An investigation of the search behaviour associated with Ingwersen's three types of information needs

Borlund, Pia; Dreier, Sabine

Published in:
Information Processing & Management

DOI:
[10.1016/j.ipm.2014.03.001](https://doi.org/10.1016/j.ipm.2014.03.001)

Publication date:
2014

Document version
Early version, also known as pre-print

Citation for published version (APA):
Borlund, P., & Dreier, S. (2014). An investigation of the search behaviour associated with Ingwersen's three types of information needs. *Information Processing & Management*, 50(4), 493-507.
<https://doi.org/10.1016/j.ipm.2014.03.001>

An investigation of the search behaviour associated with Ingwersen's three types of information needs

Pia Borlund ^{a,*}, Sabine Dreier ^b

^a Royal School of Library and Information Science, University of Copenhagen, Fredrik Bajers Vej 7K, DK-9220 Aalborg East, Denmark

^b Aalborg University Library, Aalborg University, Langagervej 2, DK-9220 Aalborg East, Denmark

ARTICLE INFO

Article history:

Received 1 July 2012

Received in revised form 3 March 2014

Accepted 8 March 2014

Keywords:

Information needs

User study

Interactive information retrieval

Internet searching

Ordinary users

Everyday-life information seeking

ABSTRACT

We report a naturalistic interactive information retrieval (IIR) study of 18 ordinary users in the age of 20–25 who carry out everyday-life information seeking (ELIS) on the Internet with respect to the three types of information needs identified by Ingwersen (1986): the verificative information need (VIN), the conscious topical information need (CIN), and the muddled topical information need (MIN). The searches took place in the private homes of the users in order to ensure as realistic searching as possible. Ingwersen (1996) associates a given search behaviour to each of the three types of information needs, which are analytically deduced, but not yet empirically tested. Thus the objective of the study is to investigate whether empirical data does, or does not, conform to the predictions derived from the three types of information needs. The main conclusion is that the analytically deduced information search behaviour characteristics by Ingwersen are positively corroborated for this group of test participants who search the Internet as part of ELIS.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In 1980 Ingwersen and Kaae (1980) report a study of user-librarian dialogues observed in the public library. Later Ingwersen (1986, p. 223) builds upon that study when he labels three types of user information needs: the verificative information need (VIN), the conscious topical information need (CIN), and the muddled topical information need (MIN). Later again Ingwersen transfers these three types of information needs into the context of online information searching and adds characteristics of information searching behaviour to each of the three types of information needs (Ingwersen, 1996, pp. 14–16; Ingwersen, 2000, pp. 163–164; Ingwersen & Järvelin, 2005, pp. 291–292). The characteristics are intellectually deduced by Ingwersen based on search experience and the nature of the information needs. The behavioural characteristics are as such not investigated empirically in the context of online searching, which is what the present study aims at. A study of the characteristics of user search behaviour with respect to the three types of information needs is important in its own right in order to either corroborate or falsify the associated characteristics outlined by Ingwersen. In addition, the study will provide insight, no matter the outcome of the study, regarding how the actual user search behaviour of the three types of information needs are characterised. An insight that is most useful, for example, for system design in terms of how best to support the user's searching of the different types of information needs. Given the search behaviour of the needs differ, then different search supports may be desirable. The results of the study are also useful in relation to evaluation of information retrieval (IR)

* Corresponding author. Tel.: +45 98 15 79 22.

E-mail addresses: sjc900@iva.ku.dk (P. Borlund), sd@aub.aau.dk (S. Dreier).

systems and user search behaviour by use of simulated work task situations hence reflecting different types of information needs. So far the test instrument of a simulated work task situation is validated only with respect to exploratory searching, which corresponds to a mix of the expected search behaviour of the CIN and the MIN (e.g., [Borlund, 2000, 2003](#)). We do, however, see a demand within the interactive information retrieval (IIR) evaluation literature for evaluation instruments that reflect the three types of information needs identified and labelled by Ingwersen (e.g., [Bell & Ruthven, 2004](#); [Toms et al., 2007](#); [White, Bilenko, & Cucerzan, 2008](#); [White, Jose, & Ruthven, 2001, 2003](#); [White, Ruthven, & Jose, 2002](#)). An investigation of the search behaviour of the three types of information needs would serve as an initial basis for how to further extend and refine the requirements for the employment of simulated work task situations with respect to the handling of the various types of information needs. Further, the demand for representation of different types of information needs noticed in the aforementioned IIR evaluation literature may be seen as a confirmation of the types of information needs identified by Ingwersen. In the present work we do not question the three types of information needs, but accept them as presented by Ingwersen.

The present paper reports the investigation of information search behaviour of users who search the Internet with the three types of information needs. The purpose is to gain insight about the characteristics of the information search behaviour of the three types of information needs in order to compare those behavioural characteristics to the corresponding descriptions by Ingwersen. Hence an empirical IIR study is carried out by use of 18 ordinary users, in the age range of 20–25, who searched three personal, self-prepared information needs corresponding to the three types of information needs: the VIN, the CIN, and the MIN. The users formulated the information needs themselves to ensure the information needs were relevant, of interest, and reflected their everyday-life information seeking needs ([Savolainen, 1995](#)). The study aims at answering the following research question:

- How does the test participants' information search behaviour correspond to the predicted search behaviour associated with the three types of information needs characterised by Ingwersen?

The remainder of the paper is structured as follows: in Section 2, we introduce in detail the three types of information needs including the associated search behaviour as described by Ingwersen (e.g., [2000](#)). However, the starting point of Section 2 is a brief introduction to how the concept of information needs is perceived within the cognitive viewpoint, which is the epistemological standpoint of Ingwersen (e.g., [Ingwersen, 1996](#)). In Section 3, we present the methodology of the study, including the test participants' search characteristics, a description of the study design and data collection methods, the overall test procedure, and the methods used for data analysis. The research question is answered in Section 4 with the presentation and comparison of the search characteristics. The paper closes with Section 5 that presents a summary and implications for future research.

2. The cognitive viewpoint and the concept of an information need

The cognitive viewpoint is user-centred and acknowledges the user's personal perception of an information need and the subjective relevance assessments of information in response to that information need ([Borlund, 2010](#), pp. 23–24). The cognitive viewpoint within IR originated as an alternative to the dominating system-driven approach to IR systems evaluation signified by the lab-based Cranfield model ([Cleverdon & Keen, 1966](#); [Cleverdon, Mills, & Keen, 1966](#)). As such the cognitive viewpoint is concerned with the concept of an information need and its formation process as perceived and acted upon by the user. At a more abstract level the information need formation process is referred to as the changes or transformations of the recipient's knowledge structures by the act of communication and the processes of perception, evaluation, interpretation, and learning ([Ingwersen, 1992](#)). The concept of an information need is defined as the outcome of a change in the state of knowledge of the user which results in an anomalous state of knowledge (ASK) ([Belkin, 1980](#); [Belkin, Oddy, & Brooks, 1982](#)). The change that results in an ASK, which is a cognitive development internal to the user, is happening due to an external situation, e.g., a given work task situation. In other words, an external situation causes a change in the knowledge state and in the user's knowledge structure, which results in an ASK. An ASK can be described as the user's recognition of an insufficient knowledge model which results in an information need, for instance, in order to reduce uncertainty. As the result of the impact of further externally added information, e.g., retrieved information, the information need may change or develop over time in order to satisfy the present problem situation as perceived by the user. This means that the concept of an information need, within the cognitive viewpoint, is understood as a potentially dynamic and developing concept, as also indicated by the cognitive revolution presented by [Robertson and Hancock-Beaulieu \(1992\)](#). Basically, an information need is born out of a situation, and may develop during the process of reaching the requirements of that situation. The user's perception of an information need is thus triggered by the perception and interpretation of a given situation, a problem to be solved or a state of interest to be fulfilled, under influence of the user's current cognitive and emotional state. This state is affected by the cultural and social context in which the user acts. In essence, the cognitive viewpoint is about the user's desire for information, and hence a frame for authentic information studies of users' retrieval, search and seeking interactions in the process of achieving the desired information ([Borlund, 2010](#), p. 24).

Having touched upon the cognitive viewpoint to IIR and how the concept of an information need is viewed it is time to take a closer look at the three types of information needs identified by Ingwersen. Thus the purpose of the following section

first and foremost is to operationalise the three types of information needs in relation to the empirical study. This is done through a description of the characteristics of each type of information need and the associated search behaviour as presented by Ingwersen (2000).

2.1. The three types of information needs

Ingwersen defines the three types of information needs (the verificative information need (VIN), the conscious topical information need (CIN), and the muddled topical information need (MIN)) according to the dimensions of well-/ill-defined and stable/variable as depicted in Table 1 (Ingwersen, 2000, p. 163). The dimensions of well-defined and ill-defined relate to how well the information need is defined in the user's consciousness at a given point in time. The dimensions of stable and variable concern the potential dynamism of the information need with reference to whether it is stable or variable in the process of satisfying the information need.

The attentive reader will notice that the well-defined CIN and the ill-defined MIN are represented twice in the matrix, which illustrate the variability or stability of the information needs, respectively. Ingwersen explains the two cases of the CINs as follows: "This type is assumed to be either of intrinsically stable nature, like the verificative one, or variable over session time" (Ingwersen, 2000, p. 164). Similarly, in common to both versions of the MIN is that they by nature are ill-defined in terms of the user's prior knowledge and perception of what information is required to satisfy the information need. According to Ingwersen (2000, p. 163) the two versions of the MIN differ with respect to how well-perceived the underlying task of the MIN is to the user. If the user's understanding of the task is muddled, then the MIN is stable; and if the underlying task is more clearly defined though still resulting in a MIN, then the information need is variable. The point is, that the user may operate within one of the illustrated types of information needs based on individual interpretation of a given work task (Ingwersen, 2000, p. 164). The VIN is described in one cell of the matrix, only, with respect to the dimensions of well-defined and stable. Despite the four-cell matrix in Table 1 Ingwersen maintains the classification of the three types of information needs, and therefore so do we. Ingwersen describes the three types of information needs as follows.

2.1.1. Verificative information need (VIN)

A user with a VIN wants to verify information objects with known non-topical (structured) data, such as author names, client address, cited authors, journal name, and facts (Ingwersen, 2000, p. 164). A VIN need is characterised by being well-defined and stable. The user's cognitive state is rich and the user possesses a strong and narrow intentionality and furthermore a low degree of curiosity (Ingwersen, 1996, p. 20). The information search behaviour of this type of information need is characterised as *querying* and *filtering* as depicted in Table 1. Due to the well-defined cognitive state, the user is expected to be less uncertain and capable of query modification and at filtering out irrelevant information (Ingwersen, 2000, p. 165). This signifies a search behaviour characterised by focused searching carried out in a relatively short time (Ingwersen & Järvelin, 2005, p. 287).

2.1.2. Conscious topical information need (CIN)

For this type of information need the user wants to clarify, review, or pursue information in a known subject matter and domain, where known subject matter signifies topical (unstructured) data about contents, such as terms, concepts, and image representation (Ingwersen, 2000, p. 164). A CIN is characterised as being well-defined, but at the same time to be of more variable nature. Consequently, the user will navigate and engage in more exploratory search behaviour and feel stages of uncertainty during the information searching (Ingwersen, 2000, p. 165). The information search behaviour of a CIN is described as *querying*, *navigation*, and *dynamic interaction* as shown in Table 1. According to Ingwersen (2000, p. 165) the user is expected to be willing (or forced) to learn and shift focus throughout the search session and the behaviour is associated with the search strategy 'berrypicking' as suggested by Bates (1989).

Table 1

Matrix of four distinct cases of human intrinsic information needs, given a perceived work task situation, and the corresponding seeking and interactive information retrieval behaviour (Ingwersen, 2000, p. 163).

Intrinsic info. need variables – given a perceived work task	Well-defined	Ill-defined
Stable	Verificative information need Conscious topical information need <i>Querying</i> <i>Filtering behaviour</i>	Muddled work task & Muddled topical information need <i>Search loops</i>
Variable	Conscious topical information need <i>Query-Navigation</i> <i>Dynamic interaction</i>	Defined work task & Muddled topical information need <i>Browsing</i> <i>Trial & error behaviour</i>

2.1.3. Muddled topical information needs (MIN)

A user with a MIN is engaged in the exploration of new concepts and relations outside known subject matter or domain (Ingwersen, 2000, p. 164). The information need is characterised by being poorly defined in the user's consciousness which results in high cognitive uncertainty (Ingwersen, 2000, p. 165). The associated search behaviour for this type of need is defined as *search loops*, *browsing*, and *trial & error behaviour* as described in Table 1. This is due to the user's cognitive uncertainty and specifically new adequate search features may be hard to recall from memory or be non-existent (Ingwersen, 2000, p. 165). Further, the user's motivation and curiosity is high and may constitute the progress of the information searching (Ingwersen, 2000, p. 165).

One final note should be made before this presentation of the three types of information needs ends. A recent tendency within the research area of information searching is to distinguish among different types of VINs, e.g., fact-finding tasks and known-item tasks (e.g., Wildemuth, Freund, & Toms, 2013). Since Ingwersen does not distinguish among different types of VINs neither do we. The following section, the methodology part of the paper, presents how the empirical study is designed and carried out.

3. Methodology

The present study is designed within the frame of the cognitive viewpoint that acknowledges the user's personal perception of the information need, and how the context that surrounds the user may impact the user in the process of searching (Ingwersen, 1992). Thus the study is designed to fulfil the following requirements:

- Involvement of real users.
- Use of real information needs.
- Focus on information searching interaction.
- Natural test situation.

The involvement of real users is a fundamental aspect of the overall research design. The presence of real users is creating a more realistic test situation simply because it is more consistent with real-life information searching. (Please note, that as soon as the users take part in the study, we refer to them as test participants.) Furthermore the use of real information needs allows the test participants to search for information of relevance and interest to them and ensure realistic search interaction and search behaviour (Ingwersen & Järvelin, 2005, p. 7). The use of the test participants' own, personal information needs also ensures a greater extent of reality to the test situation hence making it a more natural situation. So does the fact that the study is carried out in the test participants' private homes. As such the study is designed as a naturalistic study (Crystal & Wildemuth, 2009, pp. 62–72).

The remainder of this section presents the study, the test participants, and describes how they were recruited, and how they were instructed to formulate three personal information needs. The section continues with a description of the test design, the data collection methods used, and the test procedure according to the protocol. Hereafter the collected data is presented, and the methods used for data analysis are introduced.

3.1. The study

The study was carried out spring 2011 from March 20 to April 6 and consisted of a pilot study and a main study. A total of 21 test participants were recruited of which three participated in the pilot study and the remaining 18 test participants participated in the (main) study.

3.2. Test participants and recruitment

The 18 test participants were young people in the age range of 20–25 years. The sample of test participants consists of 13 females and five males, and the average age is 23.2 years. The test participants were employed in different types of professions and jobs. The majority of the test participants (11 test participants) worked as physiotherapist, planner, administrative assistant, medical technician, and warehouse operative. Three test participants studied for higher degrees, other two were receiving training as sales assistant and medical secretary. The final two test participants were self-employed in the service trade. The test participants were all Danish, and lived in the northern part of Jutland, Denmark. They assessed their searching skills as 'good' and 'very good', and they used the Internet for information searching on a regular basis.

The sample of test participants constitutes a purposive sample that is selected according to the criterion of accessibility (Kelly, 2009, p. 67). As previously mentioned, an important element of the research design was to conduct the study in a naturalistic environment. Consequently, it was decided to recruit test participants through personal relations, in order to get the test participants to open their homes for the study. The personal relations are characterised as semi-personal relations as the test participants were not selected within the investigators' close circle of acquaintances. Due to the semi-personal relations a formal contact was maintained with the test participants during recruitment and testing. The test participants were recruited in the following way:

- (1) A group of 21 potential test participants were identified. The criteria for participation were that the test participants were young people and hereby expected ELIS Internet users, and were able to read and write Danish so that they could take part in the study. They were not to be Library and Information Science or Computer Science students (or belonging to related fields of study) since this could affect the search behaviour. In addition, they were not to be of close acquaintances to the investigators.
- (2) A physical letter was sent to the 21 potential test participants. The letter contained a brief introduction to the study and an invitation to participate. Further, the letter informed that they within few days would receive a phone call that would follow-up on the invitation letter.
- (3) Few days later phone calls were made and the potential test participants were asked whether they wished to participate in the study. With no exception they all agreed to participate in the study. Hereafter, data and time for the visit to their home was agreed upon.
- (4) To confirm the agreed visit a new physical letter was sent to the test participants. In addition, the letter contained few simple instructions about the three information needs they were to formulate as part of their participation. The test participants were instructed to write down the information needs on three separate pieces of paper that were enclosed the letter.

3.2.1. Formulation of information needs

The test participants' formulation of information needs is essential because the objective of the study is to compare actual information search behaviour characteristics to the corresponding descriptions by Ingwersen. For that reason it is crucial that the three different types of information needs are present in the study.

The test participants were encouraged to select topics that they would search for anyway, and hence may as well use the opportunity of the study to do the information searching. The intention was to motivate the test participants to formulate information needs that were of personal interest, relevance, and useful in relation to their everyday-lives, e.g., job, study or leisure time. In order to ensure that the test participants were able to formulate information needs that represent the types of the VIN, the CIN, and the MIN, basic instructions with simple criteria to be met were given. The criteria derive from the theoretical assumptions about the different types of information needs and their individual characteristics, as described in Section 2.1. Table 2 depicts the criteria to be met for each of the three types of information needs. The test participants were not informed that these directions correspond to three types of information needs, but only that they were to formulate three information needs that met these criteria. To them the purpose of the study was about how well the Internet is at satisfying information needs.

No examples of the three information needs were given, in order not to dictate or influence the test participants' information needs. To obtain a full understanding of the test participants' information needs they were asked to describe: (1) what they were going to search information about; (2) why they were interested in the topic, and (3) what the retrieved information was going to be used for. Examples of the formulated information needs are: when and where LOC [artist] plays?, what public school is closest to where we live? (VINs), heart diseases, Cleopatra (CINs), and ecological milk production, and body combat (MINs). The CIN and the MIN are quite similar by both being topically focused. They differ in terms of the test participants' prior knowledge about the topic in question. The information about prior knowledge was conveyed orally during searching and via the pre-search questionnaire. This information is important, as it became apparent that we had to check whether the formulated information needs were of the intended type, and re-categorise accordingly. It was the intention that the test participants were to formulate three different information needs representing the three types of information needs. But not all managed, though they thought they did, and we did not tell them otherwise. The test participants searched whatever they had prepared, and later we re-categorised the information needs. The re-categorisation was checked for inter-coder reliability. 11 out of 18 test participants formulated information needs that represented all three types of information needs. This means that the remaining seven test participants formulated two information needs of the same type. The types of information needs searched in the study are distributed as follows: 20 VINs, 17 CINs, and 17 MINs. The unequal distribution of types of information needs is not problematic in relation to the objective of the study, as long as the three types of information needs are represented.

3.3. Study design and data collection methods

Method triangulation is used in the design of the study, and hence a combination of questionnaire, transaction logging, and interview form the basis of the employed data collection methods. The methods are employed in the following order:

Table 2
Criteria to be met in the formulation of the three types of information needs.

VIN	You would like to check a specific piece of information
CIN	You would like to find information about a topic you are familiar with
MIN	You would like to explore an unknown topic

- Pre-search questionnaire.
- Transaction logging of the information searching (video logging of screen interaction).
- Structured post-search interview.

The methods are chosen because of their ability to gather varied and complementary data of both qualitative and quantitative nature. The objective of the pre-search questionnaire is to collect data about the test participant's demographics, their self-perceived experience with searching the on Internet, and their perception of the three searches they carry out. Data about demographics and search experience are collected in order to be able to describe the test participants and to explain possible cases of unexpected search behaviour.

The test participants' interaction with the IR system is collected via transaction logging by use of Morae, version 3.2.1 (www.techsmith.com). From the video logs both qualitative and quantitative results are derived. The qualitative results are observations of the interactive search behaviour of the searching of the three types of information needs. The quantitative result take the form of classical IIR search characteristics such as: Number of unique search terms, number of search iterations, number of search terms per search iterations, number of visited web pages, number of visited web sites, number of web pages visited per web site, number of relevant web pages (that is, web pages saved as 'favourites' in the Internet browser), and time spent searching. Please note, we distinguish between websites and web pages. We define a website as a specific domain, where each website contains a number of individual web pages. Hence it becomes possible not only to analyse the activity according to visited websites, but also according to visited web pages per website.

The post-search interview is conducted as a structured interview in order to ensure consistency in the handling of the interviews (Bryman, 2008, p. 193). The structured post-search interview is used to collect information about the test participants' perception of their information searching. This type of data is of qualitative and subjective nature.

3.4. Test procedure

A detailed protocol was followed throughout the data collection (Kelly, 2009, p. 58). Before the data collection was started the test laptop was setup and placed where the test participant would usually perform the information searching. The test participants were informed about the overall study procedure and thereafter they started with filling out the questionnaire. After this the information searching began. The Internet start page was set to www.iva.dk (the homepage of the Royal School of Library and Information Science at the time), and the Internet browser was reset after searching. The searches were carried out with no time constraints, as the purpose of the study is to look at the test participants' realistic search behaviour. The test participants were further instructed to retrieve as many documents as it would take to satisfy the information need, and hence the searching stopped at the initiative of the test participants in line with the reported studies by, e.g., Borlund (2000, pp. 81–82) and Tombros, Ruthven, and Jose (2005, p. 6). Furthermore, the test participants were encouraged to search in any way they felt appropriate and natural. The only restriction was that the test participants were not allowed to consult the investigator about how to carry out the information searching. This restriction was implemented in order to ensure natural and unaffected information behaviour. Further, to make the information searching as realistic as possible the test participants were given the opportunity to save retrieved relevant information. Five test participants used this opportunity and the saved information was emailed to them. To neutralise any learning effects caused by increased familiarity with the test situation the searching of the three types of information needs were rotated between the test participants, so that fewest test participants performed the same type of information needs in identical order. The study closed with the post-search interview and the test participants were thanked for their participation with a small present worth the value of 100 Danish kroner (approx. 11 British Pounds Sterling). The test participants did not know that they would receive this token since it should not affect their participation. In all, the study took on average approximately 2 h per test participant.

3.5. Methods for data analysis

The various collected data calls for different methods for data analysis. The sub-sections below describe how the analysis of the qualitative data is inspired by Grounded Theory (Corbin & Strauss, 2008), and how the quantitative data is treated with reference to descriptive statistics.

3.5.1. Qualitative data analysis

The purpose of the qualitative analysis is to provide detailed descriptions of the test participants' information search behaviour. The analysis is carried out as a manual examination of the 54 video log files containing the information search behaviour of the 18 test participants. The analytic approach is inspired by Grounded Theory (Corbin & Strauss, 2008, pp. 51–52), and can be described as deriving concepts or themes from raw data. This approach provides a way of grouping and organising data, and is relevant to apply in relation to identifying characteristics of the test participants' information search behaviour.

The log files are recorded in a way that makes it possible to play them like a video from which the test participants' exact information search behaviour on screen can be studied. The specific information behavioural patterns are analysed at level of search sessions with respect to the three different types of information needs. More precisely the analysis is carried out by writing down characteristics of the information search behaviour, and hereafter by identifying possible patterns based on

the different test participants when information searching of the same type of information need. To facilitate the analytic process the analytical tools of asking questions and compare as recommended by [Corbin and Strauss \(2008, pp. 65–86\)](#) are employed. It is essential to ask questions and to compare when one wants to obtain an understanding of the test participants' information search behaviour with reference to characteristics, similarities, and differences, and wants to identify possible relationships. By asking questions one achieves a more exploratory approach to the data material and is also (at the same time) likely to uncover and identify connections and relationships otherwise not that obvious. In order to structure the analysis of the log files the following questions are asked:

1. How goal-oriented is the information searching?
2. How hard does the test participant find it to express search terms?
3. How much time does the test participant spend on viewing the result list?
4. How exploratory is the information searching?
5. Does the test participant leave the visited web pages quickly?
6. Does the test participant look at the second page (and potentially the following pages) of the result list?

The questions are partly identified on the basis of characteristics of the different types of information needs described by Ingwersen (e.g., see Section 2.1). The information search behaviour of the video log files is analysed in a systematic manner according to the six questions listed above. The questions 1–4 organise the observed search behaviour according to the categories low, medium, and high depending on how often the given search activity in question takes place. Question no. 1 concerns how focused the information searching is. If the information searching is easily distracted, for example by clicking on hyperlinks that are not directly related to the topic, it is assessed as 'low'. Question no. 2 considers the test participant's effort both measured by time used on formulation of search terms and how much the terms are reformulated. Question no. 3 concerns the time spent on viewing the result list. The exploratory searching of question no. 4 is defined as 'curious' and 'open'. If a test participant follows various and different paths throughout the search for instance by searching on new aspects of the topic in question and clicks on many items of interest, the behaviour is assessed as 'high'. The questions no. 5 and 6 can be answered by a simple "yes" or "no". Question no. 5 considers whether the test participant leaves the visited web pages short after entering, or if the test participant stays on the website for a longer period of time. Question no. 6 concerns whether the test participant moves to the second page of the result list (and the potentially following pages).

The 54 video log files are analysed one by one according to the six questions. The analysis of the information search behaviour is not based on particular actions, but relies on a general impression of all actions included in the information search session. The purpose is to analyse whether these characteristics are identifiable in the test participants' information search behaviour. In addition, the questions allow for the identification of other behavioural pattern(s) that may contribute to an understanding of how the search behaviour of each type of information need can be characterised and differentiated (for example leave web pages quickly, and look at 2nd page of the result list). Overall, the qualitative analysis is regarded as a process involving studying and examining data in relation to derive meaning and understanding ([Corbin & Strauss, 2008, p. 1](#)), and hence it is a dynamic and non-linear process.

The interviews are transcribed and analysed in terms of creating categories of 'meaningful meaning' about the test participants' perception of their information searching ([Kvale, 2007](#)).

3.5.2. Quantitative data analysis

Exact numbers and descriptive statistics constitute the main statistical method used given the purposive sample (non-probability sample) of the 18 test participants. All search attributes derived from the video logs are systematically integrated into tables. In one case a box plot is added to illustrate the time spent searching of the three types of information needs.

As said, we apply a purposive sample of 18 test participants in the study. Our purposive sample is therefore not a probability sample and the test participants were not randomly assigned to treatment and control groups. Despite widespread practice, it is meaningless to use statistical significance tests when the crucial assumption of randomness is missing (probability sampling and/or random assignment) (e.g., [Carver, 1978](#); [Kirk, 1996](#); [Shaver, 1993](#)). Statistical significance tests address only sampling error and it is not possible to calculate the likelihood of a sample result being due to sampling error if randomness is not present ([Cohen, 1994](#)). Further, if the assumption of randomness was fulfilled in our study, a sample size of 18 would imply that the study was under-powered and thus not able to detect statistical significance except for very high effect sizes. In other words, the study would be highly vulnerable to Type II errors, i.e., failing to reject an incorrect null hypothesis. Consequently, inferential statistics are not viable or relevant in the present study as this obviously has implications for the external validity of the results, i.e., the statistical generalisation of the results to the population.

4. Results

We start this section by presenting the qualitative analysis of the logged video files (Section 4.1). Hereafter the search characteristics based on the quantitative, measurable aspects of the searching from a classic IIR perspective (Section 4.2) are presented. The presentation of the results constitutes the answering of the research question: How does the test participants' information search behaviour correspond to the predicted search behaviour associated with the three types of information needs characterised by Ingwersen?

4.1. Search characteristics – qualitative analysis

The results from the qualitative data analysis provide a detailed insight into the test participants' information search behaviour while searching the three types of information needs. We begin by presenting the analysis of the VIN, followed by the CIN, and then the MIN. This section closes with a recap and discussion of the qualitative analyses with respect to the search characteristics predicted by Ingwersen (Section 4.1.1).

The analysis of the video log files are carried out according to the previously six listed questions. A brief, simple summary of the results are depicted in Table 3. A more detailed presentation of the results is presented below.

VIN: From the analyses of the video log files it is clear how the information searching of the VIN is the most focused of the three types of information needs (e.g. see, Table 3, question 1). The logs show how the test participants look for a specific piece of information and seem to have a pretty clear idea about which relevant sources of information to use. They are able to express quite many search terms describing their information need. They find it relatively easy to formulate the search terms (e.g. see, Table 3, question 2), and they use relatively few search iterations to retrieve the relevant information. The fact that the test participants are highly focused is illustrated by a curious search behaviour, where they do not click on links which are not directly related to the information need they are searching. The VIN is the least exploratory searched information need (e.g. see, Table 3, question 4). Compared to the CIN and the MIN this is the information need where most test participants quickly leave the visited web pages (e.g. see, Table 3, question 5). It is also characteristic for the search behaviour of the VIN that the test participants spend the shortest amount of time on viewing the result list (e.g. see, Table 3, question 3). None of the test participants look at the second page of the result list (e.g. see, Table 3, question 6). If they did not retrieve a relevant web page at once, they clicked few times on hyperlinks within the given domain to look for relevant information. If they still did not find the relevant information, they tended not to browse the domain, but did rather return to Google where they entered a new query. Another characteristic of the search behaviour of the VIN is that the test participants stopped as soon as relevant information was retrieved, and they did not pursue additional information out of curiosity.

CIN: The video log files show how the information search behaviour of the CIN is less focused than the VIN, and at the same time is more focused than the MIN (e.g. see, Table 3, question 1). The test participants spend more time on searching the CIN compared to the focused searching of the VIN. They also spend more time on the individual visited websites and they interact more with the domains compared to the VIN. From the video log files it appears that they have no problems at expressing and formulating the CIN, in that, they quickly formulate the queries.

The test participants use neither long time nor short time on viewing the result list (e.g. see, Table 3, question 3). They spend longer time viewing the result list for the CIN compared to the VIN, while shorter time compared to the MIN. In two cases do the test participants continue their relevance assessment to the second result page (e.g. see, Table 3, question 6). In general, the searching of the conscious topical information is more exploratory compared to the VIN and less compared to the MIN (e.g. see, Table 3, question 4). In more than half cases did the test participants not leave the visited web pages quickly (e.g. see, Table 3, question 5).

MIN: The analysis of the video log files of the searching of the MIN shows that the search behaviour for this type of information need is the least focused compared to the other two information needs (e.g. see, Table 3, question 1). This is primarily seen by how the test participants demonstrate a more 'open' and exploring information search behaviour and that they are not immediately able to assess whether retrieved information is relevant or irrelevant. With this particular type of information need the test participants want to obtain knowledge about a topic of a wider and more unknown nature than are the cases with the VIN and the CIN. From the video log files it is visible how the test participants find it harder to identify relevant information, both in relation to which sources to use and in relation to where the relevant information can be localised within a given domain. The searching of this type of information need presents no examples of the visited domain being left quickly (e.g. see, Table 3, question 5). The test participants seem carefully to study the visited domains. The test participants find it harder to formulate search terms than was the case with the VIN and the CIN (e.g. see, Table 3, question 2) seen by how they use fewer search terms per query and more query modifications to retrieve relevant information. The query modifications reveal partly how they conduct almost circular searching by almost identical search formulations with minor exchanges of search terms, and partly how they make tentative searches in a quick-and-dirty style of searching. It is characteristic for the MIN that the test participants spend most time on viewing the results list compared to the other two types of information needs (e.g. see, Table 3, question 3). In four cases the test participants looked at the second page

Table 3

Simplified summary presentation of the qualitative analysis according to the six questions.

Qualitative questions	VIN	CIN	MIN
1. How goal-oriented is the information searching?	High	Medium	Low
2. How hard does the test participant find it to express search terms?	Low	High (=MIN)	High (=CIN)
3. How much time does the test participant spend on viewing the result list?	Low	Medium	High
4. How exploratory is the information searching?	Low	Medium	High
5. Does the test participant leave the visited web pages quickly?	High	Medium	Low
6. Does the test participant look at the second page of the result list (and the potentially following pages)?	Low	Medium	High

of the result list, which represents the highest number of cases across the three types of information needs (see Table 3, question 6). The information search behaviour of the MIN is characterised by being the most exploratory compared to the VIN and the CIN (e.g. see, Table 3, question 4). It is furthermore characteristic for this type of information need that the test participants demonstrate a search behaviour that indicates they want to examine the topic in question further. For instance, the test participants #1 and #2 consulted an online dictionary during the information searching with the intent to obtain a greater understanding of concepts. Test participant #3 used *Google translate*, which again indicates that the test participant is engaged in a topic which he finds difficult and needs to investigate further.

The information search behaviour of the MIN is generally characterised by how the test participants acquaint themselves with a topic that they are not familiar with. The test participants read relatively more than they do while searching the other two types of information needs. Since the topic is relatively complex and unfamiliar to the test participants, they cannot confine themselves to skim, but have to read, as well as to read more in order to be able to assess the relevancy of the retrieved information. It is also characteristic that the test participants are interested in several aspects of the topic and not just one specific aspect of the information need.

4.1.1. The qualitative analysis and Ingwersen's predictions

From the video log files we see how the test participants are focused and goal-oriented in their searching of the VIN, or as Ingwersen puts it, they possess a strong and narrow intentionality and furthermore a low degree of curiosity (Ingwersen, 1996, p. 20). The test participants find the VIN easy to search, probably because they know well what they are searching for. It does, however, require few attempts of query modifications. In addition, they spend a considerable amount of time considering how to formulate the queries. The test participants spend the shortest amount of time on viewing the result list, as they are quick at assessing relevancy of the retrieved results. They formulate queries, search, assess, and reformulate. They stick to what they are searching for, and stop as soon as the information need is satisfied. Because they know what they are searching for, and therefore capable at expressing the VIN, they are also efficient at assessing relevance, or conducting information filtering, as Ingwersen puts it. This fits the information search behaviour of *querying* and *filtering* that Ingwersen associates this type of information need (Ingwersen, 2000, p. 165).

With respect to the CIN the log files show how the test participants are less focused than with the searching of the VIN, and they also spend more time on searching this information need. This probably relates to the fact that the test participants are not looking for a specific piece of information, but are interested in gaining knowledge about a much wider topic, which they already have some knowledge about. It is apparent for the CIN that the test participants possess prior knowledge about the topic, and that they search within a known subject area. This is seen by how the test participants are relatively quick at getting an overview of the searched domains and how they skim and read fragments of text in a relatively high speed. They also spend more time on the individual visited domains and they interact more with the domains compared to the VIN. From the log files it does not look like the test participants had any problems at expressing the CIN or formulating search terms that retrieve relevant results. They spend longer time on viewing the result list for the CIN compared to the VIN, while shorter time compared to the MIN. The demonstrated behaviour corresponds to Ingwersen's description of how the user for this type of information need navigate and engage in more exploratory searching and feel stages of uncertainty during the information searching (Ingwersen, 2000, p. 165). The depicted search behaviour for the CIN can be described as *querying*, *navigation*, and *dynamic interaction* as done by Ingwersen (2000, p. 165).

The search behaviour of the MIN is clearly exploratory, and the test participants show less confidence in assessing relevance and they take their time for this activity. This is probably due to curiosity and the fact that they search in a more or less unknown knowledge area. The test participants spend more time viewing the results list either because the test participants do not easily find relevant information or that it requires a greater effort on their behalf. The information search behaviour of the MIN is characterised by being the most exploratory compared to the VIN and the CIN. This is visible, e.g., from the test participants' repeatedly query modifications, the consulting of many websites, and from the test participants' view of the second page of the result list. This matches how Ingwersen (2000, p. 165) presents the MIN as being poorly defined in the consciousness of the user what makes the user experience high cognitive uncertainty. Based on the video log files all the search behaviour characteristics mentioned by Ingwersen are identified, that is, *search loops*, *browsing*, and *trial & error* (Ingwersen, 2000, p. 165). The information searching of the MIN is conducted with care and persistency by the test participants, which confirms Ingwersen's prediction of the user's high motivation and curiosity as the driving force of searching for this type of information need (Ingwersen, 2000, p. 165).

4.2. Search characteristics – quantitative analysis

In general the test participants' information searches can be characterised as *Google* searching. All test participants used *Google* at least once during the information searching. 80% of all the information searches were made via *Google*. This is consistent with results by Aula, Jhaveri, and Käki (2005) who show how 95.3% of 236 test participants report that they use *Google* as their primary search engine. Though information searching via *Google* is the most used approach of searching, direct entering of URLs of websites containing the requested information does also take place, and so does 'classic' bibliographic database search for scientific papers by use of advanced searching. In the following we take a closer look at what characterises the test participants' searching of the three types of information needs. Tables 4–11 report the quantitative search characteristics of the three information needs with respect to the number of unique search terms (Table 4), the

Table 4

The test participants' use of unique search terms in the searching of the three information needs.

Type of information need	Min.	Median	Max.	Mean	Sd.
<i>Number of unique search terms</i>					
VIN (20)	0	4	12	4.8	3.44
CIN (17)	1	5	15	6.2	4.67
MIN (17)	1	7	18	7.3	4.47

Table 5

The test participants' number of search iterations in the searching of the three information needs.

Type of information need	Min.	Median	Max.	Mean	Sd.
<i>Number of search iterations</i>					
VIN (20)	0	2	9	3.3	2.83
CIN (17)	1	3	15	4.6	4.58
MIN (17)	1	5	17	5.8	3.88

Table 6

The test participants' numbers of search terms per search iteration during the searching of the three information needs.

Type of information need	Min.	Median	Max.	Mean	Sd.
<i>Number of search terms per search iteration</i>					
VIN (20)	1.0	2.4	7.0	2.9	1.57
CIN (17)	1.0	2.7	4.6	2.7	1.11
MIN (17)	1.0	2.3	4.0	2.2	0.78

Table 7

The number web pages visited by the test participants during the searching of the three information needs.

Type of information need	Min.	Median	Max.	Mean	Sd.
<i>Number of visited web pages</i>					
Verificative (20)	2	9	43	13.7	12
Conscious topical (17)	9	25	71	30.5	18.2
Muddled topical (17)	9	36	106	47.1	32.12

Table 8

The number of unique websites visited by the test participants during the search of the three information needs.

Type of information need	Min	Median	Max.	Mean	Sd.
<i>Number of websites visited</i>					
VIN (20)	1	3.5	9	3.6	2.04
CIN (17)	1	5	12	5.6	2.98
MIN (17)	2	6	20	7.6	4.70

Table 9

The number of web pages per unique website visited by the test participants during the searching of the three information needs.

Type of information need	Min.	Median	Max.	Mean	Sd.
<i>Number of web pages visited per website</i>					
VIN (20)	1.5	3.4	6.0	3.5	1.50
CIN (17)	3.0	5.3	14.2	5.8	2.71
MIN (17)	3.0	5.3	16.8	6.4	3.22

number of search iterations (Table 5), the number of search terms per search iterations (Table 6), the number of visited web pages (Table 7), the number of visited websites (Table 8), the number of web pages visited per website (Table 9), the number of relevant web pages (Table 10), web pages saved as 'favourites' in the Internet browser), and the time spent searching (Table 11). Table 4 is presented in detail as an exemplary case of how to read the tables. The remaining tables are addressed when particular points are made, if any. Otherwise the tables are pointed to in the summary discussion of the depicted results with respect to Ingwersen's predictions (Section 4.2.1).

Table 10

The number of web pages the test participants marked as relevant with respect to the three information needs.

Type of information need	Min.	Median	Max.	Mean	Sd.
<i>Number of relevant web pages</i>					
VIN (20)	1	2	4	1.9	0.85
CIN (17)	1	3	10	3.8	2.41
MIN (17)	2	4	10	4.6	2.40

Table 11

The time measured in minutes that the test participants spent on searching the three information needs.

Type of information need	Min.	Median	Max.	Mean	Sd.
<i>Time spent in minutes</i>					
VIN (20)	01:25	03:52	15:37	05:36	04:07
CIN (17)	05:37	16:12	28:27	14:18	06:57
MIN (17)	05:02	16:38	28:30	17:17	06:02

The test participants use on average 4.8 unique search terms in the overall searching of the VIN, compared to 6.2 and 7.3 unique search terms for the CIN and MIN (Table 4).

The spread of the data (Table 4) is seen by the standard deviation (Sd) of 3.44 for the VIN. The minimum value of zero informs us that VIN searching also takes place by use of direct entering of URL(s) as no search term is registered. The maximum value of unique search terms for the VIN is 12 unique search terms. The VIN is characterised by being searched by use of fewer unique search terms compared to the searching of the CIN and MIN with average values of 6.2 and 7.3 unique search terms. Table 4 shows a distinct spread of the data with the standard deviation of 4.67 for the CIN and 4.47 for the MIN. This spread is clearly visible from the minimum and maximum values of the CIN and MIN with a range from 1 to 15 and 18 unique search terms. The median values are 4, 5, and 7 for the VIN, CIN, and MIN, respectively. The median values, as another measure of central tendency of the data set, informs us how the data is distributed when numerical ordered. In combination with the standard deviation, the median values help us interpret the strength of the mean value according to the central tendency of the data. The results of the number of unique search terms present a picture of how fewer unique search terms are used to search the VIN compared to those of the CIN and MIN. This indicates that the searching of the VIN is focused and less complicated viewed as a querying search style, and that the VIN can be so well-defined that the search starts with the entering of an URL (minimum value 0). More unique search terms are used for the searching of the CIN and again more for the MIN.

In Table 5 we see that the VIN can be satisfied in the first attempt of searching, which explains the zero value (minimum) and refers to the fact that no additional searching is carried out. We had two instances of no search iterations (test participants #10 and #18). Instead of using a search engine and formulating search terms they directly entered a known URL address and found the relevant information via hyperlinks. However, we do also see that it may take as many as 9 search iterations (query modifications) to handle the VIN. It takes more search iterations for the handling of the CIN and MIN with maximum values of 15 and 17, respectively.

Table 11 reports the test participants' overall searching time with respect to the three types of information needs. When looking at the mean values a clear and distinct picture appears with the MIN being the information need that most time is spent on searching with 17:17 min compared to 14:18 min for the CIN, and 5:36 min for the VIN. It is, however, a less clear picture when we consider the relative high standard deviation values of 4:07, 6:57, and 6:02 for the VIN, the CIN, and MIN. In particular the two topical-oriented information needs, the CIN and the MIN present very similar values, which is even more distinctive when looking at the minimum values of 5:37 and 5:02 for the CIN and the MIN, respectively. Similarly, the maximum values of these two types of information needs are close with 28:27 min spent on searching for the CIN compared to 28:30 min for the MIN.

The median values for these two types of information needs are 16:12 and 16:38 for the CIN and the MIN in that order. For a visual illustration of the distribution of the central tendencies we have added a box plot (Fig. 1) to supplement Table 11. Fig. 1 shows not only the median values (2nd quartile) but also the 1st and 3rd quartiles, which help to understand how the three types of information needs differ from each other with respect to searching time, in particular the two topical-oriented types of information needs. Fig. 1 further confirms the difference of mean values despite relatively equal minimum, median, and maximum values of the two topical-oriented information needs shown in Table 11. On average the test participants spent 2 min and 59 s more on searching the MIN compared to the CIN.

4.2.1. The quantitative analysis and Ingwersen's predictions

From Ingwersen's predictions and the qualitative analysis we know that the VIN is characterised as the information need searched that is most focused and goal-oriented. We know that the test participants stick to the topic in question and are not distracted or tempted by other links. We further know that the test participants know what to search for, are able to express

Fig. 1. Box plot of the test participants' time spent on searching the three information needs.

the information need, quick at assessing relevance, do some query modifications, and stop when the information need is satisfied. This they do in a relatively short time, actually the shortest time spent compared to the searching of the CIN and the MIN. Table 4 shows that they spend on average 4.8 unique search terms (the lowest number compared to the CIN (6.2) and the MIN (7.3)), which indicate a well-perceived and well-defined information need that they are able at expressing. This is also supported by Tables 5 and 6 that show that the test participants spend on average 2.9 search terms per iteration (the highest number compared to the CIN and the MIN) and make a minimum of query modifications (search iterations, mean 3.3, Table 5). Tables 7–10 show that it only requires, on average, few web pages (Table 7), few websites (Table 8), few visits to individual web pages of a website (Table 9) and few relevant web pages (Table 10) to satisfy the VIN, which supports the characteristics of focused and goal-oriented searching of a well-perceived and well-defined information need. This is further supported by the fact that the VIN is searched in the shortest amount of time (Table 11). The searching of the VIN can be described as formulating and reformulating queries (Tables 4–6) (querying), and assessing relevance (Table 10) (filtering). In other words, the quantitative search attributes of the VIN supports the predicted search behaviour of the targeted and focused style of searching and how information is easily assessed in the style of filtering out irrelevant information (Ingwersen, 2000, p. 165).

As for the CIN we know that it is characterised as being less focused compared to the VIN and more focused than the MIN. The information searching of the CIN is more exploratory with more interaction taking place with the retrieved websites and web pages. The test participants know what to search for and have no problems expressing the information need or assessing relevance. They spend more time searching, more time on a given domain, and more time on viewing the results than with the VIN, and less compared to the MIN. The CIN is quite similar to that of the MIN, but more well-defined though variable (dynamic) in nature (Ingwersen, 2000, p. 165), and is characterised by an exploratory search behaviour, which the present data seems to confirm. Tables 4, 6 and 10 show that the test participants use on average 6.2 unique search terms (more than the VIN and less than the MIN), 2.7 search terms per iteration and mark 3.8 web pages as relevant, which indicate a well-perceived information need that they are capable of articulating and getting satisfied. Table 5 supports the exploratory search behaviour with the average number of query modifications (iterations) of 4.6. So do Tables 7–9 with the relatively high number of average visited web pages of 30.5 (Table 7), deriving from an average of 5.6 visited website (Table 8) with an average of 5.8 web pages visited per website (Table 9) for the CIN. It takes the test participants longer to search the CIN with an average time of 14:18 min compared to the 5:36 of the VIN (Table 11). The searching of the CIN is by Ingwersen (2000, p. 165) predicted as querying, navigation, and dynamic interaction. This search behaviour is nicely supported by Tables 4–6 with the formulation and reformulation of queries (querying), and the exploratory style of searching (navigation and dynamic interaction) depicted in Tables 5 and 7–9. In short, the quantitative search attributes of the CIN supports Ingwersen's predicted search behaviour of the CIN.

Both the qualitative analysis and Ingwersen's predictions point to how the MIN is characterised as the information need searched with the least focus compared to the VIN and CIN. The test participants find the MIN the hardest to express out of the three information needs and hence harder to assess relevance of, as it is the information need they are less confident about. It is at the same time the information need the test participants explore the most, and which they spend the longest time on searching. The issue of less focus is seen by the very exploratory searching style with the highest average number of query modifications (search iterations) of 5.8 (Table 5). Tables 7–9 further support this tendency with an average of 47.1

visited web pages (Table 7), an average of 7.6 visited websites (Table 8) from where an average of 6.4 web pages are visited (Table 9). The issue of less confidence about the MIN is seen by how this type of information need is harder to express with an average of only 2.2 search terms per query modification (Table 6). The high number of average query modifications of 5.8 (Table 5) further confirms this difficulty, and so does the high number of unique search terms (Table 4), which indicate it takes many attempts to ‘zoom’ in on this type of information need. Table 10 shows that the MIN is satisfied with an average of 4.6 relevant web pages. With respect to the two topical types of information needs, the CIN and the MIN, more web pages are saved compared to that of the VIN. This may be because the topics of the information needs are multi-faceted, and/or because the test participants are less familiar with the topics and assess relevance with a larger tolerance of what may be relevant. The high number of search modifications (Table 5) and exploring interaction with retrieved websites (Table 8) and webpages (Tables 7 and 9) also shows dedication and persistency in the test participants’ searching of the MIN. The test participants are willing to spend time on this information need. Table 11 depicts that the MIN with 17:17 min is the information need they spend the longest time on searching. The indisputable difference of time spent on searching the VIN versus the MIN is addressed by Ingwersen and Järvelin (2005, p. 287) who point out that users with a VIN often perform a given underlying work task relatively quickly and without problems. In contrast to users with a MIN who spend long time on information searching because the underlying work task is complex and requires that the user acquires great amounts of knowledge (Ingwersen & Järvelin, 2005, pp. 287–288). Ingwersen also infers the factors of interest and motivation (Ingwersen, 1996, p. 20; Ingwersen, 2000, p. 165). The MIN is defined as being poorly described in the consciousness of the user that result in high cognitive uncertainty, but also processes high degrees of motivation and curiosity characterised by search loops, browsing, and trial & error (Ingwersen, 2000, p. 165). The high degree of exploratory searching of the MIN and the high amounts of search interactions combined with the low number of search terms per query modification support Ingwersen’s predicted search characteristics of the MIN.

4.3. Concluding remarks

The presentation of the search characteristics in Sections 4.1 and 4.2 answers the research question. Both the qualitative and quantitative analyses of the characteristics of the Internet information search behaviour of the three types of information need do compare and match the corresponding descriptions by Ingwersen. In other words, the analytically deduced information search behaviour characteristics by Ingwersen are positively corroborated for this group of 18 young Danish test participants who search the Internet as part of ELIS.

5. Summary and implications

The objective of the reported study has been to investigate the search behaviour characteristics that Ingwersen predicts and associates with the three fundamental types of information needs (the VIN, CIN, and MIN). The study is based on a purposive sample of 18 ordinary users that searched the Internet as part of ELIS. Ingwersen’s associated search behaviour characteristics are corroborated for this group of test participants, and an even more detailed picture of the search behaviour of the three types of information needs is reached. Based on the mean values from Tables 4–11 we see how the VIN, with one exception, always obtain the least of unique search terms, search iterations, visited web pages, unique websites visited, web pages visited per unique website, relevant web pages, and time spent searching compared to the CIN and MIN. The exception is the number of search terms per search iterations where the VIN required more than the other two types of information needs. The MIN required the most unique search terms, search iterations, visited web pages, unique websites visited, web pages visited per unique website, relevant web pages, and time spent searching, with the exception of the number of search terms per search iterations, which is signified by the least number in contrast to the VIN. The CIN positions itself with medium mean values compared to the other two types of information needs. For example, see Table 12 for an illustration.

Though the objective of the study was to empirically investigate the search behaviour characteristics of the three types of information needs, it is also worth considering the usefulness and potential application of this corroborated insight. In the introduction we motivate how this insight can be relevant for IR systems design, and from the results it is obvious how the different information needs are signified by different information search behaviours, which implicates different strategies to

Table 12
Summarization of the search behaviour characteristics of the three types of information needs.

	VIN	CIN	MIN
Number of unique search terms (Table 4)	Low	Medium	High
Number of search iterations (Table 5)	Low	Medium	High
Number of search terms per search iteration (Table 6)	High	Medium	Low
Number of visited web pages (Table 7)	Low	Medium	High
Number of unique websites visited (Table 8)	Low	Medium	High
Number of web pages visited per unique website (Table 9)	Low	Medium	High
Number of relevant web pages (Table 10)	Low	Medium	High
Time spent searching (Table 11)	Low	Medium	High

support the searching of the different information needs. The information search behaviours differ in several aspects of the test participants' interaction with the IR system (the Internet), for instance with reference to the number of unique search terms, the number of relevant web pages, the overall time spent searching, the prior knowledge of topic in question, the motivation for searching, and the degree of focus. When knowing that a user with a fact-oriented VIN often uses the search strategies *querying* and *filtering* and is goal-oriented, the IR system can be designed to support this specific type of search behaviour. In comparison an IR system that supports the user with a MIN must be tailored towards the user's expected explorative browsing search behaviour and may benefit from knowing that the user is strongly motivated, persistent and engaged in the searching.

The reported study does also present empirical evidence concerning Internet user search behaviour of the three types of information needs that is relevant for IR techniques such as, relevance feedback, user modelling, information filtering, personalisation, and similar approaches that aim to calculate and model user behaviour in order to optimise IR performance. The presented evidence can also be used for the modelling of IR system evaluation in the lab setting in line with the work of Keskustalo, Järvelin, Pirkola, Sharma, and Lykke (2009) and Keskustalo and Järvelin (2010).

The results are also relevant to evaluation of IIR systems and user search behaviour. As mentioned in the introduction, different types of information needs that resemble the VIN, CIN, and MIN are commonly employed in IIR evaluations (e.g., Bell & Ruthven, 2004; Toms et al., 2007; White et al., 2001, 2002, 2003, 2008). The present study may constitute the initial work on a potential baseline for the search behaviour of ELIS Internet searching of the three different types of information needs. Further work and validation is required. A simplification of such a baseline is Table 12. Such a baseline would be applicable for IIR evaluation by use of simulated work task situations and may replace the requirement of always to include test participants' genuine information needs as a baseline for comparison of the search behaviour of genuine and simulated information needs (e.g., Borlund, 2003; Borlund & Schneider, 2010). Furthermore, the study presents input of how to construct simulated work task situations that resemble the three fundamental types of information needs, when following the instructions the test participants received in order to prepare their information needs. That is, an investigator who wants to construct simulated work task situations that reflect Ingwersen's types of information needs ought to ensure the simulated work task description of a VIN to be about looking for factual information of relevance and interest to the group of test participants. As for the two topical-oriented types of information needs (CIN and MIN) the construction requirements concern the degree of prior knowledge of the group of test participants with reference to the interesting and relevant search topics of the simulated work task situations. Hence this study forms part of the research agenda concerning the refinement of the requirements of the application of simulated work task situations outlined by Borlund and Schneider (2010).

In conclusion, the present study empirically corroborates the analytically deduced information search behaviour characteristics by Ingwersen for this group of test participants that searched the Internet as part of their ELIS. Hereby previous research by Ingwersen is strengthened. In addition, the results of the study possess potentials for future research within a variety of research directions, and we kindly invite the IR research community to help us explore these potentials.

Acknowledgements

We are grateful to Jesper W. Schneider for the many interesting discussions on the use of statistics, the power of sample techniques, effect sizes as well as the many other interesting topics we have come across in those discussions. We also thank the test participants for their time and effort, and the reviewers for constructive input that has improved the paper. This work has been supported in part by the Danish Ministry of Culture (REX project/TAKT2011-072).

References

- Aula, A., Jhaveri, N., & Käki, M. (2005). Information search and re-access strategies of experienced Web users. In *Proceedings of the 14th international conference WWW 2005* (pp. 583–592). New York, NY, USA: ACM.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5), 407–424.
- Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *The Canadian Journal of Information Science*, 5, 133–143.
- Belkin, N. J., Oddy, R., & Brooks, H. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, 38(2), 61–71.
- Bell, J. D., & Ruthven, I. (2004). Searcher's assessment of task complexity for web searching. In S. McDonald & J. Tait (Eds.), *Proceedings of the 26th European conference on information retrieval (ECIR 2004)*. 5th–7th April 2004, University of Sunderland, U.K. (pp. 57–71). Berlin/Heidelberg: Springer Verlag.
- Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1), 71–90.
- Borlund, P. (2003). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3). <<http://informationr.net/ir/8-3/paper152.html>>.
- Borlund, P. & Schneider, J. W. (2010). Reconsideration of the simulated work task situation: A context tool for evaluation of information retrieval interaction. In D. Kelly, A. Turpin & L. Azzopardi (Eds.), *Proceedings of the third IliX symposium on information interaction in context (IliX 2010)* (pp. 155–164). New York, NY, USA: ACM.
- Borlund, P. (2010). The cognitive viewpoint: The essence of information retrieval interaction. In B. Larsen, J. W. Schneider, & F. Åström (Eds.), *The Janus faced scholar. A festschrift in honor of Peter Ingwersen* (pp. 23–34). Copenhagen: Royal School of Library and Information Science.
- Bryman, A. (2008). *Social research methods* (3rd ed.). New York: Oxford University Press.
- Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review*, 48(3), 378–399.
- Cleverdon, C. W., & Keen, E. M. (1966). *Aslib Cranfield research project: Factors determining the performance of indexing systems. Results* (Vol. 2). Cranfield, UK: The College of Aeronautics.
- Cleverdon, C. W., Mills, J., & Keen, E. M. (1966). *Aslib Cranfield research project: Factors determining the performance of indexing systems. Design* (Vol. 1). Cranfield, UK: The College of Aeronautics.
- Cohen (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997–1003.
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research* (3rd ed.). Los Angeles, CA: Sage.

- Crystal, A. J., & Wildemuth, B. M. (2009). Naturalistic research. In B. M. Wildemuth (Ed.), *Application of social science research methods to questions in information and library science*. Westport, CT: Libraries Unlimited.
- Ingwersen, P. (1992). *Information retrieval interaction*. London, UK: Taylor Graham.
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 52(1), 3–50.
- Ingwersen, P. (2000). Users in context. In M. Agosti, F. Crestani, & G. Pasi (Eds.), *Lectures on information retrieval. Third European Summer-School, ESSIR, Varenna, Italy* (pp. 157–178). Heidelberg: Springer-Verlag.
- Ingwersen, P. (1986). Cognitive analysis and the role of the intermediary in information retrieval. In R. Davies (Ed.), *Intelligent information systems* (pp. 206–237). Chichester, West Sussex: Horwood.
- Ingwersen, P., & Järvelin (2005). *The turn: Integration of Information seeking and retrieval in context*. Dordrecht, NL: Springer.
- Ingwersen, P., & Kaae, S. (1980). *User-librarian negotiations and information search procedures in public libraries: Analysis of verbal protocols, final report*. Copenhagen, DK: Royal School of Librarianship.
- Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(1–2), 1–224.
- Keskustalo, H., & Järvelin, K. (2010). Query and browsing-based interaction simulation in test collections. In *Simulation of interaction: Automated evaluation of interactive IR. Workshop of the 33rd annual international ACM SIGIR conference on research and development in information retrieval* (pp. 29–30). Geneva, Switzerland: ACM.
- Keskustalo, H., Järvelin, K., Pirkola, A., Sharma, T., & Lykke, M. (2009). Test collection-based IR evaluation needs extension towards sessions: A case of extremely short queries. In *AIRS'09 proceedings of the 5th Asia information retrieval symposium on information retrieval technology* (pp. 63–74). Berlin/Heidelberg: Springer Verlag.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 61(5), 246–759.
- Kvale, S. (2007). *Doing interviews*. London: Sage Publications.
- Robertson, S. E., & Hancock-Beaulieu, M. M. (1992). On the evaluation of IR systems. *Information Processing and Management*, 28(4), 457–466.
- Savolainen, R. (1995). Everyday life information seeking: Approaching information seeking in the context of 'way of life'. *Library and Information Science Research*, 17(3), 259–294.
- Shaver, J. P. (1993). What statistical significance testing is, and what it is not. *Journal of Experimental Education*, 61(4), 293–316.
- Tombros, A., Ruthven, I., & Jose, J. M. (2005). How users assess web pages for information seeking. *Journal of the American Society for Information Science and Technology*, 56(4), 327–344.
- Toms, E. G., O'Brien, H., Mackenzie, T., Jordan, C., Freund, L., Toze, S., et al (2007). Task effects on interactive search: The query factor. In N. Fuhr, J. Kamps, M. Lalmas, & A. Trotman (Eds.), *Proceedings of the 6th international workshop of the initiative for the evaluation of XML retrieval (INEX), Dagstuhl, Germany* (pp. 359–372). Berlin/Heidelberg: Springer Verlag.
- White, R., Bilenko, M., & Cucerzan, S. (2008). Leveraging popular destinations to enhance web search interaction. *ACM Transactions on the Web*, 2(3), 30.
- White, R. W., Jose, J. M., & Ruthven, I. (2003). A task-oriented study on the influencing effects of query-biased summarisation in web searching. *Information Processing and Management*, 39(5), 707–733.
- White, R. W., Jose, J. M., & Ruthven, I. (2001). Query-based web page summarization: A task-oriented evaluation. In W. B. Croft, D. J. Harper, D. H. Kraft, & J. Zobel (Eds.), *Proceedings of the 24th annual international ACM SIGIR Conference on research and development in information retrieval (SIGIR 2001). September 9–13 2001* (pp. 412–413). New Orleans, Louisiana, USA: ACM.
- White, R. W., Ruthven, I., & Jose, J. M. (2002). Finding relevant documents using top-ranking sentences: An evaluation of two alternative schemes. In M. Beaulieu, R. Baeza-Yates, S. H. Myaeng, & K. Järvelin (Eds.), *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval (SIGIR 2002), August 11–15 2002* (pp. 57–64). Tampere, Finland: Springer.
- Wildemuth, B., Freund, L., & Toms, E. (2013). Designing known-item and fact-finding search tasks for studies of interactive information retrieval. In *Proceedings of the second association for information science and technology ASIS&T (European Workshop), June 5–6, 2013* (pp. 131–162). Åbo: Åbo Akademi University.