Information search in a professional context – exploring a collection of professional search tasks

Verberne, S.; He, J.; Wiggers, G.; Russell-Rose, T.; Kruschwitz, U.; Vries, A.P. de

Information search in a professional context – exploring a collection of professional search tasks

Suzan Verberne
Leiden University
s.verberne@liacs.leidenuniv.nl

Jiyin He
Signal Media
jiyin.he@signalmedia.co

Gineke Wiggers
Leiden University
g.wiggers@law.leidenuniv.nl

Tony Russell-Rose
UXLabs
tgr@uxlabs.co.uk

Udo Kruschwitz
University of Essex
udo@essex.ac.uk

Arjen P. de Vries
Radboud University
arjen@acm.org

arXiv:1905.04577v1 [cs.IR] 11 May 2019

ABSTRACT

Search conducted in a work context is an everyday activity that has been around since long before the Web was invented, yet we still seem to understand little about its general characteristics. With this paper we aim to contribute to a better understanding of this large but rather multi-faceted area of 'professional search'. Unlike task-based studies that aim at measuring the effectiveness of search methods, we chose to take a step back by conducting a survey among professional searchers to understand their typical search tasks. By doing so we offer complementary insights into the subject area. We asked our respondents to provide actual search tasks they have worked on, information about how these were conducted and details on how successful they eventually were. We then manually coded the collection of 56 search tasks with task characteristics and relevance criteria, and used the coded dataset for exploration purposes. Despite the relatively small scale of this study, our data provides enough evidence that professional search is indeed very different from Web search in many key respects and that this is a field that offers many avenues for future research.

KEYWORDS

Professional search, complex search tasks, data collection

1 INTRODUCTION

Professional search is defined as the searching carried out by experts for work purposes [5, 13]. According to the literature, three specific characteristics of professional search differentiate it from web search: (a) The search tasks are **complex and specific** [10, 11]; (b) The professional searcher is an **expert** in the search domain [4]; (c) Professionals have **exploratory search needs** [3] that require

multiple queries and include browsing and analysing multiple documents [9].[1]

Search conducted in a work context is an everyday activity that has been around since long before the Web was invented. However, professional search has gained much less attention from the academic IR community than web search. As a result, many aspects of professional search are still unknown [7, 16]. Therefore even small, qualitative data sets can be valuable. They allow us to learn, case by case, how to approach difficult search tasks in realistic settings.

With this paper we contribute to a better understanding of professional search. Unlike system-oriented studies that aim at measuring the effectiveness of search methods for a particular task, we decided to take a step back by conducting a survey among professional searchers to understand their typical search tasks. We asked our respondents to provide actual search tasks they have worked on, information about how these were conducted and details on how successful they eventually were. We then manually coded the collection of 56 search tasks with task characteristics and relevance criteria, and used the coded dataset for exploration purposes. We address the following research questions:

RQ1 To what extent are the characteristics of professional search (a)–(c) reflected by the data acquired in our survey?

RQ2 Are these characteristics sufficiently pronounced to justify treating professional search as a separate search genre?

RQ3 Are the needs, goals and behaviours of professional searchers sufficiently homogeneous and consistent to justify viewing 'professional search' as a coherent, single field of enquiry?

2 BACKGROUND AND RELATED WORK

Collections of professional search tasks. Most collections of search tasks have been created in the context of TREC. A few of the relevant tracks for professional search are the total recall track, the enterprise track, and the legal track. TREC collections have been designed for evaluation purposes; they have shown to be indispensable for system comparisons and benchmarking. To the contrary, our collection is not meant for benchmarking purposes. The collection comprises user-created descriptions of a work task, created after completion of the task. Although our work does not contribute a test collection (it has no relevance assessments), the topics in our collection were collected similar to those in the iSearch collection [8], where the searchers themselves (experts) formulated

---

[1]Note that point (b) differentiates professional search from enterprise search [6].

You could help us collecting example search tasks. Please describe one search task that you have undertaken recently. Please be as specific as possible (not: "trying to find papers", but "trying to find papers that referenced Ingwersen and Järvelin (2005)").

- What was the goal of the search?
- Describe which actions you took to search (which search engine, queries, metadata filters)
- Describe how you judged the relevance of the found information. What were the relevance criteria?
- What was the outcome of the search? (did you find what you were looking for; how satisfied were you?)

Figure 1: Instructions given to respondents of our survey

topics that include fields for work task context, information need, and ideal answer.

Coding schemes for search tasks. Search tasks can be coded in a variety of ways. In early work, Bates [1] identified a set of 29 search 'tactics' which she organised into four broad categories of information seeking behaviour. Ellis and colleagues [2] developed a model consisting of a number of broad information seeking behaviours, noting also that it is possible to display more than one behaviour at any given time. Makri et al [9] extended this work, focusing on the information behaviours observed within the legal profession. More recently, Russell-Rose et al [14] used a grounded-theory approach to identify a taxonomy of information behaviours derived from a corpus of enterprise search tasks.

In many of these previous studies of information seeking behaviour, interview transcripts have served as the primary data source, offering an indirect, verbal account of end user information behaviours. By contrast, the data source used in this study represents a self-reported account of information behaviour, generated directly by end users (albeit retrospectively).

Relevance criteria. In 1998, Rieh and Belkin [12] identified seven different factors of information quality: source, content, format, presentation, currency, accuracy, and speed of loading, and two different levels of source authority: individual and institutional. Savolainen and Kari [15] found in an exploratory study that specificity, topicality, familiarity, and variety were the four most mentioned criteria in user-formulated relevance judgments, but there was a high number of individual criteria mentioned by the participants. We include the relevance factors from these prior works in the scheme for coding the relevance criteria mentioned by our respondents.

3 DATA

3.1 Collecting search tasks

The data was collected by surveying a non-representative sample of professional searchers. We defined the target group of the survey as "everyone who regularly performs complex search tasks at work in environments other than general web search". This included information specialists in various domains, but also librarians, scientists, lawyers, and other knowledge worker professions. We distributed

We manually coded the search tasks according to the following dimensions:

- Topic is well defined; keep in the set (yes/no/not sure)
- Topic domain (Computer science / Humanities / Legal / Medical / Other)
- Search for self or other (Self / Other / Unknown)
- Number of search systems/databases mentioned
- Satisfaction score (2: good; 1: sufficient; 0: undefined; -1: negative)
- Relevance criteria (Document type; Expertise level; External relevance; Language; Source reliability; Publication date; Quality; Sensitivity/Recall; Specificity/Precision; Topical relevance (True or false, multiple options possible)

Figure 2: Coding scheme used by the authors to code the collected search tasks. We refer to the first 5 as 'the main task characteristics'.

the survey in our own networks (e-mail, LinkedIn, and other social media), and through a number of professional mailing lists and newsletters that are distributed among information specialists and librarians in several domains. The instructions given to the respondents are shown in Figure 1. We also asked them about their field of expertise, their age and the number of years of professional experience.

3.2 Manual coding

In order to aggregate more structured information about the collected search tasks and quantify characteristics, we manually coded the set of search tasks. Analyzing the answers to the survey questions enabled us to transform the natural language descriptions of search tasks into quantitative measures such as the number of search systems used for the task, and the degree of satisfaction with the results obtained. An additional purpose of the manual coding was to remove under-specified search tasks, such as "search for systematic review" or "Trying to find papers and citations".

For defining the coding scheme, we followed a grounded theory approach: we based our coding scheme on existing schemes for search task coding and relevance criteria (see Section 2), making adaptations to the categories to fit our data. For example, for topic domain we include the domains that occur in our data.

Our coding process was as follows: The first round of coding on a small sample of the topics was done by the first author. Based on the findings the initial coding scheme was defined. The second round of coding on a small sample of the topics was done by one of the other authors, using the initial coding scheme. The differences between the codings of the two coders were discussed and the coding scheme was revised where needed. The third round of coding was done by all authors, on the complete set. Each search task was coded by exactly three coders. The resulting coding scheme is summarized in Figure 2.[2]

---

[2]The actual coding scheme included more elaborate explanations for each dimension, and will be shared online (URL excluded for anonymous peer review).

Table 1: Agreement statistics on the manually coded task characteristics. Underlined κ values indicate substantial or near-perfect agreement ($\kappa >= 0.6$). Italic κ values indicate moderate agreement ($0.4 <= \kappa < 0.6$).

Variable	Absolute agreement	Cohen's κ
Topic domain	84%	0.70
Number of search systems mentioned	91%	0.89
Satisfaction score	88%	0.82
Search for self or other	71%	0.56

3.3 Merging the annotated sets

The codings were merged into one set of coded search tasks. If at least one of the three coders voted to exclude the topic from the data (first question in the coding scheme), the topic was excluded. Codings for the remaining topics were combined as follows: (1) If at least two coders agree on the value for the dimension, that value was assigned (all disagreements for the – binary – relevance criteria were solved this way); (2) if all three coders selected a different value, we assigned the median in case of numeric values, or took coder 1's (the first author's) label in case of nominal values. The latter happened only for one item in our data.

4 RESULTS

71 respondents submitted a search task, out of which 15 search tasks were removed because they were not sufficiently specific. Thus, the resulting collection of professional search tasks, analysed below, contains 56 topics.

4.1 Statistics on the manual coding

We measured the inter-rater agreement for each dimension in the coding scheme using Cohen's κ on the three pairs of annotators. We report mean κ scores over the pairs. In the case of numeric variables we computed weighted κ, in which exact agreement is counted as 1, a difference of 1 is counted as 2/3, a difference of 2 is counted as 1/3 and a difference larger than 2 is counted as 0. The agreement statistics for the main task characteristics are in Table 1. The κ values indicate that agreement is substantial or near-perfect for three of the four task characteristics. The last one (search for self or other) has a moderate agreement. This characteristic is often difficult to judge from the description by the searcher. The moderate agreement is therefore caused by the occurrence of 'Unknown' as label for this dimension, occurring in 23 out of 56 search tasks.

4.2 Statistics on the respondents

The most represented age group is 46–55 (36%), followed by 36–45 (27%) and 56–65 (20%). 39% of the respondents has over 20 years of professional experience; 34% has between 11 and 20 years of professional experience. The majority of respondents (36) listed Library and Information Science (LIS) as field of expertise, followed by Healthcare/biomedical (20) and Computer science (7).[3]
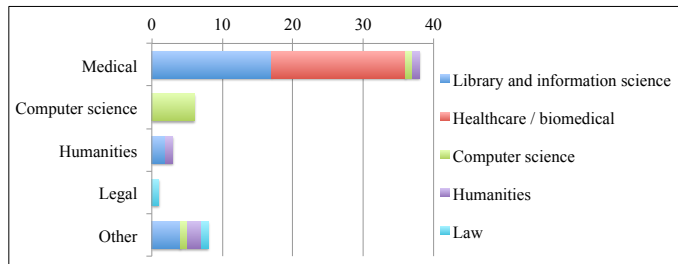

Figure 3: Frequencies of occurrence of topic domains of the search tasks. Each bar represents one topic domain; the colors within the bars represent the first field of expertise listed by the respondent for the search task.

4.3 Statistics on the search tasks

Topic domain. Figure 3 shows the frequencies of topic domains coded for the search tasks. The majority of search tasks take place in the medical domain. A substantial part of those tasks was submitted by a LIS expert. Interestingly, one task was coded as medical while the respondent listed Humanities as field of expertise. This was a task submitted by a Communications adviser who was researching the availability of online information for cancer patients.

Search for self or other. For as few as 11 search tasks, it was explicitly mentioned that the searcher searched for themselves. For 22 search tasks, the respondent explicitly mentioned searching for someone else. In the remaining 23 cases it was not clear to us whether the search task was for the searcher themselves or for an external requester. The respondents who search for an external requester largely listed Library and Information Science (LIS) as their field of expertise. If we assume that all the LIS experts search for others, 17 of the unknowns can additionally be classified as 'Search for other', which makes the majority even larger: 40 out of 56 search tasks were executed for an external requester.

Number of search systems used. The median number of search systems/databases used to complete one search task was 3 (mean: 3.1; standard deviation: 2.5). An example response to the question "Describe which actions you took to search" with 3 databases mentioned is: "Used NICE HDAS to search Embase, Medline and PubMed. Used a mixture of thesaurus terms and keywords and combined using Boolean operators".

Satisfaction with the results. The main satisfaction score over the search tasks was 1.1, indicating 'sufficient result'. For 6 search tasks, the respondent was negative about the result (score -1); for 25 search tasks the result was good (score 2).

Relevance criteria. Figure 4 shows the frequencies of relevance criteria mentioned in the search tasks. Two thirds of the relevance descriptions mention the document type. Most of the time this was a (scientific) article. Topical relevance was mentioned in 24 topics, often described in terms of 'aboutness' (e.g. "Articles about patient call, alert, alarm systems that used phones to call nurses"). A relevance criterion that is typical for information specialists is 'external relevance', denoting that someone else than the searcher themselves – the requester of the information – determines the

---

[3]The sum is more than 56 because respondents were allowed to select more than one field.
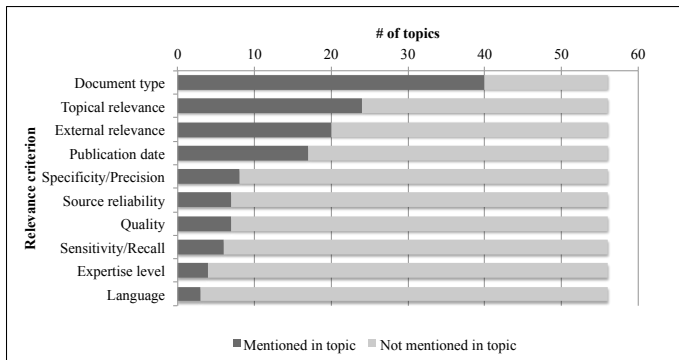
Figure 4: Number of search tasks in which each of the relevance criteria is mentioned by the respondent

relevance of the result. This mostly is evidenced by responses such as "Passed on all relevant papers for expert requester to sift". The fourth most frequently mentioned relevance criterion is publication date, mostly a mention of 'recent papers'.

5 DISCUSSION

From our collection of professional search tasks we can distill the following findings about characteristics of professional search (a)–(c) listed in the introduction:

(a) The search tasks are complex and specific. Complex search tasks are tasks that need multiple steps to be completed, by collecting information from multiple sources [3]. From our data, it is clear that multiple sources were needed to complete the tasks. The median number of search systems/databases mentioned in one search task was 3 and there were search tasks with over 8 search systems required to complete the task. Some of the respondents explicitly show the complexity of their queries; 9 search tasks mention or show the use of Boolean operators. With respect to the specificity of the search tasks, we see that the tasks are clearly domain-specific, and addressing highly specialized topics.

(b) The professional searcher is an expert in the search domain. In the set of responses to our survey this is not necessarily the case. Figure 3 showed that a large portion of the search tasks in the medical domain was completed by LIS experts. The recruitment of respondents explicitly included information specialists and librarians. We did not foresee that this target group would make up the majority of our respondents. This aspect of the survey does not change the other characteristics of the search tasks. In addition, LIS experts are experts in information search, which makes them different from users in the context of enterprise search (see footnote 1), in which any staff member is a potential user.

(c) Professionals have exploratory search needs that require multiple queries and include browsing and analysing multiple documents. As said above, multiple sources are used in many of the search tasks, probably partly caused by the importance of completeness of the result set (systematic review tasks, which are inherently recall-oriented). We do not have quantified results on the exploratory nature of the search tasks, but most of the topics seem well-defined and specific rather than exploratory.

6 CONCLUSIONS

RQ1. To what extent are the characteristics of professional search (a)–(c) reflected by the data acquired in our survey? The results of our survey reflect that the search tasks in professional search are complex and highly specific, but not necessarily exploratory. The results also show that not every professional searcher is an expert in the search domain; they can also be LIS experts.

RQ2. Are these characteristics sufficiently pronounced to justify treating professional search as a separate search genre? The characteristics that we found confirm the differences between professional search and web search mentioned in the literature. These have implications for the design of professional search systems. First, the finding that many professional searchers search for others means that the searcher may not be in a good position to assess the relevance of results. For that reason it might be a good idea to provide additional information in the interface of the IR system based on possible relevance criteria (e.g. publication date, expertise level) [17], to aid the user in creating the (short)list for the client. Second, the complex information needs of professional searchers, for example in systematic reviews, means that professional searchers are searching for information spread across multiple documents. This means that there is no one particular document that best provides the information, and that there is no clear requirement what document should be ranked first. The user interface of a professional search system should ideally be adapted for this characteristic, and show a result set covering a diverse set of aspects to the information need.

RQ3. Are the needs, goals and behaviours of professional searchers sufficiently homogeneous and consistent to justify viewing 'professional search' as a coherent, single field of enquiry?

Based on the data of this research, it seems that the same characteristics of professional search apply to professional searchers from all the domains that participated in our study. Further research is required before it can be determined whether all groups can be treated as one field of enquiry for these purposes. A prominent finding from our survey is the evidence of multiple sources being used – and that the tools do not support this that well. As a result, many different search engines are used to search the relevant sources.

REFERENCES

[1] Marcia J Bates. 1979. Information search tactics. *Journal of the American Society for information Science* 30, 4 (1979), 205–214.

[2] David Ellis and Merete Haugan. 1997. Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of documentation* 53, 4 (1997), 384–403.

[3] Jiyin He, Marc Bron, and Arjen P de Vries. 2013. Characterizing stages of a multi-session complex search task through direct and indirect query modifications. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval.* ACM, 897–900.

[4] Birger Hjørland. 2015. Classical databases and knowledge organization: A case for boolean retrieval and human decision-making during searches. *Journal of the Association for Information Science and Technology* 66, 8 (2015), 1559–1575.

[5] CHA Koster, NHJ Oostdijk, S Verberne, and EKL D'hondt. 2009. Challenges in Professional Search with PHASAR. In *Proceedings of the Dutch-Belgian Information Retrieval workshop (DIR 2009).* Enschede, Nederland:[sn], 101–102.

[6] Udo Kruschwitz and Charlie Hull. 2017. Searching the Enterprise. *Foundations and Trends in Information Retrieval* 11, 1 (2017), 1–142.

[7] David Lewis. 2018. We Don't Know S***** (Search). In *Proceedings of the First Biennial Conference on Design of Experimental Search & Information Retrieval Systems, Bertinoro, Italy, August 28-31, 2018. (CEUR Workshop Proceedings)*, Omar Alonso and Gianmaria Silvello (Eds.), Vol. 2167. CEUR-WS.org, 107. http://ceur-ws.org/Vol-2167/short10.pdf

[8] Marianne Lykke, Birger Larsen, Haakon Lund, and Peter Ingwersen. 2010. De-
     veloping a test collection for the evaluation of integrated search. In *European
     Conference on Information Retrieval*. Springer, 627–630.
[9] Stephann Makri, Ann Blandford, and Anna L Cox. 2008. Investigating the
     information-seeking behaviour of academic lawyers: From Ellisfis model to
     design. *Information Processing & Management* 44, 2 (2008), 613–634.
[10] Dean Mason. 2006. Legal Information Retrieval Study fi?! Lexis Professional and
     Westlaw UK. *Legal Information Management* 6, 4 (2006), 246. https://doi.org/10.
     1017/S1472669606000831
[11] Peter Morville and Jeffery Callender. 2010. *Search patterns: design for discovery.* "
     O'Reilly Media, Inc.".
[12] Soo Young Rieh and Nicholas J Belkin. 1998. Understanding judgment of infor-
     mation quality and cognitive authority in the WWW. In *Proceedings of the 61st
     annual meeting of the american society for information science*, Vol. 35. 279–289.
[13] Tony Russell-Rose, Jon Chamberlain, and Leif Azzopardi. 2018. Information
     retrieval in the workplace: A comparison of professional search practices. *Infor-
     mation Processing & Management* 54, 6 (2018), 1042–1057.
[14] Tony Russell-Rose, Joe Lamantia, and Mark Burrell. 2011. A Taxonomy of
     Enterprise Search.. In *Proceedings of the 1st European Workshop on Human-
     Computer Interaction and Information Retrieval*. 15–18.
[15] Reijo Savolainen and Jarkko Kari. 2006. User-defined relevance criteria in web
     searching. *Journal of Documentation* 62, 6 (2006), 685–707.
[16] Suzan Verberne, Jiyin He, Udo Kruschwitz, Birger Larsen, Tony Russell-Rose,
     and Arjen P. de Vries. 2018. First International Workshop on Professional Search
     (ProfS2018). In *The 41st International ACM SIGIR Conference on Research & De-
     velopment in Information Retrieval - SIGIR '18 (SIGIR '18)*. ACM, New York, NY,
     USA, 1431–1434. https://doi.org/10.1145/3209978.3210198
[17] Gineke Wiggers, Suzan Verberne, and Gerrit-Jan Zwenne. 2018. Exploration of
     Intrinsic Relevance Judgments by Legal Professionals in Information Retrieval
     Systems. In *Proceedings of the 17th Dutch-Belgian Information Retrieval Workshop*.
     5–8.