

## Scientific information retrieval behavior: A case study in students of Philosophy

Jesús Tramullas, Ana Sánchez Casabón  
Department of Information Sciences, University of Zaragoza  
Ciudad Universitaria s/n 50009, Zaragoza (Spain)  
{tramullas, asanchez}@unizar.es

**Abstract.** The behavior and patterns of recovery and processing of digital information by users is a recurring theme in the literature. The study of these behaviors are carried out through observation techniques and analysis of processes, actions and decisions undertaken by users in different situations. This paper presents the data resulting from the study of patterns of recovery and management of reference information of three consecutive courses of a specialized subject. The findings obtained showed a clear difference between patterns of information retrieval and obtained prior to the end of the training process, but there has been a significant change in the ultimate goal of users or appreciable changes in their prospects for application in other environments

**Keywords:** Information seeking behavior, information user studies, information literacy

### 1 Studies into information behaviour

The analysis and study of users' information behaviours is a classic theme which has been widely covered by investigation into information and documentation sciences and is reflected in the corresponding scientific publications. These studies have been covered with different methodological approaches [1] over a 50-year period, and have been extensively dealt with by Wilson [2], [3]. An international conference has been consolidated which is specifically dedicated to such studies, this being ISIC (*Information Seeking in Context*). Moreover, numerous specialised scientific journals have published studies into this matter.

One result of this evolution has been the increasing importance of qualitative study approaches. Most studies combine observation methodologies and the acquisition of quantitative data with qualitative group and individual assessment methods. Some examples of this kind of studies into information seeking behaviours analyses among university students are those of Zhang, Angehelescu and Yuan [4] and Heiström [5]. These works conclude that the level of command of a theme conditions the users' search behaviours rather than their skill to search, and that the way to approach study processes has an impact on the way in which users develop search processes, respectively. Junni [6] stressed that the amount of scientific bibliography used in

master's degree theses has increased with the growing use of digital information resources, but not the number of specialised sources consulted. Nicholas, et al [7] analysed the use of digital journals among the academic community, mainly from a quantitative perspective. If behaviour is limited to searching and discovering books, Rowlands and Nicholas [8] determined a predictive model for the university community according to the gender, discipline and academic status factors.

## 2 Study approach

The studies cited in the previous paragraph draw various conclusions about search patterns and use of information in different university contexts. Nonetheless, the strong contextual influence underlying them all means we have to accept their conclusions in the context they were done in. According to this principle, an analysis has been proposed in a Spanish university setting to identify any patterns and behaviours shared with other studies as far as possible or, conversely, to identify any typical contextual characteristics which may affect the environment they have been carried out in.

The curriculum of the Philosophy Degree taught at the University of Zaragoza contemplates training students in basic retrieval, processing and usage techniques to be used with scientific information. For such purposes, an optional course subject is taught with six credits known as *Information and Documentation for Humanities*. The Department of Information Sciences is responsible for teaching this optional course subject, which has been taught continuously between 2007-2010. The results of the learning established for this course subject are the following:

- Knowing the principles and foundations of scientific information and documentation processes.
- Developing scientific information search, selection and management processes.
- Include scientific documentation in all academic activities.

In order to accomplish these objectives, a series of learning activities are carried out which are included in the course subject's academic guidelines. The evaluation of these objectives and the determination of the level reached by students are performed by means of a personal scientific information retrieval and processing project.

Teaching this course subject in the aforementioned period has enabled us to not only determine the initial information retrieval patterns of a group of university students in relation to both reasoning and interfaces interaction processes, and their capacity to act from using these patterns, but to also obtain results in terms of the effectiveness of improvement actions. The following data acquisition techniques were used:

1. Direct observation of students' behaviours.

2. Personalised interviews and questionnaires about the activities carried out and the results.
3. Reports about the development and execution of personal projects.

### 3 Initial situation and analysis

Analysing the responses and observing the initial learning activities have enabled us to establish a low level of knowledge about the structure, organisation and usage of the digital information resources that are considered basic. The results are presented in Table 1.

**Table 1.** Preliminary use of information resources.

	Google	Advanced Google	OPAC	Reference Databases	Repositories	Subject Gateways
2007-08	12	1	2	0	0	0
2008-09	15	2	3	0	0	0
2009-10	14	4	4	0	0	0

The main source of information for students was the generic search engine, Google. Nonetheless, of the 47 students, only 7 reported a limited use of the advanced search interface and that they had not employed the thematic possibilities offered by Google (Table 2).

**Table 2.** Use of Google Services.

	Basic	Advanced	Scholar	Books	Blogs	Video	Image	News	Alerts
2007-08	12	1	0	0	0	0	1	0	0
2008-09	15	2	0	0	0	0	0	0	0
2009-10	14	4	0	1	0	1	2	0	0

This lack of knowledge about the advanced information retrieval services offered by Google was also accompanied by a mistaken personal perception of the students' own knowledge. At the beginning of the subject, the students were questioned about the level they believe they had of the knowledge and the use of the possibilities that this search engine offers. Of all the students, 86% stated that their level of competence was high or very high. This information was then verified with a series of 6 searches which entailed using different formulations of equations involving the selection of search terms, the application of several operators in the equations and the use of constraints. The results obtained reveal that the students' level of success did not reach 30%, and they showed they had no knowledge most of the time about the operators and options this search engine offers as they frequently employed a series of

terms or a given sentence as a search expression to select one option or another intuitively.

Their use and interpretation of the search results coincided with their shortcomings during the interrogation processes. Only 33% of the students were able to correctly identify the different elements shown in the list of hits and to use them properly. Along the same lines, only 40% of these students went on to a second page of hits, and only 3% went on to a third page if they thought it necessary. The navigation processes they started to access original documents with were performed in the same window or tab where the search results were displayed and, on most occasions, this was the reason why students seemed disoriented, leading them to recommence or abandon their search processes. They simply made a decision in virtue of the higher or lesser extent they needed the information they were searching or depending on their weariness of obtaining poor results. As Table 1 shows, when students did not obtain relevant results with Google, they used no other information sources because they did not know they existed.

Their use of library catalogues was rather limited and clearly insufficient. We noticed they only accessed the University of Zaragoza Library OPAC, and OPAC was only used to check the availability of material to loan, and never as a scientific information tool. However, it is logical that they do not use this tool very much because their knowledge of specialised resources is limited. Therefore it is impossible to assess the users' application and satisfaction with these tools in this first stage, at least in terms of them employing such resources as an information retrieval tool.

In relation to this perception, the students have a vague notion of the quality information concept. Although this concept has been the object of theoretical and practical formulations in highly structured information settings, studying it proves much more complex in non homogeneous and poorly structured information settings. The students did not show concern for the quality of the information, except for one of its basic aspects: reliability. The review and assessment of the information quality criteria were carried out as a work group, and the results enabled us to state that the students almost totally lack the logic and basic criteria which may be applied constantly and homogeneously.

#### **4 Developing the teaching-learning process**

The subject that was used as a base for this study addresses the informational literacy of the Philosophy Degree students. Its objectives are to improve the information retrieval, processing and assessment processes in general, and skills in searching, selecting and incorporating scientific information and documentation in academic work in particular. In accordance with this, the course subject established learning activities to:

1. Develop advanced information retrieval processes.
2. Identify and make good use of the information retrieval interfaces.

3. Know the main scientific information and documentation resources for their discipline.
4. Use optimised information retrieval processes.
5. Include the scientific information obtained in other products or services.

The intention of combining theory and practice is, on the one hand, for students to gain a clear vision of the structure and content of the information resources they employ and, on the other hand, to learn the technical characteristics of both the interface and the retrieval language of all these resources. The expression of learning is carried out by developing a personal information retrieval and inclusion project about a specific theme of the degree the students were studying. The aim of the project is to create a personal bibliographical information resource by means of a bibliographic management platform which is fed with the information covered by the resources used in the learning process.

This project also entails having to write a technical report which specifies the information sources or resources used, the information search processes carried out, the problems considered in the interaction process, the qualitative assessment of the validity and usefulness of the information retrieved, and the problems encountered to include or re-use the information in other products or services. The information contained in the reports which was collected in the three academic years (2007-2010) have enabled us to extract a series of results which are dealt with in the next section.

## 5 Results of the project assesment

The information collected in the review of the results obtained in the projects and the corresponding reports has enabled us to state that the information search behaviours conducted by students have constructively progressed. We stress that this progress is limited by the contextual factors relating to former training, the students' perception of the subject, and all the students' personal objectives.

The use of information resources has led to a remarkable change as most students started to use specialised resources, as Table 3 depicts. The changes are significant if compared with the data presented in Table 1.

**Table 3.** Use of information resources.

	Google Scholar	OPACs	Scientific Search eng.	Reference DB	Repositories	E-Journals	Subject Gatew.	Bibliogr. servers
2007-08	12	10	12	11	10	12	6	2
2008-09	15	11	15	15	10	13	9	7
2009-10	14	12	14	14	12	14	7	4

The training students received has enabled them to carry out information retrieval processes in a higher number of resources, particularly bibliographic and referential

databases, as well as in repositories. In principle, this greater use could indicate an improvement in the scientific information assessment and retrieval processes. Nevertheless, when we break down the data presented in Table 3 and bear in mind other factors, we find factors that are external to the process and limit their success.

Firstly, the languages used are a barrier for the students. A large number of students (56%) only used resources in Spanish and, consequently, they only included Spanish references in their bibliographies. When questioned about this, many students mentioned that this limit was owing to them not knowing other languages. This language-based restriction also affects the resources used themselves as those students with language problems searched and selected scientific information almost exclusively in OPACs and referential databases whose content was in Spanish.

Secondly, the theme and object of the study and, consequently, the information search, lack definition. Students lack specific training in the methodology to perform scientific works. This deficiency makes a suitable formulation of the informative need difficult, and leads to an insufficient capacity to select the terms that could better represent it. This consideration is reflected in the set of terms used in search equations. The reports reveal that the students define an average of between four and six terms which they believe may cover their information needs (Figure 1). Then they use them systematically in search interfaces in the search field of all the fields, if available, or in the title and/or descriptor fields.

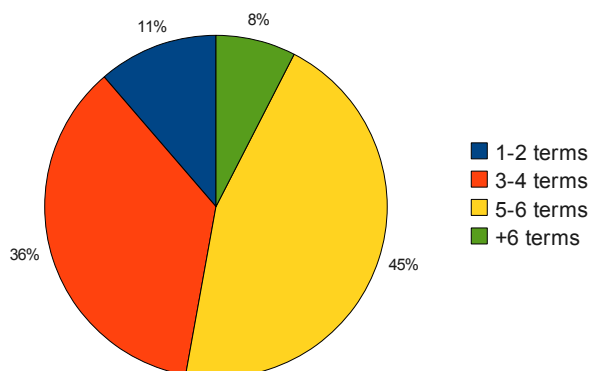


Fig. 1. Number of terms used in retrievals

When the students were asked about modifying the terms according to the relevance of the hits obtained, 64% reported they do not modify the terms if they do not obtain satisfactory hits because they think that the hits do not exist, whereas the remaining 36% modify terms if they consider the hits inadequate. The operator that students used in the equations was the AND-OR logic product, followed by the exact phrase operator. They barely use the remaining operators, and they do not consider the time limit operators given the longer durability of scientific publications on the matter.

Finally, we may also consider several matters from the reports regarding information retrieval and processing whose valuation by the students is totally qualitative:

- Information retrieval interfaces in specialised resources are complex and not very clear, and a model based on a simple interface like Google is preferred.
- The quality of the contents of the resources used is most irregular.
- The services offered by electronic journal suppliers are preferred.
- The contents offered by the OPACs are insufficient compared with other resources.
- Improved interoperability between resources would be desirable to automatically re-use contents.
- Repeating the contents in electronic journals, referential databases and repositories leads to confusion.

## **6 Conclusions**

Wilson [2] brought attention to the relative value of informational behaviour studies which were carried out with small or specialised groups of users, defending the need for studies in broader fields so that a more extensive outlook is available to make an assessment. However, we consider that this approach cannot exclude that used in works like the present one because information retrieval has a strong contextual component which affects users' behaviours. Generalisation tends to avoid the fact that retrieval processes are eminently personal and individual. As O'Brien and Symons pointed out [9], the course level achieved and the academic discipline covered influence users' decisions, although it has been detected that the Web and classmates have a greater influence.

Two kinds of conclusions may be drawn from the research carried out: the first relate to the individual processes carried out by students, while the second correspond to the group's social and educational context. As for the former kind, the following is stressed:

1. Students simplify retrieval processes in terms of both time and resources.
2. Simple interfaces with a few visual or interaction-type elements are preferred.
3. Terms for the search expressions are selected in the initial stage of the process, and the trend is not to modify the selection in subsequent situations.
4. Search processes do not tend to be extended by navigating or exploring.
5. The whole process is like a linear episode rather than a recurrent one, and reconsidering search actions is rare.

Second searches, that is, those relating to the context that are mainly of a social and education kind, particularly influence users' behaviours:

1. The confusion between digital literacy and informational literacy in the educational context means students are efficient when it comes to using tools, but inadequate when considering processes and assessing them.
2. Lack of training in scientific work processes has an impact on their lack of knowledge about them existing, their use and the possibilities that specialised resources offer.
3. The training for users that university libraries provide does not fulfil the objectives and results required.
4. The interest and development of social networks have led students to consider them and use them as a leading platform to make the most of Internet resources..

## Referencias

1. Fisher, K., Erdelez, S. y Lynnc, E.F. (eds.) *Theories of Information Behavior..* Medford, NJ: Information Today for the American Society of Information Science and Technology, 2005. (ASIST Monograph Series)
2. Wilson, T.D. Information needs and uses: Fifty years of progress. En B.C. Vickery, (ed.). *Fifty years of information progress: A Journal of Documentation review*, pp. 15- 51. London: Aslib (1994)
3. Wilson. T.D. Fifty years of information behavior research. *Bulletin of ASIST*, February/March 2010, [http://www.asis.org/Bulletin/Feb-10/FebMar10\\_Wilson.html](http://www.asis.org/Bulletin/Feb-10/FebMar10_Wilson.html) (2010).
4. Zhang, X., Anghelescu, H.G.B., y Yuan, X. Domain knowledge, search behavior, and search effectiveness of engineering and science students: An exploratory study. *Information Research*, **10**(2), paper 217, <http://informationr.net/ir/10-2/paper217.html> (2005)
5. Heinström, J. Fast surfing for availability or deep diving into quality – motivation and information seeking among middle and high school students. *Information Research*, **11**(4), paper 265. <http://informationr.net/ir/11-4/paper265.html> (2006)
6. Junni, P. Students seeking information for their Masters' theses: the effect of the Internet. *Information Research*, **12**(2) paper 305. <http://InformationR.net/ir/12-2/paper305.html> (2007)
7. Nicholas, D., Huntington, P. Jamali, H.R. Y Watkinson, A. The information seeking behaviour of the users of digital scholarly journals, *Information Processing & Management*, **42**(5), pp. 1345-1365 (2006)
8. Rowlands, I. y Nicholas, D. Understanding Information Behaviour: How Do Students and Faculty Find Books?, *The Journal of Academic Librarianship*, **34**(1), pp. 3-15 (2008).
9. O'Brien, H.L y Symons, S. The information behaviors and preferences of undergraduate students, *Research Strategies*, **20**(4), pp. 409-423 (2005).