

Talking about the problem: a content analysis of pre-search interviews

[T.D. Wilson](#)

Högskolan i Borås, Borås, Sweden
and Leeds University Business School
Leeds, UK

Abstract

The pilot study reported here applies content analysis techniques to twenty interviews carried out in preparation for mediated searches as part of the 'Uncertainty in Information Seeking' project. The tools used in the analysis (Atlas.ti and TEXTStat) are described and their contribution assessed. TEXTStat, a free program, was used to produce frequency counts of the words used in the interview and to examine the context of those words. Atlas.ti was used to assign codes to the interview transcripts and to model the relationships among these codes. The mode of analysis of the cases is qualitative and interpretative and the results reveal the complexity of the information problems, the variety of motivations for undertaking a mediated search, the difficulties of expressing the search concepts, and the relationship between the need for a search and previous information seeking behaviour.

Introduction

Research on information behaviour is concerned with the behaviour of persons in seeking to find answers to their problems, with attempts to model that behaviour (e.g., [Wilson, 1999a](#)), with models of the actual search process and its characteristics (e.g., [Kuhlthau, 1994](#); [Ellis, 1989](#)), and with analyses of behaviour in particular work settings (e.g., [Huotari and Wilson, 1996](#); [Byström, 2002](#); [Heinström, 2003](#)) or in everyday life (e.g., [Chatman, 1987](#); [Savolainen, 1999](#)). Most of this literature is concerned with discovering the fundamental basis of behaviour in respect of information seeking, but how people conceive of the problems with which they have to deal has been subject to rather less investigation.

This paper attempts to explore the ways researchers talk about their problems in the course of pre-search interviews and to derive from their problem descriptions some understanding of the contexts of their problems and of the difficulties they experience in carrying out their own searches. The 'Uncertainty Project' revealed that researchers, generally, had not previously sought help in the conduct of searches, relying upon their own skills. Their self-efficacy in this will be compared with previous work on information searching strategies and the nature of the task they are seeking to carry out will be addressed and related to ideas of task complexity.

Two previous areas of research are relevant for the work reported here: the reference interview, and that of intermediary interaction in the search process.

The reference interview has long been regarded as an essential stage in the process of discovering information for people (e.g., [Taylor, 1962, 1968](#); [CACUL, 1979](#); [Jennerich, 1987](#); [Dewdney, 1988](#); [Ross, 2000](#)) and with the coming of online resources, and more recently, the development of digital libraries and 'virtual reference work', the subject has resurfaced as critical for the discovery of the information user's needs (e.g., [Tibbo, 1995](#); [Abels, 1996](#); [Giannini, 1999](#); [Martin, 2001](#); [Janes and Hill, 2002](#)). There have been a number of evaluations of the performance of reference librarians which show that, at least in the libraries covered by the evaluation, the reference interview is more honoured in the breach than in its observance. For example, Burton pointed out that these studies have shown that libraries: '...provide correct responses to between 50 and 60 per cent of the questions put to them by users. So frequently has this result been found that it has been enshrined in the literature as the "55 per cent rule".' (Burton, 1990: 203)

Whatever the reality of the reference interview in practice, however (and Hauptman, (1987) suggests that, to a large degree, it is unnecessary), there appears to have been relatively little use of reference interview data in research on information seeking behaviour. A study by Radford (1998) observed the reference interaction and interviewed users, but the focus of this work was on non-verbal behaviour and a content analysis of the interviews carried out after the observation. Earlier work by the same author ([Radford, 1996](#)) applied relational theory from the field of communication science to recollections of critical incidents by both librarians and library users, but not to transcripts of the interview. Recently, Wu and Liu (2003) examined the interaction of patron and reference librarian, but in order to establish the *elicitation styles* of the librarian, rather than to analyse the utterances of the information seeker. Dewdney and Mitchell (1996) analysed librarians' self-reports of *accidents* in the reference interview, rather than transcripts of the incidents. The closest work to that reported here appears to be that of Solomon (1997) who tested an analytical framework in a sample of nine conversations from two

information-seeking contexts. However, Solomon, was interested in the interaction process and the conversational elements, rather than in the determination of the information need.

The field of interaction between the information seeker and the intermediary is well researched and there is overlap between this field and that of the reference interview - the major difference is that research on user-intermediary interaction is more concerned with the dynamics of the situation throughout the search process. Various studies demonstrate that the process of interaction in the course of the search affects the user's perception of need and the expression of that need (e.g., [Markey, 1981](#); [Saracevic, et al., 1997](#); [Ellis, et al., 2002](#); [Wu and Liu, 2003](#)). Some work in this area, however, is devoted to the *pre-search interview*, which, conceptually, is identical in aim to the reference interview. For example, the well-known 'anomalous state of knowledge' (ASK) model ([Belkin, et al., 1982](#)) aims at identifying, through interviews with users, the anomalies in their representations of the problems they are seeking to resolve through a search. The proposition was that, if information retrieval systems could be designed to focus on these anomalies, they would be more effective in satisfying users' needs. Mokros, et al., ([1995](#)) also distinguish between the two: 'Most generally, the search interaction may be divided into two stages: the presearch interview and the online computerized search.'

The study reported here applies content analysis techniques ([Krippendorf, 1980](#); [Weber, 1990](#)) to transcripts of pre-search interviews in an attempt to discover, first the characteristics of the problem statement that may provide a framework for problem elicitation and secondly, a theoretical framework that might help to explain the motivation to seek mediated searches.

Methods

The study is based on interviews carried out in the course of the 'Uncertainty Project' ([Wilson, et al., 1999](#)) at the University of Sheffield. The enquiry population consisted of a self-selected sample of respondents who replied to a call for participation in the project. Consequently, it does not attempt to generalise but to discover, as far as possible, common characteristics in the enquirer's exposition of his or her information problem, such as the context of its occurrence, the extent of prior investigation of the problem, and the hoped-for results.

The transcripts show that the method used to elicit information to help the researcher search for information on behalf of the client was very successful. The clients were eager to express themselves and responded fully to the probing questions of the interviewer.

The data consisted of 111 transcripts of pre-search interviews, from which twenty were selected for in-depth analysis, in this pilot, which is in preparation for a larger scale analysis of the data. (The cases are listed in [Appendix 1](#).) Two methods of content analysis were applied: qualitative analysis was carried out with the help of [Atlas.ti](#) as the software tool and [TextSTAT](#) (a freeware program) was used for the analysis of word frequencies in the texts.

Each transcript was analysed by assigning codes indicative of:

- the factors driving the search for information;
- the characteristics of the information problem;
- the previous search activities, if any;
- factors that could affect the search process; and
- desired outcomes.

Not all of the transcripts contained information that represented these different factors, but, as may be seen from the transcript given as [Appendix 2](#), they could be very detailed. The accounts related mainly to the first question in the interview:

1. Client's description of the problem and nature of the enquiry

Before we begin the reference interview I'd like you to tell me in your own words a little about your problem and where you are in relation to it.

[Interviewer to probe once the client has made a statement to gain information about the nature of the topic, how it fits into the person's research generally, what the context is. How significant is the work to you - personally? What pressure, if any, are you under to carry out this work? What the purpose of the search is and whether the topic is based in a single discipline or is multi-disciplinary - In either case, which discipline(s) is/are involved?]

The TextSTAT software package was used to provide a frequency count of the words in the transcript and this was edited to merge singular and plural terms and other grammatical variants, and to remove common words, such as *and*, *to*, *from*, etc. Table 1 shows the frequency distribution for the transcription analysed in Figure 2.

From the frequency distribution alone, it is possible to see that the subject has something to do with policies for natural resources in Africa, with connections to environmental and cultural aspects, and that the disciplines likely to be of interest include geography, development studies, and agriculture.

Word	Freq.
policy(ies)	14
resource(s)	6
land	5
natural	5

environment	4
search	3
Africa	3
Namibia	3
province	2
international	2
Botswana	2
cultural	2
game	2
development	2
geography	2
farming	2
agriculture	2

Table 1: Word frequencies

TextSTAT also has a concordance feature, which provides a keyword in context type of format to show the location of keywords in the sentence, with the capability of sorting on either the word immediately before or immediately after the keyword. It is also possible to truncate a word and use the concordance to group variant forms. Figure 1 shows the concordance for the word-stem DEVELOP from transcript 418:

re were big political exchanges between DEVELOPED countries and developing countries, s
vision between the North, the core, the DEVELOPED countries, the rich countries on the
h relations, i.e. relations between the DEVELOPED countries and developing countries, h
and the developing, the south, the underDEVELOPED, the poor countries. I particularly
ars on the role of what we used to call DEVELOPING countries in international politics
orld Trade Organisation and the role of DEVELOPING countries or southern countries. So
there are a number of key southern or DEVELOPING countries that because of their size
ons between the developed countries and DEVELOPING countries, however we define those t
changes between developed countries and DEVELOPING countries, so if one thinks of conte
e definitions and on the other hand the DEVELOPING, the south, the underdeveloped, the

Figure 1: Concordance output from TextSTAT

This feature, together with the straight frequency count, is useful when identifying codes, merging codes, or building code *families* in Atlas.ti. In the example, we find 'developed countries', 'developing countries', and 'underdeveloped' - all of which we might wish to use to code this text, or which we might wish to merge into a single family.

Atlas.ti gives one the capacity to scroll through a text and select codes to identify the concepts used by the interviewee in talking about the problem. Atlas.ti maintains a list of the codes and of the text fragments (*quotations*) to which they apply. Figure 2 shows a part view of the main coding screen of the program.

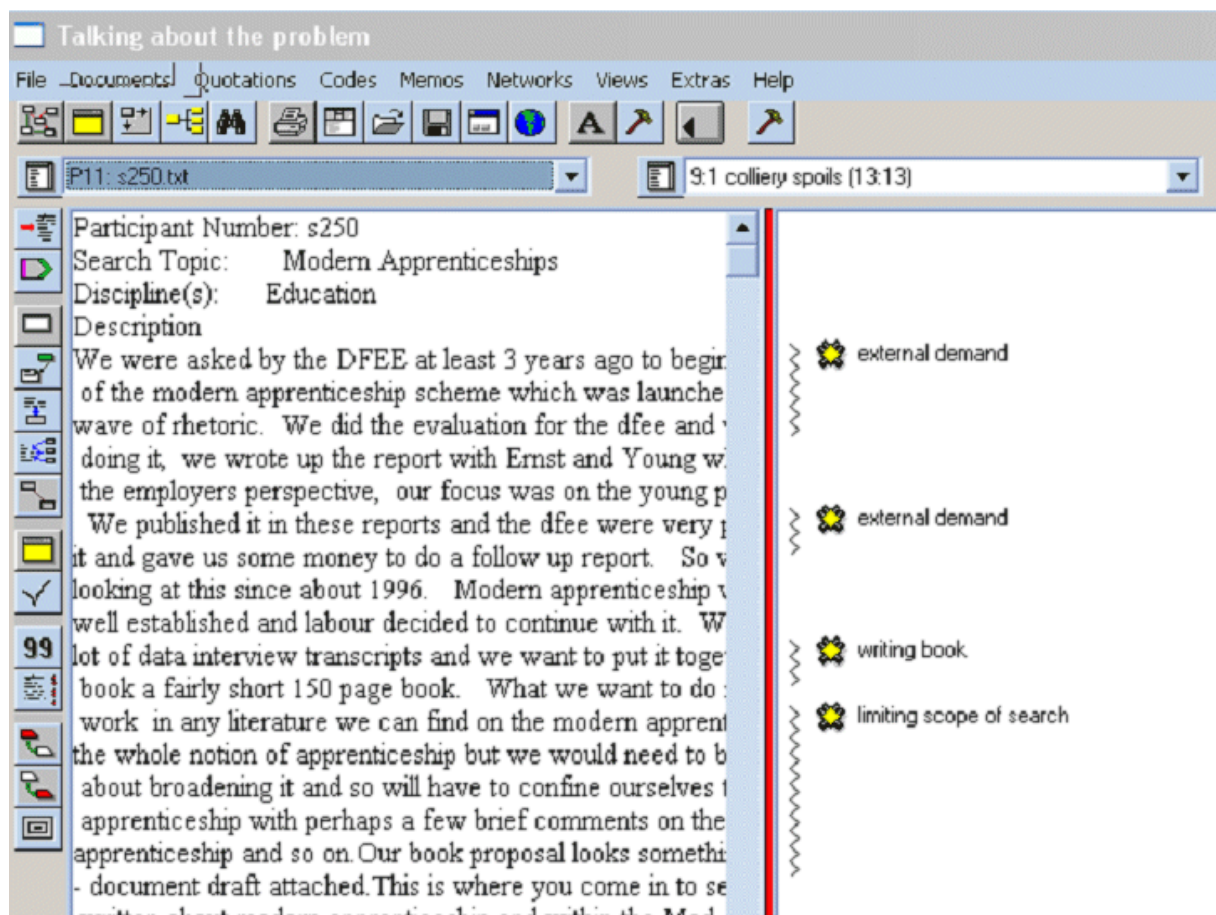


Figure 2: The Coding process in Atlas.ti

The *network* feature of Atlas.ti was used to construct network diagrams, linking coding terms, as a means of suggesting fruitful relationships to explore. These networks varied in complexity, from the very simple diagram shown in Figure 3 to the more detailed network shown in [Appendix 3](#).

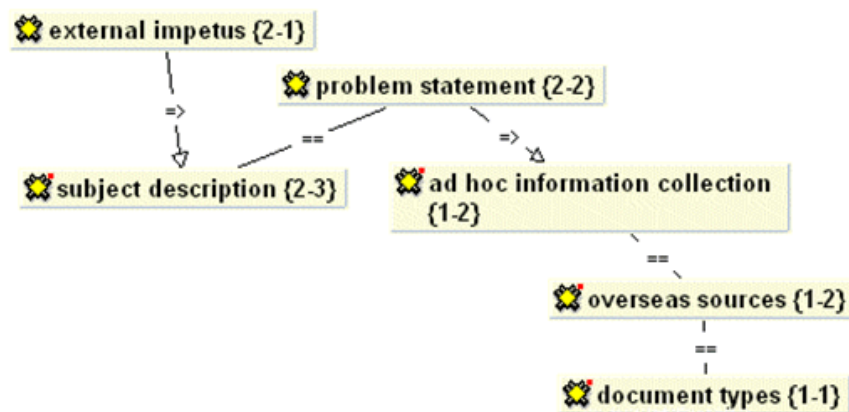


Figure 3: A simple network view

The network shows that the transcript contained an account of the external driver for the search, a description of the subject involved, a statement of the specific problem, an account of ad hoc information collection that had taken place, and an indication that the search should concentrate on certain overseas countries and document types.

Results

Disciplines represented

The twenty cases in this study spanned a wide variety of disciplines, as shown in Table 3 below. In fact, it is rather difficult to assign the cases to specific disciplines, since the tendency was for the interviewees to characterise their work as inter-disciplinary. For example, one of the Archaeology topics related archaeology and agriculture and one of the Management topics involved the relationship between management and music—the PhD student concerned had two supervisors, one in the School of Management and the other in the

Department of Music. However, regardless of the inter-disciplinary character of the research, the table shows a wide range of subject areas.

Aquatic botany
Archaeology (2)
Civil Engineering (2)
Control Systems
Cultural Studies
Education
Geography
Housing Design
Journalism
Management (2)
Neurophysiology
Physics
Politics
Psychology (3)
Social Psychology

Table 3: Disciplines represented in the study

(The no. in parentheses is the number of persons in that discipline.)

Previous information seeking activity

Table 4 shows the data on previous information seeking activity for all the cases in the Uncertainty Project (excluding missing cases). The table shows that the two most popular ways of seeking information were to carry out a database search independently of any intermediary and to carry out a Web search. Given that the majority of cases were from the University of Sheffield, the lack of mediated searching is not surprising, since financial stringencies over several years have not encouraged the development of library-based information searching. At the same time, the availability of access to a range of databases through the national, centrally-funded Bath Information and Data Services (BIDS) encourages academics to undertake their own searches.

Type of search	Search activity carried out?			
	Yes		No	
	No.	%	No.	%
Web search	86	46.0	101	54.0
Mediated search of database	15	8.0	172	92
Own search of database	100	53.2	88	46.8
Search of printed indexes	17	9.1	170	90.9
Library catalogue search	42	22.5	145	77.5
Library browsing search	38	20.3	149	79.7
Search of own collection	48	25.8	138	74.2
Search of colleague's collection	31	16.6	156	83.4

Table 4: Previous information seeking

(Note: more than one response possible.)

The following were fairly typical of the previous information seeking reported by interviewees:

'Probably very very limited if I am honest, I have used the STAR library catalogue, I have messed around with BIDS but probably not very successfully... I think I prefer going straight to the shelves and finding something. ...stuff is going to appear in certain journals, so there are quite familiar things like that and then there are things like conference papers... So, searching-wise I can't say I've searched that extensively. Perhaps the best way to describe it, is if you find one paper say via Star then using that bibliography to point you in new directions, I think more have been generated that way than going through the system. So I am happy to be surprised.' (s424)

'[Previous searching has been]...mainly through BIDS, and we have this electronic journal database and you can pull out 20 environmental journals and abstracts and there is also a hard copy of journal abstracts which we get circulated. Of course we all take journals of some form, but BIDS is definitely the one that we use most of all. We look at the different environmental web sites such as the EPSA Website which publishes their work and we do searches on that. (s322)

Motivations for the search request

From the analysis, several factors emerged as motivating the interviewees to request a search. These can be seen as of two types: the motivation for the research, and the motivation for the search.

The twenty cases represent a diverse set of motivations for the research (Table 5). In seven cases the research was for the PhD, with people at different stages in the process, some just beginning, some having already completed one or two years. The remaining thirteen cases represented academic staff in various stages of their careers, including one seventy-six-year-old Professor Emeritus, who was still researching and giving conference papers.

Type of project	No.
PhD project	7
Externally funded project	4
Developing or redirecting research area	4
Continuing research interest	2
Writing a book	1
Preparing patent submission	1
Database development project	1

Table 5: Motivation for the research

Turning to the more immediate motivation for the research, in some cases the researcher was simply taking advantage of an opportunity that happened to coincide with his or her work. For example:

'What I am trying to find out is the available literature that looks at the question of autism from a social perspective or a biographical perspective. I am trying to get away from the predominant psychological and social factors to see what other factors are involved in the field... when you sent the letter out this sounded like an interesting possibility of combining the two things.' (s428)

In the majority of cases, however, enquirers were expressing some dissatisfaction with their own capacity to search the relevant information sources: they were afraid that they would miss key papers, or they believed that they lacked the necessary search skills, they had difficulty in determining the appropriate keywords for searching, and expressed dissatisfaction with systems that offered little more than simple keyword searches. In the latter case, of course, they may not have bothered to explore any advanced search capability that did exist. Typical comments included:

'We have done various searches but we are not sure that we have covered all of the areas.' (s287)

'...if you put in keywords it is not as flexible as the search you describe to me where you... use... phrases. That has more potential to look at a wider range of references. So, in short, I am pretty successful when I see a reference and am using the right keywords, but getting the stuff out is pretty hard.' (s322)

'...it may be my fault that I am not very good at putting into keywords how to actually do the search. If someone shows me a title or an abstract I can quickly disregard it or not, yet... if you put the same keywords in that would get you what you wanted..., you would get a lot of other things, you would get a mass of other things.' (s278)

'I think in a way, perhaps there is a way of searching that I'm not familiar with, or a database or a magic bank of data that I don't know about....' (s425)

The difficulties of keyword searching are discussed further below, but it is worth noting, at this point, studies of user-intermediary interaction also identify this area as problematic for both user and intermediary. For example, the study by Saracevic *et al.*, (1997: 50) found that the category *terminology and restrictions* ('Elaboration on and modification of concepts, terms, keywords and descriptors; generation of terms; specification of borderlines; restrictions such as with respect to language, years; technical term spelling.') accounted for the largest proportions of elicitations by both users (32%) and intermediaries (55%).

In some cases, the enquirer believed that the Project might be able to access a wider range of sources:

'The reason we were interested in this was that it might give us access to different databases because a lot of the literature that we need is based overseas.' (s253)

'What I really need for my work to be convincing and for it to be accepted for publication I need to have engaged in a very thorough search strategy, i.e., identified all of the papers that actually measure people's past and future behaviour in a longitudinal study.' (s263)

The Professor Emeritus had a very wide knowledge of the literature relevant to his topic:

'Some of the best papers cited come from journals of about 1910, 1911, 1912 etc.'

but:

'What I would really like to find out is if there is any modern work on transpiration losses from shingle plants and whether there is any modern experimental work on the hydrology of shingle. I suspect there will be very little but these would be the things to home in on.' (s275)

The case of the researcher involved in a potential patent application was instructive:

'So this [a document held by the researcher] is what the UK patent agent threw up in his standard search and we found a couple of things in here that we were aware of generally, but again, a rather restricted set of search words. In no way do I think that this is what's out there. In a sense I don't know, of course, I don't know how you approach this in general terms, but I assume that at least one seed paper in the patents and the literature would be helpful, but on the other hand I am not looking for a duplication of what we have done already, I am looking for something completely fresh and much better and comprehensive.' (s.403)

Here we have doubts about the efficacy of a previous, professional search; a perception that the set of search terms used was inadequate; and an admission of a lack of competence to do better himself. All resulting in a fairly high level of uncertainty about the whole exercise.

Task characteristics

The relationship between task complexity and information seeking behaviour has been explored by a number of researchers, most notably by researchers at Tampere University (e.g., [Byström & Järvelin, 1995](#); [Vakkari, 1999](#); [Byström, 2002](#)). However, that work has mainly been concerned with tasks of an administrative or educational character, and here we are concerned with tasks of varying degrees of intellectual demand. There is a problem in trying to categorise tasks of this kind within the general framework proposed by Byström & Järvelin (1995), since they do not possess the same kind of decision-type characteristics as those involved in administrative tasks. In fact, it can be argued that the information seeker in these interviews is less interested in decision-making than in discovery.

It may be possible, however, to devise a categorisation of tasks from the search information given by the information seeker. For example, in the case of a search for existing patents relating to a discovery, the user wishes to be certain that no relevant patent exists—the user might be said to have an *exclusive* motive, since the failure to find patents means that the process of patenting the invention can proceed, while the existence of a patent would mean that the process would have to cease. In the vast majority of cases, however, the motive is *inclusive*—the user wishes to discover, if not everything about a problem, then, at least, is open to the possibility that more exists than s/he has been aware of.

Xie (2000), in an analysis of forty cases of users' information search behaviour, discusses the types of goals that may be ascribed to users: the long-term goal, which the user will pursue over a lengthy period of time; the leading search goal, that is the immediate task that gives rise to the search; the current search goal, or the specific search results hoped for in the current search; and *interactive intention*, or sub-goals associated with the search. The interview transcription given as [Appendix 2](#) can be interpreted in these terms. The long-term goal is defined in the statement: 'Myself and a number of other colleagues have been researching properties of colliery spoils for a number of years now primarily focusing on the construction of landfill liners.' The leading search goal by the statement: 'Now we have been asked by the Environment Agency to perhaps put a document together on best practice using colliery spoils, not only for landfill liners, but for other earthworks and construction aspects.' and the current search goal by, 'As part of that we need to draw together all of the literature that is available from all the different sources.'

Determining whether and how these purposes vary in *complexity* would be very difficult, however, since the complexity perceived by the user is likely to vary with the degree of expertise they already possess in the field of investigation. There is some work on intellectual task complexity but the tasks tend to be more specific than the very diffuse task of, say, researching for one's Ph.D. For example, Chan *et al.* (1999) examined task complexity defined in terms of simple and complex queries put to a database, finding that the more complex queries took longer to perform and were less successful. Bagga *et al.* (1999) developed a categorisation of the complexity of questions put to a question-answering system, concluding that the more complex a question, the more difficult it was to extract an answer. Research of this kind is of little help in seeking to assess the complexity of more generalised tasks and there may be a useful line of research in seeking to fill the gap.

Specifying the problem for the search

When we compare the kinds of search strategies that professionals can adopt with the more advanced databases with those adopted by searchers in general, the contrasts are striking. Various studies (e.g., [Spink et al. 1998](#); [Spink and Xu 2000](#)) have shown that those who search the Web, for example, adopt very limited strategies, with the mean number of words being between two and three. Less than 30% of searchers use any Boolean operators even where this is possible. This is confirmed by an examination of the search strings used when people search the electronic journal Information Research: data were collected from searches carried out by users on the 5th and 6th January 2004 and Table 6 shows the first ten search expressions (alphabetically ordered):

analysing interview data
architect
archiving historic maps
archiving maps
attitudes of teachers

big five factor
canon
case studies
content management system
data mining

Table 6: Search expressions

Whatever the complexity of the users' information problems, they appear to seek information by using quite simple formulations.

At times, one can see that the same person is trying out different search strategies: it seems unlikely that the following search strings were used coincidentally by different searchers:

- information seeking behaviour of contractor
- information seeking behaviour of construction industry
- information seeking behaviour contractor construction professionals

Boolean expressions are not possible with the search engine used, so it appears that a single user is trying differently phrased versions of the same search problem.

A skilled intermediary, however, can elicit a great deal of information from a client, as we have seen, and can use a wide range of techniques (when the database allows) to try to discover information of relevance, and, where the client requires it, a limited range of information of relevance. Even very simple steps can significantly reduce the volume of material retrieved: one of the phrases used by a searcher of Information Research was *digital divide*, which returned 177 hits; simply treating the two words as a phrase reduced the number of items retrieved to eight. Similarly *big five factor* refers to the well-known five factor personality test and, if it is expressed as "Big Five factor" and searched for as a phrase, one document is retrieved instead of the user having to scan through 244 largely irrelevant hits.

Turning to the cases explored here, topic 275 was on the 'eco-vegetation of shingle plants', but if that phrase is used in either Web of Science or in a Web search using Google, it returns no hits. However, the pre-search interview revealed a great deal more about the subject under investigation and it was possible, through an analysis of the transcript to arrive at a number of search options:

shingle AND plants AND water
 (shingle OR pebbles) AND plants AND water
 shingle AND plants AND (water OR tidal)
 ;(shingle OR pebbles) AND plants AND (water OR tidal)
 shingle beach AND plants AND nutrient

all of which succeeded in retrieving varying numbers of relevant documents from Web of Science. The use of the fifth formulation (shingle beach AND plants AND nutrient) also succeeded in retrieving fifty-five documents from the Web, a number of which would have been of at least contextual interest to the enquirer.

The significance of the appropriate choice of search terms is confirmed in a recent study of students' information searching by Pennanen and Vakkari (2003: 764) which notes that, '...students' ability to express their knowledge structure in query terms affects search success directly without the mediation of the number of facets and terms or the search tactics used.'

Discussion and conclusion

The immediate motivation for taking advantage of the opportunity presented by the Project to have an intermediary carry out a search, seems generally to have been a recognition on the part of clients that their own attempts at searching (and occasionally attempts by others) had been less than completely effective. The earlier searches had generally been carried out on databases subscribed to by the University Library and available through the campus network. Comments show that clients were well-aware of their 'self-efficacy' in this respect and the findings of related research suggest that searchers in general tend to use very simplistic approaches to the task.

Nevertheless, the data reveal that this specific decision to seek a mediated search was embedded, generally, in previous attempts to locate information. Fifty-three percent of all of our clients had carried out a database search of their own, and 46% had carried out a Web search (note that the categories were not mutually exclusive and some clients had done both and had also investigated other ways of discovering information.)

The general nature of the problem presented also had an effect on the time-pressure under which clients were working: the PhD students generally had a rather unhurried approach to the information searching part of their work, since they knew that either, it was the beginning of their work and that they would be revisiting their searches throughout their work, or it was at the end of their work and they were doing a final check on the current state of their original literature review. A similar view of time-pressure was found in those who were in the course of redirecting their research - they were, essentially, seeking a direction and knew what would be helpful when they found it, rather than having a clear idea of what they were expecting to find. Those involved in funded research projects were under greater pressure because of deadlines, either for the submission of proposals, or because of the completion of the work or stages in the research.

Figure 4 suggests a general model of the factors that appeared to affect the decision, at least of some clients, to take advantage of the offer

of a mediated search:

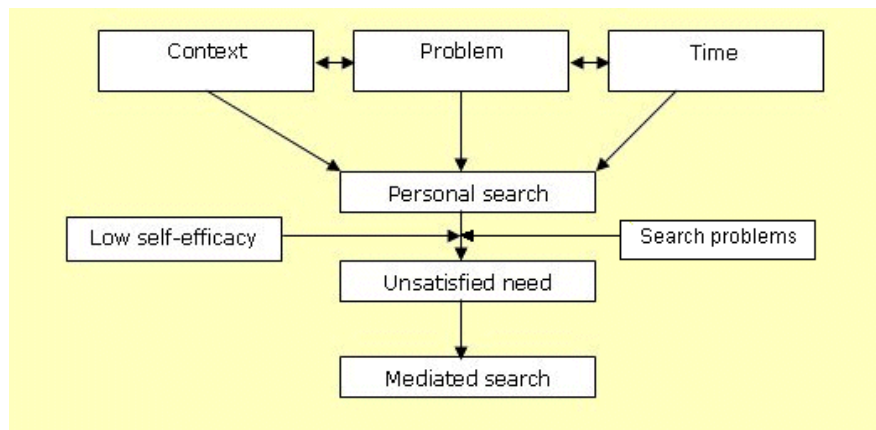


Figure 4: The demand situation

As a consequence of the financial pressures experienced by universities in the UK over the past twenty years, it is probable that many are in the situation of having to limit the extent to which trained librarians can conduct searches on behalf of users. However, this work suggests that, in spite of the popularity of personal searching, there is likely to be a strong latent demand for mediated searches.

The transcripts reveal in some considerable detail the complexity of the problems explored by the clients. The degree of complexity varies; in some cases, it is because of the context of the enquiry (for example, the agricultural development work in Africa, which was complicated by the foreign sources of useful documents), while in others it is a consequence of the problem area itself (for example, the growth of plants on shingle beaches), and in some, both context and content (for example, the patent search problem relating to a complex problem in electronics). Uncovering this complexity and ensuring that the search process discovers useful sources is the function of the skilled librarian, who can convert the problem description into search formulations appropriate to the sources used. Information users themselves do not possess these skills as a matter of course and, in any event, often lack means of access to some of the databases that might be of greatest use to them. In addition, unless strongly motivated to master this new medium, they are likely to lack the time to acquire the skills - this is particularly the case in the UK where financial pressures and assessment programmes have severely restricted the time available to academics.

Finally, we have seen that the tools used in this exercise, qualitative analysis (Atlas.ti) and word-frequency analysis (TextSTAT) are effective in revealing underlying conditions in the information seeking process and that, in particular, they can be used to analyse the 'reference interview' in ways that are instructive to the research, if not readily applicable to library practice.

Acknowledgements

My thanks to Allan Foster, the Research Assistant on the Uncertainty Project, to the former British Library Research and Innovation Centre for their support for the project, to the anonymous referees for their comments on the draft, and to Professor Pertti Vakkari for further helpful comments.

References

- Abels, E. (1996). The e-mail reference interview. *RQ*, **35**(3), 345-358.
- Bagga, A., Zadrozny, W., and Pustejovsky, J. (1999). Semantics and complexity of question answering systems: towards a Moore's Law for natural language engineering. In *Proceedings of the 1999 AAAI Fall Symposium on Question Answering Systems*. (pp. 1-10) North Falmouth, MA: American Association for Artificial Intelligence. (Technical Report FS-99-02)
- Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, **38**(2), 61-71.
- Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). ASK for information retrieval: Part II. Results of a design study. *Journal of Documentation*, **38**(3), 145-164.
- Burton, P. (1990). Accuracy of information provision: the need for client-centered service. *Journal of Academic Librarianship*, **22**(4), 201-215.
- Byström, K. (2002) Information and information sources in tasks of varying complexity. *Journal of the American Society for Information Science and Technology*, **53**(7), 581-591.
- Byström, K. & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management*, **31**(2), 191 - 213.
- CACUL Symposium on the Reference Interview (1979). *The reference interview: proceedings of the CACUL Symposium on the Reference Interview, held at the Annual Conference of the Canadian Library Association, Montreal, P.Q., June 9 to 15, 1977*. Ottawa: Canadian Library Association.
- Chan, H. C., Tan, B. and Wei, K. (1999). Three important determinants of user performance for database retrieval. *International Journal of Human-Computer Studies*, **51**, 895-918.
- Chatman, E.A. (1987). The information world of low-skilled workers. *Library and Information Science Research*, **9**(4), 265-283.
- Dewdney, P. (1988). The effective reference interview. *Canadian Library Journal*, **45**(3), 183-184.
- Dewdney, P., and Michell, G. (1996). Oranges and peaches: understanding communication accidents in the reference interview. *RQ*,

- 35(4): 520-523, 526-536.
- Ellis, D. (1989). A behavioural approach to information retrieval design. *Journal of Documentation*, **45**(3), 171-212.
 - Ellis, D., Wilson, T.D., Ford, N., Foster, A., Lam, H.M., Burton, R., & Spink, A. (2002). Information seeking and mediated searching: Part 5. User-intermediary interaction. *Journal of the American Society for Information Science and Technology*, **53**(11), 883-893.
 - Giannini, T. (1999). Rethinking the reference interview—from interpersonal communication to online information process. In *Knowledge: creation, organization and use. Proceedings of the ASIS Annual Meeting. vol. 36*, (pp. 373-80). Washington, DC: American Society for Information Science and Technology.
 - Hauptman, R. (1987). The myth of the reference interview. In B. Katz & R. Fraley (Eds.), *Reference services today: from interview to burnout*. (pp. 47-52). Binghamton, NY: Haworth Press.
 - Heinström, J. (2003). [Five personality dimensions and their influence on information behaviour](http://informationr.net/ir/9-1/paper165.html). *Information Research*, **9**(1) paper 165. Retrieved 27 December, 2003 from <http://informationr.net/ir/9-1/paper165.html>
 - Huotari, M-L., and Wilson, T.D. (1996). The value chain, critical success factors and company information needs in two Finnish companies. In P. Ingwersen & N.O. Pors (Eds.), *Information science: integration in perspective*. (pp. 311-323). Copenhagen: The Royal School of Librarianship.
 - Janes, J. and Hill, C. (2002). Finger on the pulse. *Reference and User Services Quarterly*, **42**(1), 54-65.
 - Jennerich, E.Z. (1987). *The reference interview as a creative art*. Littleton, CO: Libraries Unlimited.
 - Krippendorff, K. (1980). *Content analysis: an introduction to its methodology*. Beverly Hills, CA: Sage Publications.
 - Kuhlthau, C. C. (1994). *Seeking meaning: a process approach to library and information services*. Norwood, NJ: Ablex Publishing.
 - Markey, K. (1981). Levels of question formulation in negotiation of information need during the online pre-search interview: a proposed model. *Information Processing & Management*, **17**(5), 215-225.
 - Martin, K. (2001). Analysis of remote reference correspondence at a large academic manuscripts collection. *American Archivist*, **64**(1), 17-42.
 - Mokros, H.B., Mullins, L.S., & Saracevic, T. (1995). Practice and personhood in professional interaction: social identities and information needs. *Library and Information Science Research*, **17**(3), 237-257.
 - Pennanen, M. & Vakkari, P. (2003). Students' conceptual structure, search process, and outcome while preparing a research proposal: a longitudinal case study. *Journal of the American Society for Information Science and Technology*, **54**(8), 759-770.
 - Radford, M.L. (1996) Communication theory applied to the reference encounter: An analysis of critical incidents. *Library Quarterly*, **66**(2), 123-137.
 - Radford, M.L. (1998) Approach or avoidance? The role of nonverbal communication in the academic library user's decision to initiate a reference encounter. *Library Trends*, **46**(4), 699-717.
 - Ross, C.S., Nilsen, K., and Dewdney, P. (2000) *Conducting the reference interview : a how-to-do-it manual for librarians*. New York, NY; London: Neal-Schuman.
 - Saracevic, T., Spink, A. & Wu, M-M. (1997). [Users and intermediaries in information retrieval: what are they talking about?](http://www.win.tue.nl/~laroyo/2ID10/Resources/users-and-intermediaries.pdf) In Anthony Jameson, Cécile Paris, and Carlo Tasso (Eds.), *User Modeling: Proceedings of the Sixth International Conference, UM97*. (pp. 43-54). New York, NY: Springer. Retrieved 20 May, 2004 from <http://www.win.tue.nl/~laroyo/2ID10/Resources/users-and-intermediaries.pdf>
 - Savolainen, R. (1999). Seeking and using information from the Internet. The context of non-work use. In T. D. Wilson & D. K. Allen (Eds.), *Exploring the contexts of information behaviour: Proceedings of the 2nd International Conference on Information Seeking in Context, August 12-15, 1998. Sheffield, UK*. (pp. 356-370.). London: Taylor Graham.
 - Solomon, P. (1997). Conversation in information-seeking contexts: a test of an analytical framework. *Library & Information Science Research*, **19**(3), 217-248.
 - Spink, A. and Xu, J.L. (2000). [Selected results from a large study of Web searching: the Excite study](http://informationr.net/ir/6-1/paper90.html). *Information Research*, **6**(1), paper90. Retrieved 6 January, 2003 from <http://informationr.net/ir/6-1/paper90.html>
 - Spink, A., Bateman, J. and Jansen, B.J. (1998). [Searching heterogeneous collections on the Web: behaviour of Excite users](http://informationr.net/ir/4-2/paper53.html). *Information Research*, **4**(2), paper 53. Retrieved 6 January, 2003 from <http://informationr.net/ir/4-2/paper53.html>
 - Taylor, R.S. (1962). The process of asking questions. *American Documentation*, **13**(4), 391-396.
 - Taylor, R.S. (1968). Question-negotiation and information-seeking in libraries. *College & Research Libraries*, **29**(3), 178-194.
 - Tibbo, H.R. (1995). Interviewing techniques for remote reference: electronic versus traditional environments. *American Archivist*, **58**(3), 294-310.
 - Vakkari, P. (1999). Task complexity, information types, search strategies and relevance: integrating studies on information seeking and retrieval. In T. D. Wilson & D. K. Allen (Eds.), *Exploring the contexts of information behaviour: Proceedings of the 2nd International Conference on Information Seeking in Context, August 12-15, 1998. Sheffield, UK*. (pp. 35-54). London: Taylor Graham.
 - Weber, R. P. (1990). *Basic content analysis* (2nd ed.). Newbury Park, CA: Sage Publications.
 - Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, **55**(3), 249-270.
 - Wilson, T.D., Ellis, D., Ford, N. & Foster, A. (1999). *Uncertainty in information seeking. Final report to the British Library Research and Innovation Centre/Library and Information Commission on a research project carried out at the Department of Information Studies, University of Sheffield*. Sheffield: University of Sheffield, Department of Information Studies for the British Library Board. Retrieved 27 December, 2003 from <http://informationr.net/tdw/publ/unis/>
 - Wu, M.M., & Liu, Y.L.I. (2003). Intermediary's information seeking, inquiring minds, and elicitation styles. *Journal of the American Society for Information Science and Technology*, **54**(12), 1117-1133.
 - Xie, H. (2000). Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *Journal of the American Society for Information Science and Technology*, **51**(9), 841-857.
-

Appendix 1: List of case topics

s250 - Modern apprenticeships
s252 - Marketing and change within the music industries: analysis and future options
s253 - Land use policies and natural resource practices in Southern Africa
s263 - Habits.
s275 - Eco-vegetation of shingle plants
s278 - Optimisation of heat treatment of steel using techniques such as neural networks
s287 - Colliery spoil
s302 - The highland clearances
s314 - The agricultural economy of Western Asia in the last 10,000 years
s322 - Electron and carbon balances for understanding degradation of pollution
s403 - Pulsed plasmas
s411 - Psychological aspects of risk and safety in the work place
s413 - Human factors in the use/adoption of intranet technology
s414 - Crime reporting in media
s418 - North/South dimension of global political economy
s420 - Neuroimaging, neurotransmitters and neurodegenerative disease
s421 - Transformation in South African public services since 1994
s425 - The role of cultural industries in urban regeneration
s426 - Tenant participation in housing design
s428 - Social autism

Appendix 2: Transcript of a problem description

Client: s287

Search Topic: Colliery Spoil

Discipline(s): Civil Engineering, Geology, Geo-chemistry, Soil Science, Microbial-pyrites, Geography

Description:

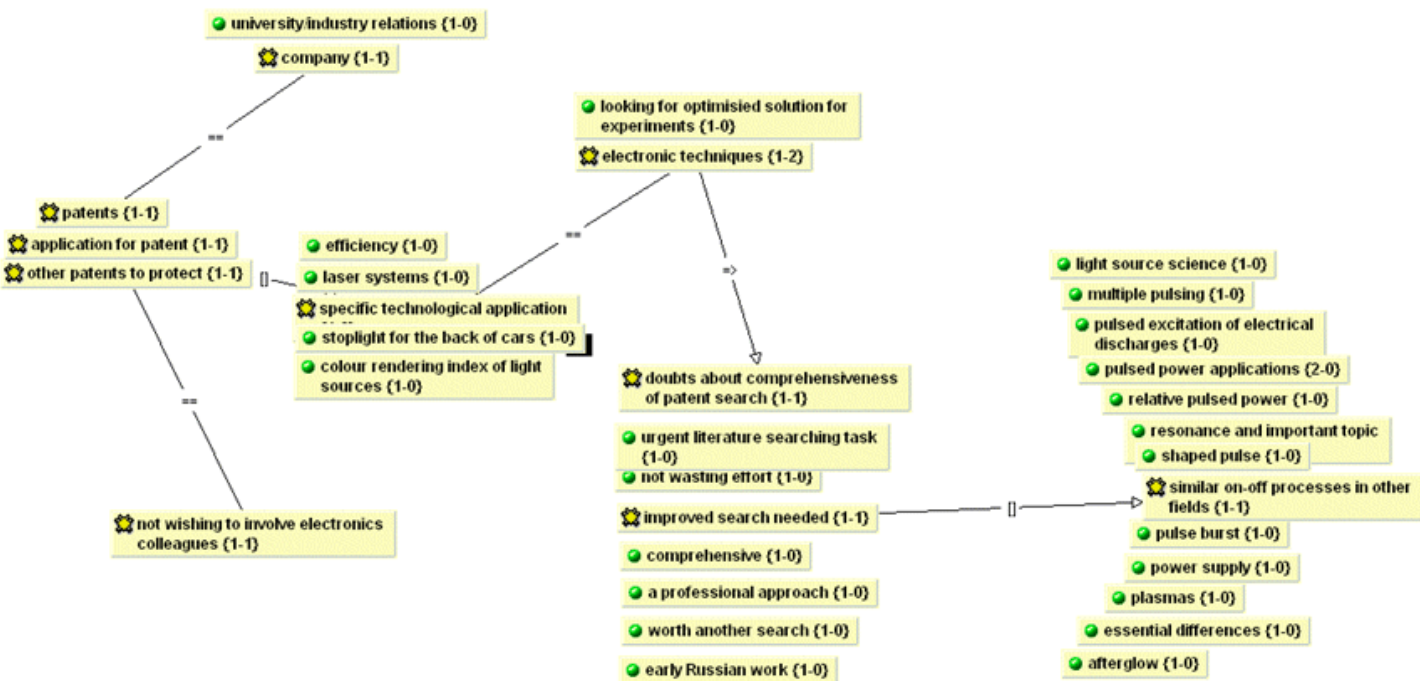
Myself and a number of other colleagues have been researching properties of colliery spoils for a number of years now primarily focusing on the construction of landfill liners. Now we have been asked by the Environment Agency to perhaps put a document together on best practice using colliery spoils, not only for landfill liners, but for other earthworks and construction aspects. As part of that we need to draw together all of the literature that is available from all the different sources.

The problem is that colliery spoil does come under a number of different headings, you can class it as a mine waste, you can look at related topics, you can look at properties of clays and soils in terms of their capacity as a landfill liner. Now a landfill liner, I don't know if you know anything about landfills, but [they are] big holes in grounds where people throw waste. That's what they used to do in the past, but people realised that as rain comes into the waste you form a fairly nasty mix called leachate, a chemical that trickles into ground and eventually maybe gets to an aquifer. Now they line everything usually with a metre thick layer of clay, but you can use colliery spoil, which implies waste rock and material arising from coal mining.

So we are after really all relevant literature, it's not only to do with the material, the main property we are interested in is its permeability, its ability to transmit water. In our case we are trying to minimise the transmission of water and fluid but we are also interested in the strength and stiffness properties of that material, how it's constructed, how to place it. There is quite a lot of science in the compaction process and getting that optimised. So there are a lot of strands to draw on. But the primary aim is to get the permeability of this material as low as possible and as reliably as possible. But there are issues of how you construct it, which can draw from other fields or issues of reaction with chemicals: does it get degraded by the leachate that comes out of the landfill? does it in itself produce chemicals? It is known that there are a number of chemicals that can be leached out of coal by rain - not a huge problem but they can be a problem in the short term.

There will, I imagine, be related research on other mine waste, we are focusing mainly on coal-mine waste, but there are other mine wastes which people may have done work on and the primary properties of colliery spoil is that it is a mix of clay, silt, sands and gravel, so it is a whole mix of different particle sized soils and people have done work on those sorts of soils with a large variation in grade size so those areas would come in as well. So there are a lot of areas to draw on and to sort of group together the best research or data that is available that we can pull together into what at first sight might seem a fairly specific topic. Clearly there is work in journals and conferences that we want to get hold of. We have done various searches but we are not sure that we have covered all of the areas. That is probably the gist of it.

Appendix 3: A more complex network



Note: the links have been simplified in this network to show linkages between groups of codes