iterate or stop seeking. The results show that people interpret the interaction with web pages as social situations, use their social knowledge, and actively seek for social information.

Keywords

Information seeking, Internet, social knowledge, social navigation

1. INTRODUCTION

In today's information society dealing with information has become a primary activity of life for more and more of us. Information seeking is precursory to other information management activities, like generation of information, and thus understanding information seeking behaviour enables supporting information management at large.

Although the Internet has evolved into a major information source, it is widely acknowledged that information seeking is very difficult, in general and in

particular in the Internet. There have been several attempts in understanding the nature of information seeking better by creating process models and phasing the process (e.g. Saito and Ohmura 1998, Golovchinsky and Belkin 1999, Kuhlthau 1996, Marchionini 1995). These models can also be used to improve designers' understanding of user behaviour and further to design systems that would diminish the difficulties in information seeking.

Most of the information seeking research so far has concentrated on behaviour and cognition in information seeking. Only recently have affective and social aspects

begun to draw researchers' attention (e.g. Munro et al. 1999a). However, social aspects are in evidence also in information seeking even if individuals are not in direct interaction with each other. The members of communities of expertise (cf. Wenger 1998) can use social knowledge of their field to guide their information seeking behaviour. Scientific communities are a good example.

We believe that by combining the two approaches (process models and focus on social aspects in information seeking) research can produce valuable information about information seekers' behaviour, also giving ideas for design enhancements for information providers.

1.1 Information seeking process

Information seeking has been defined as *"a process in which humans purposefully engage in order to change their state of knowledge"* (Marchionini 1995, 5). There have been several attempts to describe what are the stages or subprocesses of the information seeking process.

The process can be divided into two major tasks - defining the target and locating information that meets the definition (Saito and Ohmura 1998). There are advantages of defining the target before gathering new information, since it increases the information seeker's satisfaction with the results (Saito and Ohmura 1998). Kuhlthau (1996) also recognised the stage of clarifying the focus of the information seeking as a critical phase for successful completion of the activity. In addition, social aspects are implied to be related mainly to this stage in her model.

Kuhlthau (1996) studied library users engaged in information seeking (as part of a project) from constructivist viewpoint and revealed more detailed structure of information seeking process than Saito and Ohmura (1998). The model has seven distinct stages: task initiation, topic selection, prefocus exploration, focus formulation, information collection, search closure, and start of writing. Affective, cognitive, and behavioural changes occur while a person moves from one stage to another. Topic selection, prefocus

exploration, and focus formulation are mostly dealing with target definition, although some information is also located in these stages. Thus, the two major tasks, target definition and information location, seem to be intertwined.

Marchionini (1995) has proposed a model for the information seeking process, which proposes several subprocesses that may proceed in parallel and nearly any of them can be the most prominent at any time. So, a sequential view of the information seeking process is abandoned in order to describe the main structures of the process. The model is based on studies of people using electronic information retrieval systems and consists of eight subprocesses: recognising and accepting an information problem, defining the problem, selecting a source, formulating a query, executing search, examining results, extracting information, and reflecting and making a decision to iterate or stop. Like Kuhlthau (1996) Marchionini (1995) also mentions some social aspects in the subprocesses related to target definition (define the problem and reflect, iterate/stop), but also in selecting source, formulating query and examining the results. However, neither of the researchers has examined the use of social knowledge further.

Both Marchionini's and Kuhlthau's models include the division of defining the target and seeking for information that matches with the target, although acknowledging that the tasks are actually composed of subtasks. Although there are differences in these models, they are not mutually exclusive because of the different timeframes: Kuhlthau's model depicts longer scale *information seeking project*, while Marchionini's model is more relevant to *individual instances of information seeking*. So, at the stages of topic selection, prefocus exploration, and information collection information seekers actually perform several instances of information seeking as described by Marchionini.

The target definition has been acknowledged as a very important stage in the process (Saito and Ohmura 1998, Kuhlthau 1996) and social aspects have been linked to it (Marchionini 1995, Kuhlthau 1996). However, there has not been any research on how social knowledge is used in the different stages of information seeking process.

Individual instances of information seeking were selected to be investigated in this study rather than an information seeking project, because it is easier to control different variables in short term tasks, which enables understanding more clearly the dependencies between variables of interest. Thus, Marchionini's model was used as the main theoretical tool.

1.2 Constructing comprehension using social knowledge

The cognitive processes involved in information seeking behaviour are constructive in nature, that is, knowledge and understanding are built based on available information as the information seeking proceeds. Information seeking behaviour requires construction of cognitive models of sought information and plans to execute searches. Potential information sources include task descriptions, information extracted from search results and retrieved documents.

Information seeker's current knowledge and skills are also used. In particular, domain knowledge has been found to be an important factor affecting effectiveness of information seeking (Marchionini 1995, Hoelscher and Strube 1999). Typically the focus has been on the content matter of domain knowledge ignoring the social knowledge so closely related to expertise in a particular area. Social knowledge related to domain expertise includes, for instance, information about the players in the field are, their activities and reputation. In fact, the domain knowledge as a whole is a result of a social process, i.e. of collaborative construction of knowledge.

The emerging field of social navigation (Munro et al. 1999a) offers a new way to look at information use. Understood widely, "*[s]ocial navigation considers the creation of social settings and "places" in information space and behaviour in them, the sociality of information creation, people as members of groups and nature of information itself, its location, evaluation and use*" (Munro et al. 1999b, 2-3). With this perspective all the phases of information seeking in the Internet can be viewed as social situations, even if users do not interact in real time.

Harper (1999) gives an example of the importance of the use of social knowledge in information seeking. He observed how IMF specialists used their social knowledge successfully to construct an understanding about a country's financial situation. IMF specialist dismissed what might have seen to be logically valuable source of information, because they possessed knowledge of the social and power relations in the country. Instead of information provided by an institute publishing macroeconomic reports about the country, they valued higher the official paper of the party in power. In order to understand the decision making of the people in power this paper was much more important than the reports by the institute, which did not support official politics.

The previous research on the information seeking process has mainly concentrated on the behavioural and cognitive changes that occur during the process. In addition to this, the use of available information has been studied largely as related only to the results of the process, e.g. will persons with domain knowledge conduct more effective information seeking than novices. This study combines the two aspects of information seeking - the process and the information used to support it - to reveal at a more detailed level how social knowledge is used in the information seeking process.

2. METHODS

2.1 Participants

Seven participants representing a variety of professions (product development, marketing, visual design, usability and informatics) performed tasks of searching information from the Internet. Four of the participants were women, three were men. Participant's ages varied between 28 and 47. All had long experience in computer use. Participants' background information is summarised in Table 1.

Table 1. Participant's experience with computer use and their knowledge in the domains related to areas used in the search tasks. Knowledge was measured by participants' own judgement with a scale from 1 (poor) to 5 (excellent). D / wk = computer use in days per

week, h / d = computer use hours per day, I-tech = Internet technology.

	Years	d / wk	h / d	I-tech	Usability
P1	15	5	> 4	3	4
P2	10	6-7	> 4	3	3
P3	7	5	2 - 4	1	2
P4	9	6-7	_ - 2	1	3
P5	10	6-7	> 4	4	3
P6	15	6-7	> 4	5	1
P7	8	6-7	> 4	4	4

All participants were familiar with Internet searches and most of them did searches every day they used computers. All participants had created web pages and were familiar with html, but two of them only superficially.

3.2 Tasks

Participants performed five tasks, the first of which was for practising thinking aloud and was not included in the analyses. All tasks were open-ended, without one single correct answer. The tasks were:

- Find a html guide that you would like to use when making web pages.
- What advantages and disadvantages do CSS (cascading stylesheets) have?
- Find a style guide that you would like to use when making web pages.
- What sort of things affects usability of web pages?

3.3 Procedure

At the beginning participants filled in a questionnaire about their background skills and knowledge. The test leader explained the procedure and asked participants to think aloud while completing the tasks. The tasks were performed in a usability lab with a PC. Each participant

performed the tasks in a different order. The participants were invited to the lab one by one.

Participants' performance and thinking aloud was video and audio taped, which allowed the results to be analysed using detailed transcribed verbal protocols. Participants were encouraged to use their preferred search engine, link index or bookmarks that they were accustomed to use. The majority of them used Altavista, some others used also Yahoo, Google or Ihmemaa (a Finnish search engine). In most cases, a participant used only one search facility. However, three of them used more than one, especially if the first few search attempts were not producing desired results.

After completing each task they were asked to estimate how difficult the search had been for them and why. After completing all the tasks participants were briefly interviewed about their habits of doing Internet searches.

3.4 Analysis

Transcripts containing word to word protocols of participants' thinking aloud, records of actions taken by them and running time stamps were produced. Verbal protocols were analysed with a content analysis.

Verbal protocols were divided into propositions, which were then categorised based on the following information:

- in which of Marchionini's subprocesses the propositions fell
- whether the propositions contained indications of use of social knowledge.

Social knowledge was further divided into sub-categories containing references to the participant himself or herself, individuals, and groups. Unlike the categories of the phases of information seeking process, the categorisation of social knowledge was derived from the data.

Two independent evaluators categorised the content of one participant's protocol in order to estimate an inter-rater reliability. The reliability for search process phase categorisation was 66% and for use of social knowledge 91%. The relatively low reliability for search process phases is explained with disagreements about

proposition at phase transition points. Phases 6 and 8 seemed to be the most difficult to interpret.

In addition to content analysis, the frequencies of the propositions in the categories and their proportions were counted for quantitative analysis.

3. RESULTS

There were not much verbalisation for the subprocess of recognising the information need (34 or 1.0% of all the propositions). These propositions were mainly concerned with commenting about the task description, e.g. Participant 5 said “okey so let’s seek for an html-guide”. More examples of the propositions reflecting information seeking subprocesses are listed in the table 2 and those reflecting use of social knowledge in the table 3.

Slightly more propositions fell into categories of defining the problem, selecting the source, and formulating the query, 226 (7.0%), 279 (8.6%), and 197 (6.1%) respectively. When the participants defined the problem, they mostly reflected on the task descriptions and tried to gain better understanding of them. This can be illustrated with a quote from Participant 7 “guidelines guidelines they can be whatever / but when there is design then it might indicate that you know you can find also something like appearance stuff”. An example of selecting source category was selecting to search for material in certain language like Participant 6 did “let’s try to find it in plain Finnish / well this sort of search inet.fi search engine”. The propositions that were categorised as subprocess formulating the query included verbalisations of trying to find appropriate keywords and a way to combine them – “so then I want very basic / if I type in beginners plus css in altavista / and any language” (Participant 2).

Executing the query, mainly a behavioural subprocess, was verbalised quite rarely: only 82 (2.5%) propositions were categorised as such. Verbalisation seemed to occur mainly when there were some problems in executing the query, for instance in the case of Participant 5 “err / now I don’t remember whether it is spelled with g”.

Table 2. Examples of propositions reflecting different information seeking subprocesses.

Subprocess	P	Example
1. Recognise, accept	P4	these are just you know your own stuff / why can’t I search for you know basketball stuff / usability
2. Define problem	P6	what I would like to use / well then I’d like to have a little more / because there are probably loads of basic html-guides / but I want more specific information
3. Select source	P3	hey these virtual libraries / there might also be in here well / some material in electronic format / or are these just started
4. Formulate query	P1	or then I could go and try to find separately / cascading style sheets limitations / let’s say / limitations of
5. Execute query	P7	or you know you can use plusses and like in altavista or there / I don’t know how they work in this search engine
6. Examine results	P4	style guide of yale university design / I think I’d check that one out / here seems to be some style guides / I’d check that / let’s view that
7. Extract info	P3	Images / how to add in which format / and then there is well / let’s see what’s in here
8. Reflect, stop	P5	okey here’s one / but you know of course I wouldn’t just start with one / then I’d probably search for others

The clear majority of the propositions were classified as reflecting subprocesses of examining the results, extracting information, and reflecting and deciding to iterate or stop the seeking, 904 (27.8%), 742 (22.9%), and 676 (20.8%) respectively. When the participants examined the results, they commented their first impressions, “with these keywords I get just consultant companies / at least the first ones seem to be the kind of companies that offer these services” (Participant 2). The propositions like “there isn’t anything about usability here / what is this anyway / virtual library / they tell about style sheets / about html” (Participant 1), which

implied more thorough investigation were classified as extracting information. After gathering some information the participants reflected on their progress, e.g. Participant 7 said “but these are still articles they don’t quite answer to my need I have now”.

107 (3.3%) propositions could not be classified in any of the subprocesses. They were not related to the information seeking; for instance comments about the hardware used or getting a cup of coffee to drink.

Table 3. Examples of propositions reflecting different social knowledge used.

Social knowledge	P	Example
Participant	P6	oh yeah I’ve been here already at some stage
Individuals	P4	I would certainly get something more for support / I’d call to you for instance / I wouldn’t go just blindly to seek
	P5	for example this jukka korpela who’s manual I’ve used when studying html
Groups	P2	there are quite a lot of same themes that we have here emphasised
	P2	what user interface engineering this refers to that same company that user interface engineering

There were 400 (12.3%) propositions that contained indications to use of social knowledge. Among them, 46 (11.5%) referred to the participant him or herself. They usually implied that the navigational history of the participant was used as reference, e.g. “that url looks familiar so I think I’ve been here before” (Participant 7). The 111 (27.8%) propositions related to other individuals mainly contained comments about some well-known persons in the domain or a reference to a person the participant knew personally. For example, “jakob nielsen’s website / well that is of course one place to re look / naturally” (famous individual, Participant 1) and “as the matter of fact I could try that google / at least my husband praises it” (personally known individual, Participant 5). The majority of the proposition reflecting use of social knowledge (243, 60.7%) were related to groups. The participants seemed mainly to comment on the information providers, such

as an institute creating a web site like Participant 7 when saying “let’s see if I like their philosophy about this this issue or if I agree with it I’ll read more what they say”.

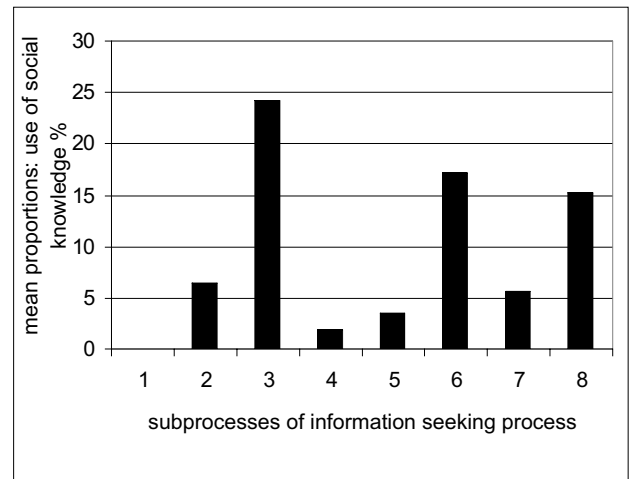


Figure 1. The distribution of the mean percentages of the propositions reflecting usage of social knowledge along Marchionini’s information seeking subprocesses (1 – recognise, accept, 2 - define problem, 3 - select source, 4 - formulate query, 5 - execute search, 6 - examine results, 7 - extract information, 8 – reflect, iterate/stop).

The distribution of the propositions reflecting usage of social knowledge in the different subprocesses is shown in Figure 1. One-tailed t-tests were conducted to see whether the mean percentages of the propositions reflecting the use of social knowledge differed from zero in any subprocess ($\alpha=0.05$). Social knowledge was statistically significantly used in the subprocesses select source ($t_{(12)}=-3.285$; $p=0.003$), examine results ($t_{(12)}=-3.068$; $p=0.002$), extract information ($t_{(12)}=-5.411$; $p<0.001$), and reflect, iterate/stop ($t_{(12)}=-2.097$; $p=0.029$).

4. DISCUSSION

This study aimed to clarify how social knowledge is used in information seeking in the Internet. The main result was that social knowledge is statistically significantly used when selecting the source or place

where to start looking for the information, while examining the results and extracting information, and when reflecting and making the decision to iterate or stop seeking.

Information seeking starts with recognising a gap in the knowledge. In this study the tasks were given to the participants. The tasks did not directly refer to any persons nor groups so it is natural that no indication for using social knowledge was found in recognising and accepting information need. However, this might not be the case in a more usual situation where the recognition of the information need might occur in a social interaction situation.

Although the target definition phase was identified both by Kuhlthau (1996) and Marchionini (1995) as a phase when information seekers might use social knowledge, no evidence for this was found. The participants mainly tried to describe the task in their own words and make sense of the task description given to them. With the participants who were not experts in the domain, it is understandable that they did not elaborate the problem definition to include social aspects, because they might lack the needed knowledge. However, the case of the participants who did know the domain, the lack of using social knowledge is surprising and needs more research. One explanation is that the situation, task descriptions given written on the paper, made the participants focus mainly on them, which might not happen in an everyday situation.

Typically when information is sought in the Internet, the target is not clearly defined (Saito and Ohmura 1998) and especially for the domain novices defining the problem might elicit another information seeking instance within the broader task. Golovchinsky and Belkin (1999) called this type of information seeking instance as information exploration that occurs when the information seeker has *“an incomplete or inadequate mental model for the purpose of addressing some information need”* (p. 23). Maybe for some participants the problem definition subprocess was replaced by another information seeking process during which they gathered information needed to define the problem. This possibility was not taken into account in the classification of the propositions and thus the use of

social knowledge in defining the problem might have been obscured.

Some participants clearly selected a source based on their personal connection to certain web sites, for instance participants knew the people who had created the content of the web site or regarded them otherwise highly.

No evidence for use of social knowledge when formulating the query was found, although social networks can be used as pointers to relevant information (Erickson and Kellogg in preparation). Although people definitely formulate the queries on the basis of domain knowledge, they do not seem to use social domain knowledge at this stage. This apparent controversy can be explained with the fact that people use their social knowledge when deciding where to look for the information. For instance, they might select a source because a person they respect has recommended it or they know a group that has provided a place where the information is likely to be found. Thus, when formulating the query they are already in a place that has been partly selected based on social information, so they might feel more comfortable using now other kinds of information available to them. It has also noted that not all participants were experts in the domains related to the tasks, so they might need to rely mainly on the information given in the task description when formulating the query.

The phase of executing the queries is mainly behavioural, thus no indication of use of social knowledge was found.

When the participants examined the results, e.g. scanned a listing of search results, they used social knowledge in various ways. The most usual ways were spotting any familiar names of individuals or groups and checking the URLs to see whether the page was a personal home page or a more formal one. People clearly used their social knowledge to filter out potentially unimportant results.

Extracting information has been described as reading, copying etc. with no implied use of social knowledge (Marchionini 1995). However, the results of this study indicate that people pay attention to social information,

e.g. the creator of the information or references to certain persons or groups, while extracting the information. For some participants knowing something about the people who had created the information was essential. For instance, when asked whether it was important for Participant 2 that there is personal information about the content producers, she answered “it gives... yes it sort of gives a face to the page, so that it’s not only some unknown institute that maintains it but this is some guy who is motivated and... he probably updates the pages regularly when there is something new”.

Every now and then during the information seeking the participants took a closer look where they were and if they had gathered enough information to be satisfied. They talked about the target they had defined and the information they had found up to that point. Since social knowledge was used to gather the information, it is quite understandable that people mentioned it also when reflecting on the information seeking.

In our social world, the players are individuals and groups. It was also possible to find evidence of social players in the verbalisations of our participants. Participants used themselves (“I”) as social references in the process of information seeking for both selecting favoured sources and guiding navigation. Other individuals (“you”) were most often either trusted or renowned information providers or sources of reference. But there were also comments that showed distrust about individuals, e.g. it seemed to be a widely held assumption that personal home pages were not a good source for valid information.

References to groups (in most cases “they”) were made the most. This is natural as organisations, rather than individuals, maintain a vast number of web sites. Our tasks also quite often lead the participants to sites maintained by universities and Web organisations (e.g. W3C). The participants used social knowledge to evaluate the content of the found documents based on information about the group of individuals or organisation providing the content. Groups and individuals with an affiliation to a renowned group were treated as more trustful sources of information.

4.1 Conclusions and implications

Our results are a good example of how seeking information in the Internet is more than just getting the information – people interact with the pages using their natural abilities to interpret social situations and taking advantage of their social knowledge. This can happen even if the other persons are represented in the situation only as a name in the text or even in a part of the URL of the page. Thus we are not “socially blind” in digital systems as described by Erickson and Kellogg (in preparation), but we are able to extract social information and use it even if it is not explicitly provided.

In fact, people actively seek for social information in the Internet. Thus, providing such information that helps people to select where to start the search, evaluate and examine results, and extract information could facilitate information seeking. It should be clearly stated who provided the content and their affiliations. It is also important to indicate the criteria and the process that were used when the content was created (e.g. use of review process). With such social information readily present the information seeker saves time and effort and the reflection on the relevance of the gathered information becomes easier.

4.2 Future research topics

Although the data seemed to fit reasonable well with Marchionini’s model of information seeking process, i.e. different subprocesses were identifiable from the data, there were some problems related to the analysis method used in this study. Because only verbalisations were included in the content analysis, some subprocesses, especially executing query, probably could not be analysed fully since they might be associated with considerable cognitive load which reduces verbalisation or be mainly expressed in overt behaviours. What was said and what was done were also in asynchrony every now and then. For instance, a participant was explaining how she selected to seek for information in a certain place and at the same time typing in a search query, thus the verbalisation was classified as the subprocess of selecting a source, although she was at that time actually formulating and

executing a query. A classification system that would combine the overt behaviours and verbalisations might capture better all subprocesses.

This study did not examine the differences between experts and novices. However, it came evident that it is worth investigating whether there are differences between experts' and novices' use of social knowledge in information seeking. An interesting subprocess is especially the problem definition that could be studied with information exploration framework (Golovchinsky and Belkin 1999) with novices.

Due to the preliminary nature of the study, the number of participants was limited and the results need to be confirmed with a larger population. In addition, clearly not all social aspects of information seeking can be studied in a laboratory, thus an ethnographically inspired follow-up study is being planned to verify the results and enrich the understanding how social aspects are manifested in information seeking.

REFERENCES

- Erickson, T. and Kellogg, W. A. (in preparation). Social Translucence: An approach to designing systems that support social process. submitted to *Transactions on Computer-Human Interaction*. Retrieved from http://www.pliant.org/personal/Tom_Erickson/st_TOCHL.html 23rd April 2000.
- Golovchinsky, G. & Belkin N. J. (1999). Innovation and evaluation of information. *SIGCHI Bulletin* 31(1), 22 - 25.
- Harper, R. H. R. (1999). Information that counts: a sociological view of information navigation. In *Social navigation of information space*. Springer, London. 80-89.
- Hoelscher, C. and Strube, G. (1999). Searching on the Web: Two types of expertise. *Proceedings of 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 305-306.
- Kuhlthau, C. C. (1996). *Seeking meaning: a process approach to library and information studies*. Ablex Publishing Corporation, Norwood, USA.
- Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge University Press, Cambridge.
- Munro, A. J., Höök, K. and Benyon, D. (Eds.) (1999a). *Social navigation of information space*. Springer, London.
- Munro, A. J., Höök, K. and Benyon, D. (1999b). Footprints in the snow. In *Social navigation of information space*. Springer, London. 1-14.
- Saito, M. and Ohmura, K. (1998). A cognitive model for searching for ill-defined targets on the web: the relationship between search strategies and user satisfaction. *Proceedings of 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 155-163.
- Wenger, E. (1998). *Communities of practice. Learning, meaning, and identity*. Cambridge University Press, Cambridge.