



This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



**CC creative commons**  
COMMONS DEED

**Attribution-NonCommercial-NoDerivs 2.5**

**You are free:**

- to copy, distribute, display, and perform the work

**Under the following conditions:**

**BY:** **Attribution.** You must attribute the work in the manner specified by the author or licensor.

**Noncommercial.** You may not use this work for commercial purposes.

**No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

**Your fair use and other rights are in no way affected by the above.**

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

## **A Study of Students' Information Searching Strategies**

### **Andrew N Baldwin, Professor of Construction Management**

Dept of Civil & Building Engineering, Loughborough University,  
a.n.baldwin@lboro.ac.uk

### **Elizabeth Gadd, Academic Services Manager (Engineering)**

University Library, Loughborough University

### **Panos Balatsoukas, Research Student**

Department of Information Science, Loughborough University

### **Abstract**

Concerns have been expressed with respect to students' ability to search for information using electronic search engines and databases. This research adopted a structured method comprising a combination of questionnaire surveys, an observational study and a 'sense making' interview to assess the information searching skills of a group of 14 students undertaking their final year dissertation studies on undergraduate programmes within the Department of Civil and Building Engineering at Loughborough University. The findings reveal that the participants encountered problems with each type of search engine used (Google, Metalib, the Library OPAC system, and individual databases) and lacked knowledge of how to use advanced search strategies. All the participants formulated queries using simple words or free text and there was no evidence of consideration of structured word searching using systematically selected keywords. The results indicate priority areas for additional tuition in information literacy.

**Keywords:** Information Literacy, Information Searching, Dissertations, Literature Reviews

## Introduction

Despite the availability of sophisticated tools for information searching students' literature reviews seldom meet the expectations of their supervisors. Literature reviews are frequently criticised by researchers for their poor quality, lack of a range of sources, lack of depth, and insufficient analysis of the information located. Poor citation remains an ongoing problem. In what is now accepted as the 'information age' opportunities exist for students to make better use of the sources of information available. How adept are they at doing this?

The aims of the research reported in this paper were to investigate students' information searching strategies when undertaking literature reviews for individual, 'final-year' dissertations/projects and to design new approaches to literature reviews that take account of both students' existing and required search skills. The research team was seeking a better understanding of students' knowledge of and ability to use the search engines and databases commonly available to undergraduates. These include OPAC, (an online database of materials held by a library), Metalib (a 'gateway' to academic information obtained via library catalogues, databases and e-journals), and Google (the internet search engine). This experimental study focused on the information searching behaviour of students in terms of several characteristics including: the type of information systems used; the number and type of queries submitted; the number of successive searches undertaken; the use of relevant feedback mechanisms; and the number of search results viewed.

This aspect of 'information searching' was investigated in the wider context of the students' understanding of what comprises a literature review; the value of the literature review to their dissertation; their understanding of the importance of citations; and their awareness of bibliographic management tools and electronic search engines. The research included observational studies to confirm/reject these perceptions. To place these findings in context a review was made of the teaching relating to the information seeking and information searching provided to students both immediately prior to their dissertation/project literature review and in the earlier stages of their undergraduate education. Studies were also made of past dissertations/projects. Staff views of students' information searching skills were also solicited.

The sample of students selected for the study comprised students from seven undergraduate programmes within the Department of Civil and Building Engineering at Loughborough University. These programmes cover a range of studies within the Built Environment Discipline including: Civil Engineering; Construction Engineering Management; Commercial Management and Quantity Surveying, Architectural Engineering and Design Management, Transport and Business Management and Air Transport Management. The total number of students currently engaged in their dissertation/project, i.e. the total sample size, was 206.

The findings from this research have identified priorities for improving the teaching provided to students when undertaking the literature reviews. These changes to the teaching will enable students to undertake more efficient, effective searches for not only the literature reviews required for their final year dissertations but also for other assignments throughout their undergraduate studies. This paper includes details of all the research undertaken but focuses in particular on the results of the student survey and observational studies as they are the dominant part of the research. Additional findings are reported from the qualitative research undertaken to supplement the observational study. Definitions of each of the search systems reviewed and specific terms used in the analysis of the observational studies are included in Appendix 1.

### Literature Review

The widespread adoption of the world wide web, WWW, resulted in a range of subsequent research as to how users interact with web based systems. The primary driver for such research was originally improved interface design. Belkin *et al.*, (1995) reviewed information seeking strategies for the design of effective interaction with systems and information retrieval (IR). This research, based upon a model of information seeking behaviour (Cool and Bryne, 2002), assumed that user interaction with systems to seek information can be viewed as information-seeking strategies which may be modelled as a 'dialogue' suitable for both the design of systems and the analysis of system users' performance.

Bryne *et al.*, 1999, present a 'Taskonomy', a taxonomy of tasks, of WWW use. This provides a range of tasks and sub tasks including: use information, locate on page, go to page, provide information, configure browser, and react to environment. Their research concluded that WWW browsing is a complex cognitive process and that, rather than dwell on research into specific aspects of the design of web pages, further research should include analyses at higher levels of abstraction e.g. users' strategies and patterns of behaviour. Subsequent research focused on individual task analysis e.g. Xie (2009) found that although research had demonstrated that task is the driving force for information retrieval few studies investigate how people engaged in search tasks or define their decisions and behaviours in the information retrieval process. Her research based on quantitative and qualitative data collected by multiple methods including web surveys, questionnaires, diaries, and telephone interviews, found that the nature of the task ( routine, typical and unusual), the stage of the task ( pre-focus, formation, and post-focus), the timeframe (extremely urgent, urgent, non-urgent), together with the task type, impact the information searching strategy.

Research within an academic environment has reviewed students' search processes and outcomes in a range of contexts. In their study of information needs and information seeking behaviour of scholars at a Turkish University, Ucak and Kubanoglu (1998), found major differences in the need and information seeking behaviour among engineers, scientists, social scientists, and humanities scholars. Within each discipline there are strong similarities. These similarities were most notable between engineers

Formatted: Not Highlight

and scientists followed by social scientists and humanities scholars. The greatest differences which were observed were those between engineers and humanities scholars.

Huuskonen and Vakkari (2008), examined students' information searching for the writing of an essay on evidence-based medicine. The study involved the completion of a set of tasks using the medical database Medline. Precision and relative recall were not found to be associated to the way information in retrieved items was used for performing the new tasks. The students appeared simply to look for a sufficient number of documents containing enough information to complete their task.

Griffiths and Brophy (2005) recognised that although there is a considerable body of research in respect of information retrieval (IR) systems, similar research on retrieval from the WWW is still not as advanced. Their research found that the majority of participants did not perform complex or advanced searches. Studies conducted into students' use of electronic resources and their searching behaviour on the web indicate that students rely on the web to find information about their everyday needs and that the time spent searching for information can significantly vary. Other researchers have found that students feel overconfident when they search for information on the web but in many cases they find difficult to evaluate its quality (Cmor and Lippold, 2001).

In the research undertaken by the EDNER project each student participant searched for fifteen test queries and completed questionnaires for each task undertaken. Two results were of particular interest. First, students preferred to locate information via a search engine above all options (Google being the search engine of choice) and secondly, whilst 70% of participants felt that they were successful, only half of them thought that it was easy to locate information, Griffiths and Brophy (2005).

Modelling the information seeking patterns of engineers and research scientists in an industrial environment Ellis and Haugan (1997) found that, 'although there were differences in the features of the information patterns of the research scientists and engineers, the behavioural characteristics were similar'.

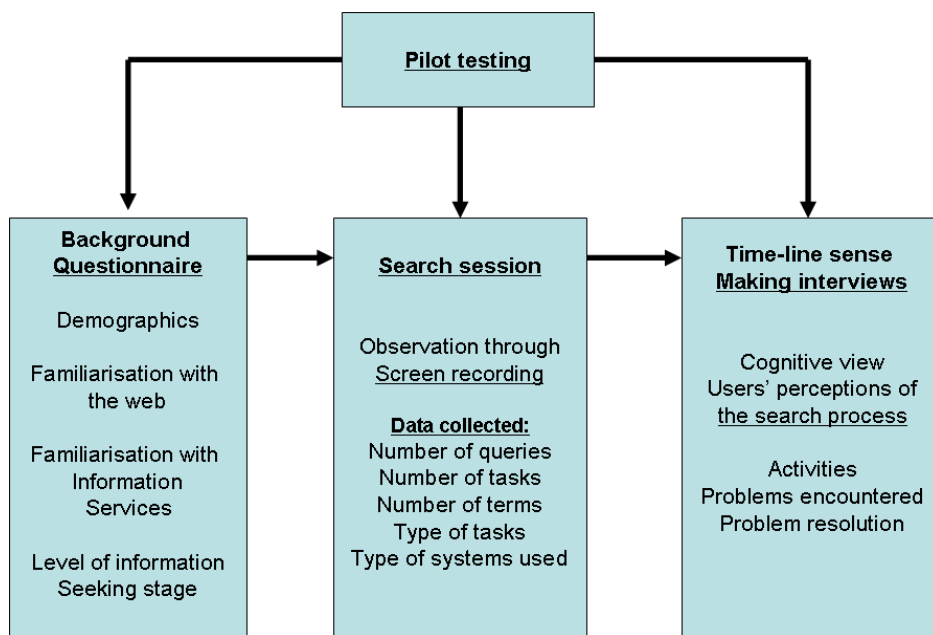
Other studies in an academic environment have identified different types of problems experienced by students, such as lack of training (Barsky and Bar-Ilan, 2005); lack of awareness about the use of library sources (Coleman and Sumner, 2004), lack of usability of IR systems (Morris and Balatsoukas, 2006), the impact of different cognitive models on the information searching behaviour (Ford *et al.*, 2003) and problems related to different expectations and mental models of the search process (Nicholas *et al.*, 2006).

Against this background the research reported in this paper sought to investigate the information searching skills required by students in undertaking their dissertation during the final part of their undergraduate studies. The research sought to confirm, or refute, the findings of the research described above and to provide new insight into how current students search for information. This, it was anticipated, would both improve

the quality of their dissertation, and also the quality of information searching for other assignments. To achieve this required data on students' search strategies, the systems used, the types of queries initiated, the number of results retrieved, and how information was selected and reviewed.

### Methodology

The methodology, shown diagrammatically in Figure 1, comprised a combination of two questionnaire surveys, an observational study and a 'sense making' interview with each study participant.



**Figure 1: Summary of the data collection process**

The participants were selected from the group of students undertaking their final year dissertation studies on undergraduate programmes. A check of the teaching and learning within the undergraduate programmes was undertaken to confirm that all the students had received the same basic level of tuition and support with respect to information literacy and therefore could be considered as a unified sample.

To recruit participants for the study all students undertaking a dissertation were emailed and invited to participate in the study. To ensure that they met a set of specific criteria they were required to undertake an on-line screening questionnaire to confirm that: they were familiar users of the web; they were frequent users of a set of information services, such as online library catalogue, search engines and scholarly databases; they were at the beginning of their information seeking process for their dissertation studies; they were pursuing research in either a construction management

topic or an experimental topic. After this screening process candidate participants were contacted through emails and meetings were arranged. The meetings took place at the Department of Civil and Building Engineering of Loughborough University. After participants' induction to basic concepts of this study, they were required to complete a consent form and a short background questionnaire.

They were then asked to search for information relating to their dissertation topic using the desktop computer and internet services provided in an office location. No specific task was set. The participants were asked to search for a real information need, rather than a predetermined task. Their searching behaviour was captured through the use of screen recording software. After the search session students were invited to participate in a semi-structured sense-making interview. During this interview they were asked about specific instances in their search and their decision making process. The purpose of the interviews was to increase the understanding about how participants made sense of their search process as it progressed and the decisions that they made within their search. The combination of behavioural and cognitive data gave a more holistic understanding of the students' information searching process. At the end of the data collection process each student was paid a small honorarium.

### **Background questionnaire**

The main function of the background questionnaire was to provide an in-depth knowledge about the participants' profile in areas that may influence and provide explanations of the way they performed during the search process. In addition, the items in the questionnaire ensured that participants met the criteria of this study by having the minimum skills and experience to undertake the experiment. The data collected provided demographic data about the participants; indicated their familiarity with personal computers as well as their use of the internet and the WWW and presented a synopsis at an individual and group level. The questions used in the questionnaire came from Spink (2002) and Spink and Dee (2007).

A variety of factual information and self-perception questions were used for structuring the body of the questionnaire. The questions used were 'structured' questions requiring the selection of one or more neutral fixed categories and questions based on Likert-type scales (Fife-Schaw, 2000). One question, the final question, was open-ended, inviting the respondent to add additional comments to support the answers already provided.

### **Observation**

The observation was facilitated through the use of screen recording software (the software product Camtasia was selected for this task). This software facilitated the collection of both quantitative and qualitative data. The quantitative data collected were: number of queries; number of words/terms per query; number of thesaurus terms (if applicable); number of information system used (e.g. Google or Metalib); the number of times natural language, browsing, Boolean operators and other search commands

were used; and the number of tasks performed during the search process. The qualitative data collected were: the type of information systems used (e.g. Google or Metalib); the type of tasks performed during the search process; and users' preference and comments.

The screen recording was initiated by the researcher prior to each student's information searching period. During this period, (a minimum of 30 minutes continuous operation) each participant was observed by the researcher who noted in a manual log significant events for later discussion. This was supplemented by the time-line sense making interview.

### ***Time-line sense making Interview***

Several information scientists have advocated the importance of combining objective data of user performance with cognitive data of a user's interaction with a system (Kulthau, 1991). After the search session, each participant was asked to participate in an interview. The purpose of the interview was to collect further qualitative data about users' perceptions of their searching behaviour. These interviews provided further details about users' cognitive processes during the search process. Dervin's time-line sense making interview was selected as a means of systematising the data collection process during the interviews. Dervin's time-line interview is an accepted method to facilitate data collection about users' activities during the web searching process (Jacobson 1991; Dalrymple, 2001). The participants are asked to recall the search process through a set of three repeated question statements. In the first question, participants were asked to recall their search process as a set of events. These are called *time-line events*. An example of a user's response might be: "I clicked on the internet explorer, then I typed in the search box: [www.google.com](http://www.google.com), then I typed in the query box: information science AND Information seeking, then...then...then". Each *'then'* statement represents a time-line event. Then, for each time-line event the user is asked to recall any questions, puzzles or confusions about the specific event. For example, for the time-line event: "...*Then I evaluated the results displayed by the search engine*", a user could possibly respond that she was confused by the presence of the 'Similar pages' function displayed in the search result page, or about the way the results have been sorted. Finally, for each question, puzzle or confusion statement the user is asked to indicate whether a response was obtained. The data collected from these questions provided a complete description of users' searching behaviour from a cognitive perspective and enabled a more informed analysis of the quantitative data.

### **Analysis and Discussion**

The data were analysed to study the systems used, the queries performed, the search strategies used and the results retrieved. These results were analysed across all the search engines used, both collectively and individually.

A total of 14 final year students from the department of Civil and Building Engineering participated in this study. All participants indicated that they had been instructed to do a



literature review, while the majority of them, 13 out of the 14 participants, had been told to go to the library if they needed advice as to the type of information sources they should be looking for in the context of their study. Twelve participants mentioned that they had not received any advice about how many references they should be aiming to locate and use. In addition, the majority of participants had not been given any books or photocopies of specific journal articles to read (10 out of 14) or, names of specific authors that they should look up (11 out of 14). Only three participants indicated that they had been given titles of past dissertations and encouraged to refer to them. The analysis of the quantitative data collected from the observational study involved simple descriptive statistics and a set of Spearman's correlation tests, while the analysis of the qualitative data followed a content analysis of users' responses during the interviews.

### Systems used

Participants in the study used an average of 2.3 systems per search session. The maximum number of systems used was four, the minimum number was one. In particular, 10 participants used the Google search engine to search for information, while six participants used Metalib, the library online catalogue and individual databases respectively. In the case of 11 participants, these systems were used in combination. The most frequent combinations were between Google + individual databases (three participants), as well between Google + OPAC + individual databases (three participants). Less frequently occurred combinations were: Google + Metalib (two participants), Metalib + OPAC (two participants) and OPAC + Google (one participant).

Table 1 shows the order with which these systems were used by participants to search for information. Google was the first choice for the majority of students who participated in the study. Five students started searching for information using Google and another four participants used it as a second choice. Although five participants used Metalib either as their first or second choice, four participants preferred to access directly individual databases at the beginning of their search (such as Emerald, Loughborough institutional repository, or professional databases).

**Table 1: Order of use of IR systems by participants**

Systems used	First choice	Second choice	Third choice	Fourth choice
Google	5	4	1	0
Metalib	3	2	1	0
OPAC	2	3	1	0
Individual dbs	4	0	1	1

Although student's used a combination of systems in their searching there was little evidence that this was based on an informed structured approach. The following quote is typical:

"I was looking for any reports or sites that describe the methodology I want to use in more detail, there was nothing relevant in Metalib so I turned to Google hoping that there would be more information about my topic."

### Queries performed

A total of 98 queries were formulated by the participants. The lowest number of queries per session was two, the highest was 12. The mean number of queries per search session was seven. Of these queries a total of 58 were new queries while the remaining 40 were queries that refined an initial query. The results showed that participants used an average of 10 words per search session. A total of 13 refined queries were used in order to broaden a search, the remaining 27 queries were made to narrow it down. A closer examination of the effectiveness of the refined queries showed that the majority of participants achieved their goal to either to narrow down or broaden up their searches. However, in 13 cases participants' additional queries failed to refine their searches.

Participants used several strategies to refine their queries. These involved:

**Searching for the same term using different databases.** For example, one participant searched for: *construction relationships* using Metalib. In the main search, the participant searched under the *General* category the following databases: *Article first*, *Compendex* and *Civil Engineering Abstracts (Illumina)*. However, the participant retrieved 3760 documents. The refinement process involved searching the same keywords selecting only the *Civil Engineering Abstracts*. As anticipated, the participant managed to successfully narrow down the search output by reducing the number of selected databases.

**Narrowing down a search by adding more terms.** A participant performed a search in Metalib using the following terms as keywords: *hand arm vibration*. The system retrieved 1792 results for this search. During the refinement process the participant added the term *management* in the initial query: *management of hand arm vibration*. The system retrieved only 118 documents. The addition of more terms was the most frequently used refinement strategy by the participants of this study. A similar behaviour was observed in the case of another participant. This participant used Google to search for: *site waste management cost*. The search engine retrieved a total of 2,630,000 results. The participant needed to refine this query two times in order to reduce the amount of results retrieved in Google search engine. In the case of the first refinement, the participant added the term *benefits*: *site waste management cost benefits*, and retrieved 488,000 results. In the second refinement the term *analysis* was added: *site waste management cost benefits analysis*. Based on this refined query the system retrieved 11,500 results.

**Broadening a search by removing terms from the query.** For example, a participant retrieved only two documents for the query: *health and safety benefits of offsite construction*, using Metalib. In order to broaden this search the participant removed the

term *benefits* and repeated the search in *Metalib* using the same databases. The refined query retrieved nine documents.

**Broadening a search by replacing terms.** In this case participants replaced the terms used in the main query with broader terms. For example, a participant formed the query: *risk in construction*, using the OPAC. For this query, OPAC retrieved a total of 19 results. The participant attempted to retrieve more results by replacing the terms of the initial query with the much broader term: *risk assessment*. The refined query matched to a total of 268 documents.

Incorrect attempts for refining queries resulted from:

**Problems with the selection and use of databases in Metalib.** For example, a participant searched for: *offsite construction and precast concrete* using Metalib. The participant selected to search under the *Civil Engineering* category and *Civil + Build* sub-category. In the case of the main search the participant searched the following databases: *Civil Engineering Abstracts (Illumina)*, *Compendex*, *health and safety science abstracts*, *INSPEC* and *ZETOC*. However, the participant retrieved only six documents. The refinement process involved searching the same terms using only ZETOC. During the interview, the participant justified this action by arguing that the system should retrieve more results. The refined search, however, retrieved only two relevant documents.

The qualitative data obtained from the students highlighted difficulties with database selection when using Metalib. A total of 13 participants mentioned that they found problems selecting appropriate databases in which to search. A specific problem identified was a lack of knowledge about which databases were more relevant to their topic:

“There is a list of different databases, honestly I had no idea which one I should pick...I randomly selected three databases to search hoping that they would help me find something useful...”

The second difficulty was an inability to perform a search without first selecting a database:

“The system somehow forces you to select a database before searching...it should be possible for me to perform a general search without selecting a specific database.”

A third, the absence of a mechanism to provide students with the opportunity to search across data bases that were unique to different subject categories:

“Difficult to select a category to search since my topic is relevant to many disciplines...”

**Problems with the use of field searching and the evaluation of the results in OPAC.** In this case, a participant formulated a query using a combination of Boolean operators (the Boolean operator AND) and field searching. The refined query was:

*Swimming pool AND renewable energy.* Although the application of the Boolean operator AND was correct, however, the participant did not notice that the term *renewable energy* was entered in the wrong field (in the Author field instead of the Keyword field). This query was not successful and the participant retrieved no documents. The same problem was repeated for the majority of participants who used the OPAC in order to search for information. This happened because the students searched the library catalogue by keyword rather than the author's name.

The students did not like the absence of a description summary from the list of results retrieved in OPAC. Examples of comments made included:

"there was no description of the contents."

"There is no abstract of the book or a table of contents, I read only the title but the title is a bit general...still don't know whether the book is relevant."

**Problems related to the use of broader and narrower terms.** The most common problem faced by the participants of this study related to the inappropriate use of terms in order to broaden or narrow a search. A typical example of this type of problem was observed in the case of a participant who searched in Google search engine using the main query: *risk management in construction*. The system retrieved a total of 9,300,000 results. However, the participant expected to narrow down the search using the refined query: *risk in construction management*. In this case the system retrieved 70 million more results than the first search. The participant by removing the term *management* reduced the specificity of the query and increased the amount of results retrieved.

This led to problems in evaluation:

"The page [i.e. results page] does not give much information to decide whether there is something relevant or not...I have to go and check its website individually which is going to take a lot of time."

"The main problem was that Google retrieves too many results. I guess that most of them were relevant but I didn't spend a lot of time checking the results."

There is a clear indication that participants in the study lacked knowledge about how to use advanced search strategies such as Boolean operators and truncation. All participants formulated their queries using simple keywords and free text. Boolean operators were only used by three participants while truncation was used by two participants only. Finally, one participant searched for a specific phrase (by including text in quotation marks: "...") and one other participant preferred simply to browse for information rather than formulate a query.

### **Results retrieved and the process of judging their relevance**

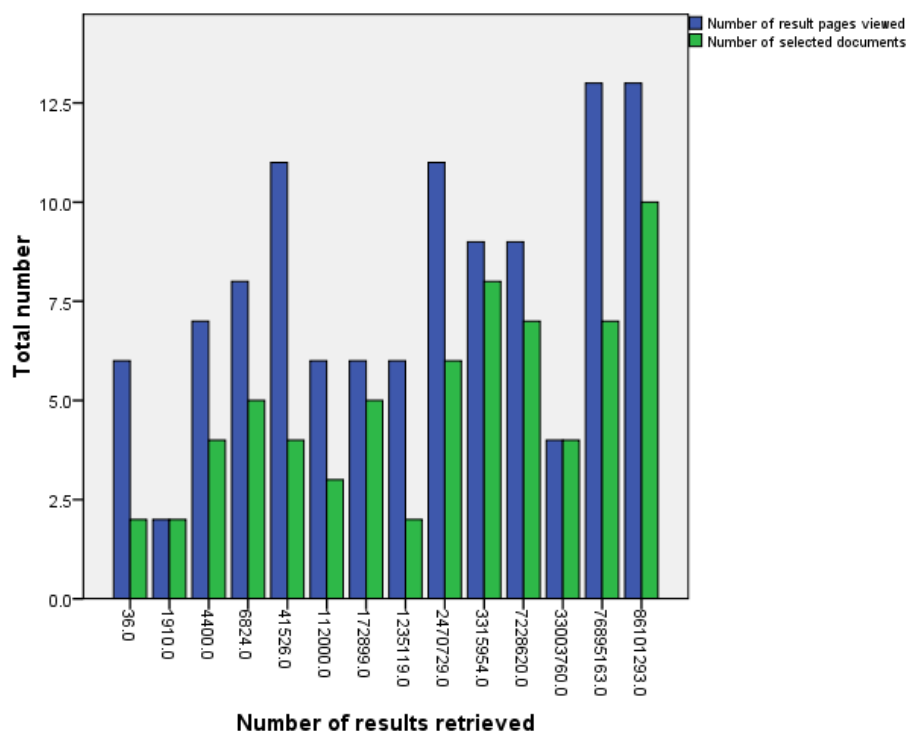
There was a wide variability in the number of results retrieved per query. The minimum number of results retrieved was 36 while the maximum approximately 86,100,000. This variability can be attributed to the type of systems used by participants to search for information. For example, the queries submitted in Metalib or the OPAC usually

retrieved no more than few thousands of metadata surrogates, but this number increased considerably in the case of the Google search engine. The fact that participants used the Google search engine frequently had an effect on the large mean number of total results retrieved per query. This number was estimated to be approximately 15,000,000 documents per query.

However, the variability of the total number of results retrieved did not influence participants in judging their relevance which was stable across queries. For example, participants reviewed a mean number of eight pages of results retrieved by each information retrieval system. A closer examination of the results showed that the majority of participants in the study (ten out of 14) reviewed between six and 13 pages of results. These numbers are very low given the huge amounts of results retrieved per query. However, this finding challenges ad hoc characterisations with respect to students' decisions when judging relevance and especially the perception that students do not usually assess the relevance of the results presented to them beyond the first page (Loringo *et al.*, 2008).

The number of documents selected as relevant by participants was also small. This number ranged between two and ten documents accessed or downloaded per search session. The mean number of documents selected by participants in the study was five. The results also showed that the majority of these participants (eight out of 14) accessed-downloaded between four and seven documents per search session.

The relationship between the number of pages viewed and the number of documents downloaded or accessed by participants is represented in Figure 2. Two Spearman's correlation tests were conducted in order to investigate whether there was a relationship between the number of pages viewed and documents accessed, and the number of pages viewed or documents accessed-downloaded and the number of documents retrieved.



**Figure 2: User behaviour with search results**

The results showed that the relationship between the number of pages viewed and documents accessed was significant at the 0.01 level. This means that as the number of pages viewed increases the number of relevant documents accessed increases too. Also statistically significant, at the 0.005 level, was the relationship between the number of documents retrieved and the number of pages viewed, and the relationship between the number of documents retrieved and the number of documents accessed-downloaded.

### Results by type of system

Having considered the results in a single sample the results were analysed for each of the three main systems used.

#### ***Participants' information searching behaviour in the Google search engine***

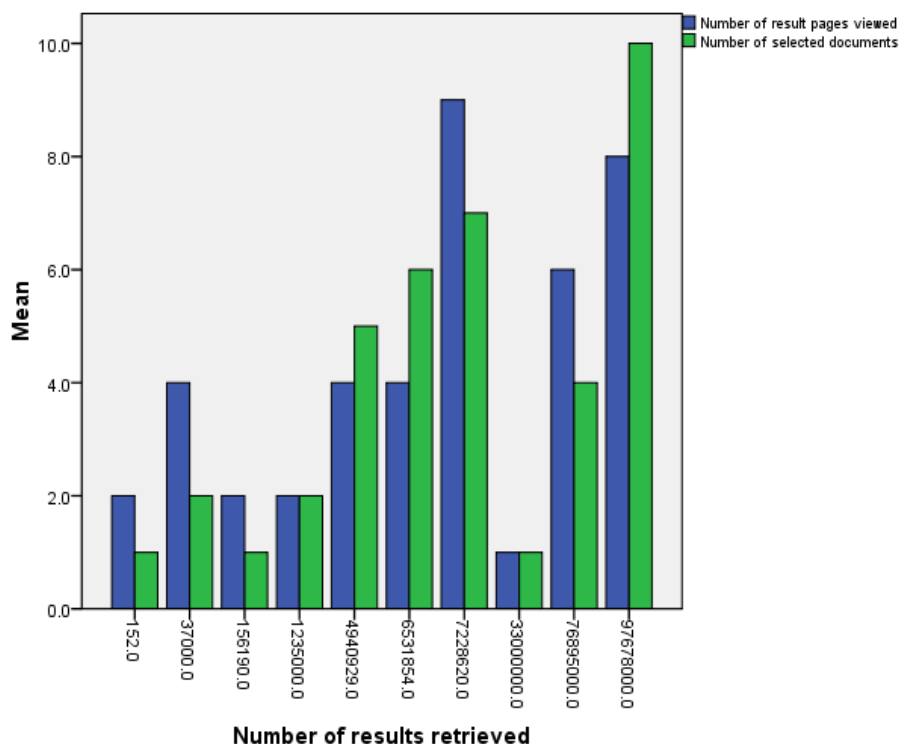
As already stated, the Google search engine was the most frequently selected system. Nine participants in the study used the simple search interface of Google search engine. Only one participant used Google Scholar, no participant used the advanced search interface. A total of 36 queries were formulated by participants using Google. The lowest number of queries was one while the highest was seven queries. However,

the mean number of queries per search session was 3.6. Of these queries a total of 19 were new queries while the remaining 17 were queries that refined an initial query. The results showed that participants used an average of 6.8 words per search session. All ten participants formulated their queries using simple keywords and free text. Truncation was used by one participant. None of the participants who used Google formulated queries based on Boolean operators.

The results showed a wide variability in the number of results retrieved per search session. The mean number of results retrieved per participant was 22,000,000 with a minimum number of 152 results retrieved and a maximum of 97,678,000 results. Participants reviewed a mean number of 4.2 pages of results using the Google search engine. A closer examination of the results showed that the majority of participants who used Google reviewed less than four pages per search session. It is noted that these numbers are very low given the huge amounts of results retrieved per search session. The results of the interviews showed that some of the participants were not satisfied with the presentation of the results in the search result interface of Google and this factor could have influenced the amount of pages evaluated by participants.

The number of documents selected as relevant by participants was again small. This number ranged between one and ten documents accessed or downloaded per search session. The mean number of documents selected by participants in the study was 3.9. The results also showed that half of the participants who used the Google search engine (five out of 10) accessed only one or two documents per search session. The low level of access to full text document can be attributed to the fact that Google does not provide full text access to many academic papers or corporate reports. This was an issue subsequently raised by participants in the interviews because it did not align with their expectations of the 'one-stop' online searching provided by Google.

The relationship between the number of pages viewed and the number of documents downloaded or accessed by participants is shown in Figure 3. Two Spearman's Correlation tests were conducted in order to investigate whether there was any relationship between the number of pages viewed and documents accessed, as well as between these variables and the number of documents retrieved. The results showed that the relationship between the number of pages viewed and documents accessed was significant at the 0.01 level. This means that as the number of pages viewed increases, the number of documents accessed as relevant increases too. However, no significant relationships were observed between the number of documents retrieved and the number of pages viewed, as well as between the number of documents accessed or downloaded and the total number of documents retrieved.



**Figure 3: User behaviour with search results (Google)**

The correlation tests were repeated, however, this time a distinction was made between new and refined queries. The results showed no significant relationship between the number of retrieved documents and the number of documents accessed as relevant, as well as between the number of retrieved documents and the number of pages viewed by participants in the study for both new and refined queries. Similarly, no significant association existed between the number of pages viewed and documents accessed.

**Participants' information searching behaviour in Metalib**

Six participants in the study used Metalib as a source for finding information. Since this study was focused on civil and building engineering students, four participants searched Metalib by specifying the category: *Civil Engineering* and the sub category: *Civil + Build*. This selection presented participants with a list of databases that were relevant to the specific category and sub category. However two users of Metalib did not specify any category or sub-category and searched under the *General* category of Metalib. The participants who used Metalib selected specific databases. The databases most often used by participants were: *Civil Engineering Abstracts (Illumina)* and *Compendex*. These were selected by four participants. Three participants selected

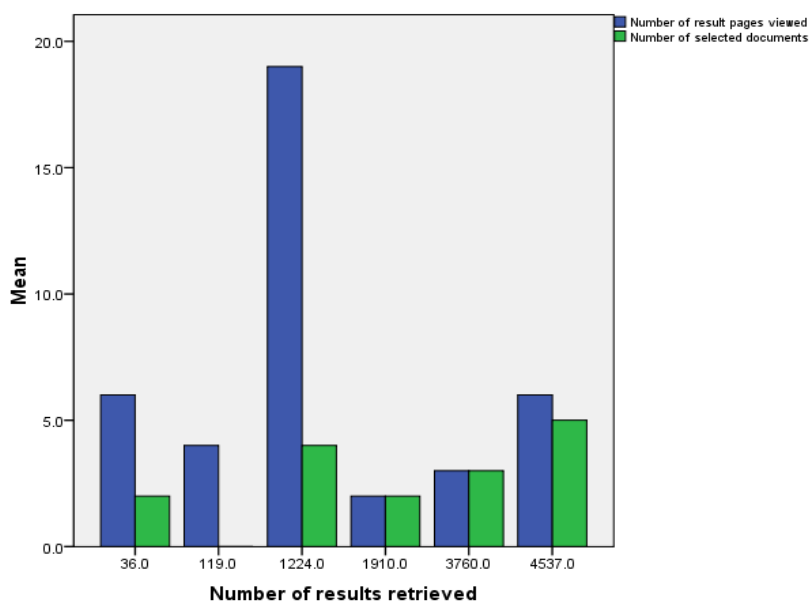


to search for information in *ArticleFirst*. Also, *Zetoc*, *INSPEC* and *Web of science* were preferred by two participants. Finally, *Intute*, *Institutional repository*, *OPAC*, *Aqualine*, *Health and safety science abstracts* and *SPiE digital library* were the databases that only one participant in each case preferred to search for information. However, the findings of the interviews showed that the process of selecting databases in the majority of cases was made randomly because participants lacked a clear understanding about the coverage of these databases. Participants selected databases either based on the relevance of their title (for example, Civil and Building Engineering Abstracts) or based on previous familiarity with specific databases (for example, OPAC, Institutional repository, INSPEC).

A total of 23 queries were formulated by participants in Metalib. The minimum number of queries was one while the maximum was six queries. However, the mean number of queries per search was 3.8. From these queries a total of 12 were new queries while the remaining 11 were queries that refined an initial query. Also the results showed that participants used an average of 8.5 words per search session. The six participants formulated their queries using simple keywords and free text. The Boolean operator AND was used by one participant. None of the participants who used Metalib formulated queries based on the use of advanced search techniques, such as truncation and filtering, or searched by phrase.

The mean number of results retrieved per participant was 1931 with a minimum number of 36 results retrieved and a maximum of 4537 results. Participants reviewed a mean number of 6.6 pages of results in Metalib. A closer examination of the results showed that the majority of participants who used Metalib reviewed more than four pages per search session. The number of documents selected as relevant by participants was again small. This number ranged between zero and five documents accessed or downloaded per search session. The mean number of documents selected by participants in the study was 2.6, which is lower than Google. The results also showed that half of the participants who used Metalib search engine accessed one or two documents per search session. This can be attributed to the lack of familiarity with the presentation of results in Metalib and the cumbersome process involved for downloading a full text document using Metalib. These factors, which were highlighted by participants in the interviews, explain why participants reviewed less of the results and accessed less relevant documents in Metalib rather than in Google.

The relationship between the number of pages viewed and the number of documents downloaded or accessed by participants is represented in Figure 4.



**Figure 4. User behaviour with search results (Metalib)**

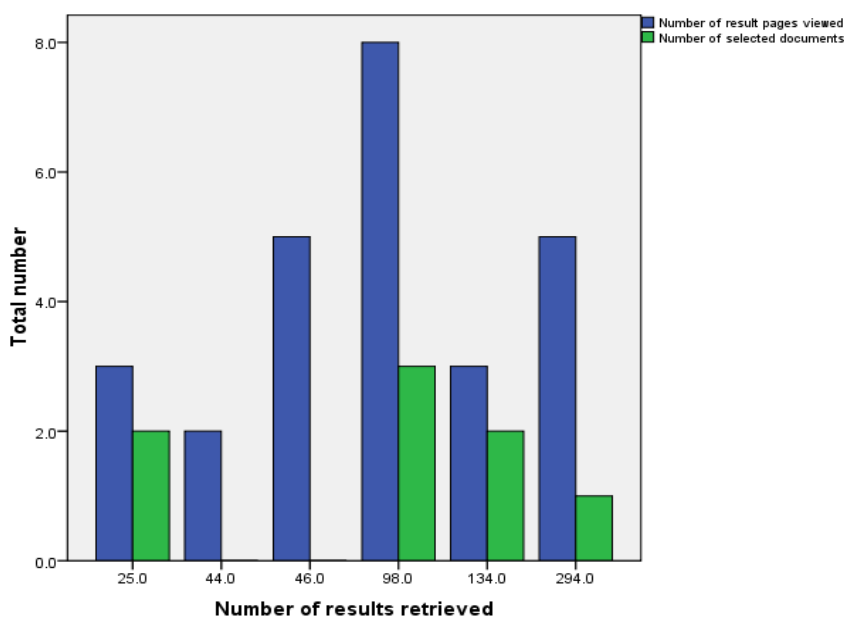
The Spearman's Correlation tests showed no significant relationships between the number of retrieved documents and the number of documents accessed as relevant, as well as between the number of retrieved documents and the number of pages viewed by participants.

***Participants' information searching behaviour in the library OPAC***

Six participants in the study used the library catalogue in order to search for information about their project. A total of 17 queries were formulated by participants using Google. The lowest number of queries was one while the highest was five. However, the mean number of queries per search was 2.8. From these queries a total of nine were new queries while the remaining eight refined an initial query. Also the results showed that participants used an average of four words per search session. All six participants formulated their queries using simple keywords and free text. Only one participant used Boolean operators. None of the participants used any type of advanced search technique or browsed for information. Finally, it is worth mentioning that two participants made use of metadata fields in order to restrict their queries to documents that contained the same terms in the title only.

The mean number of results retrieved per search session was 107 with a minimum number of 25 results retrieved and a maximum of 294 results. Participants reviewed a mean number of 4.3 pages of results. A closer examination of the results showed that the majority of participants who used the library OPAC reviewed between three and five pages per search session. The number of documents selected as relevant by participants ranged between zero and three documents per search session. The mean

number of documents selected by participants in the study was 1.3. Finally, the relationship between the number of pages viewed and the number of documents downloaded or accessed by participants is represented in Figure 5.



**Figure 5: User behaviour with search results (OPAC)**

As with the Metalib system, no significant relationships were observed between the number of documents retrieved, the number of result pages viewed by participants and finally the number of documents accessed as relevant.

### Summary and implications for informing student learning

As detailed above, previous research into students' information searching behaviour indicates that when searching for information students avoid the use of complex search strategies, such as the use of advanced search techniques and the formulation of complex queries: ' (Griffiths and Brophy, 2005). Also, previous findings have shown that students tend to use Google when searching for information on the web but only a few of them found easy to locate useful information (Griffiths and Brophy, 2005). 'Also, other researchers have found that students apply "satisficing" approaches to their searching behaviour which means that they usually complete a task as soon as they have identified a sufficient, but not an optimal, number of resources. (Huuskonen and Vakkari, 2008).

Our research endorsed these findings. The students observed had no overall strategy for their information search and showed no consideration of a structured approach to searching using predefined keywords. All the participants in our study commenced their search using a simple string of words and then expanded or narrowed the search

by the addition or truncation of this initial word string. They showed a preference for using the Google search engine but had an awareness of, and were prepared to use, other search facilities either as a first choice or in conjunction with Google. Students lacked the knowledge and skills to make efficient effective use of the search facilities. They had little or no knowledge of the use of Boolean operators. This confirms the results of Barsky and Bar-Ilan (2005), and Morris and Balatsoukas (2006). They revealed a lack of training and a lack of experience in the use of information retrieval systems. This lack of experience was amplified by 'problems' resulting from misplaced expectations of the systems being used.

Whichever system used for their search there was a wide variability in the average number of results retrieved per query. This variability can be attributed to the type of systems used by the participants to search for information. When using Google the mean number of results retrieved per participant was 22,000,000. However, the variability of the total number of results retrieved did not influence participants' judgment with respect to the relevance of the results. The participants reviewed a mean number of eight pages of results retrieved by the information retrieval system. This was consistently higher than may have been expected from the findings of previous research (see Loringo *et al.*, 2008), and the commonly held belief that when searching for information students simply select from the first page of results presented to them. When presented with extensive numbers of results the participants had no strategy for evaluating the quality of the information retrieved other than its apparent relevance to the subject area. This may be attributed to the fact that the participants were commencing their dissertations and their literature reviews and the fact that they were unfamiliar with either recent research projects and/or established researchers in the subject area. Consequently there was little or no evidence of student consideration of the usefulness/quality of the results accessed and downloaded as presented to the students. They approached the search problem with a view simply to locate a sufficient number of documents containing enough information to complete their task. This confirms the findings of Huuskonen and Vakkari (2008).

## Conclusions

The participants used a range of systems to search for information. Google was the students' preferred system for searching, although not to the exclusion of other systems. Metalib, the Library OPAC system, and individual databases were also used. Having conducted their initial search several strategies were used by students to broaden or narrow their search. The adoption of these strategies revealed a lack of understanding of selection methods. The participants encountered problems with each type of search engine used and lacked the knowledge of how to use advanced search strategies. All of the participants formulated queries using simple words or free text and there was no evidence of consideration of structured word searching using systematically selected keywords. There was a wide range in the number of results

retrieved per query. This variability can be attributed to the type of systems used by participants to search for information.

These overall findings from the study inform student learning requirements. The students participating in this observational study all encountered similar problems. These problems may be grouped into four main categories: problems relating to unrealistic expectations of the systems used; problems with the search process adopted; usability problems that arise from a lack of training in systems use; and problems relating to information overload. There is a clear need for all students to acquire a comprehensive understanding of the different search systems available for their literature review and the relative strengths of these systems which should include OPAC, Metalib, search engines including Google, and subject specific databases. Specifically, students need training in how to select the relevant system to use to find the information they require. They need to learn how to use systems effectively in both their basic and advanced modes of working. Guidance must be provided in how to identify and select the keywords with which to search these systems. This includes the use of Boolean operators and other key search parameters. Guidance must be given as to how to assess the quality of the information retrieved. This extends beyond the source of the material obtained to include the ability to critically assess the quality of the research findings reported in the document

### **Acknowledgements**

The research reported in this paper was funded by The Centre for Built Environment Education under their project funding reference RES6276 and the authors would like to thank the CEBE staff for their assistance and guidance in the research programme.

## References

Barsky, E. & Bar-Ilan, J. (2005) From the search problem through query formulation to results on the web. *Online Information Review*, 29 (1), 75-89.

Belkin, N. Cool, C., Stein, A. and Thiel, U. (1995). Cases, scripts and information-seeking strategies: on the design of interactive information retrieval systems, *Expert Systems with Applications* vol. 9, no. 3, pp. 379–395.

Byrne, M., John, B., Wehrle, N. & Crow, C. (1999) A tangled web we wove: A taskonomy of web use. In: *Human Factors in computing systems: Proceedings of CHI 99*. Reading, MA: Addison Wesley. pp. 544-551

Cmor, D. and Lippold, K. (2001). Surfing vs. searching: the Web as a research tool. Presented at the 21st Annual Conference of the Society for Teaching and Learning in Higher Education.

Cool, C., and Bryne, N.J.,(2002), 'A Classification of Interactions with Information', in H. Bruce, R. Fideel, P. Ingwersen & P.Vakkari, (Eds), *Emerging Frameworks and methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (COLIS4)* pp1-15, Greenwood Village, CO: Libraries unlimited, 2002.

Coleman, A. & Sumner, T. (2004) Digital libraries and user needs: Negotiating the future. *Journal of Digital Information*, 5 (3). Available at: <http://journals.tdl.org/jodi/article/view/119/123> [Accessed 1 November 2009].

Cool, C. and Belkin, N. (2002). A classification of interactions with information. In: *Proceedings of the Fourth International Conference on Conceptions of Library and Information Science* (Libraries Unlimited).

Dalrymple, P. W. (2001) [A quarter century of user-centered study: The impact of Zweizig and Dervin on LIS research](#). *Library and Information Science Research*, 23 (2), 155-165.

Ellis, D. and Haugan, M. (1997), "Modelling the information seeking patterns of engineers and research scientists in an industrial environment", *Journal of Documentation*, Vol. 53 No. 4, pp. 384-403.

Fife-Schaw, C. (2000). Questionnaire design. In: G. Breakwell, S. Hammond & C. Fife – Schaw (Eds.) *Research methods in psychology*. London: Sage.

Ford, N., Miller, D. & Moss, N. (2003) Web search strategies and approaches to studying. *Journal of the American Society for Information Science and Technology*, 54 (6), 473–489.

A.N. Baldwin, E. Gadd & P. Balatsoukas: A Study of Students' Information Searching Strategies

Griffiths, J.R., Brophy, P. (2005), "Student searching behaviour and the web: use of academic resources and Google", *Library Trends*, Vol. 53 No.4, pp.539-54.

Huuskonen, S. & Vakkari, P. (2008) Students' search process and outcome in Medline in writing an essay for a class on evidence based medicine. *Journal of Documentation* vol. 64(2), pp287-303

Jacobson, T. (1991) Sense-making in a database environment. *Information Processing & Management*, 27 (6), 647–657.

Kuhlthau, C. (1991). Inside the search process: information seeking from the user's perspective, *Journal of the American Society for Information Science* , vol. 42, no. 5, pp. 361–371.

Loringo, L., Haridassan, M., Brynjarsdóttir, H., Xia, L., Joachims, T. Gay, G., Granka, L., Pellacini, F., & Pan, B. (2008) Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, 59 (7), 1041-1052.

Marchionini, G. (2007) *Toward multimedia surrogates*. Available at: [http://ils.unc.edu/~march/multimedia\\_surrogates.pdf](http://ils.unc.edu/~march/multimedia_surrogates.pdf) [Accessed 2 November 2009].

Morris, A. & Balatsoukas, P. (2006) Usability evaluation of ebrary and overdrive online systems. In: P. Brophy, J. Craven & M. Markland (Eds.), *Libraries without walls 6: Evaluating the distributed delivery of library services*. London: Facet Publishing. Pp 211-223.

Nicholas, D., Huntington, P., Jamali, H. & Watkinson, A. (2006) The information seeking behaviour of the users of digital scholarly journals. *Information Processing and Management*, 42 (5), 1345–1365.

Spink, A. (2002). A user-centered approach to evaluating human interaction with web search engines: an exploratory study, *Information Processing and Management* , vol. 38, no. 3, pp. 401–426.

Spink, A., and Dee, C. (2007). Cognitive shifts related to interactive information retrieval. *Online Information Review*, vol.31, no.6, pp. 845-860.

Ucak, N.O., Kurbanoglu, S.S. (1998), "Information need and information-seeking behaviour of scholars at Turkey University", IFLA General Conference, Amsterdam, August, .

Xie, I. (2009). Dimensions of Tasks: Influences on Information Seeking and Retrieving Process. *Journal of Documentation*. Vol. 65(3), pp339-366.

## Appendix 1- Definitions:

**Documents:** This term refers to the full text documents (for example, web pages, journal papers or e-books).

**Google:** is the name of a multi-national public Internet search and advertising technologies corporation which has produced and operated the most widely recognised and publically used information search engine in the world.

**Individual databases:** The term represents any database or website that had been used by participants directly and was not accessed through an intermediary system or database, such as Metalib, Google or the library OPAC. In the case of individual databases, the user provided a URL, accessed and submitted a query directly to the database.

**Metalib:** A gateway to good quality academic information obtained via library catalogues, databases, and e-journals.

**OPAC:** An Online Public Access Catalogue comprising a database of materials held by a library or group of libraries which is searched primarily to locate books or other material physically located at a library.

**Query:** A specific, single search episode within a search session.

**Search session:** This term refers to all the queries submitted by a participant to one or more systems during the user study. Therefore, the term *search session* differs from the term *query* that refers to a specific, single search episode within a search session.

**Surrogates:** The term refers to the bibliographic records that are displayed to the user in a search result interface after a search for information has been performed. According to Marchionini (2007), the main role of surrogates is to help users to evaluate the relevance of the retrieved information to their information needs (i.e. sense making). Bibliographic or metadata records, however, are primarily used for storing information and facilitating information searching, but they are not necessarily appropriate for presentation in search result interfaces.

**System:** A type of information retrieval system